```
                     3074. .3167,3234. .>3351)
                     /gene="v"
                     /product="vermilion"
     gene            <1856. .>3351
                     /gene="v"
     CDS             join(1856. .1881,1971. .2131,2187. .2320,2377. .2986,
                     3074. .3167,3234. .3351)
                     /gene="v"
                     /note="tryptophan-2,3-dioxygenase"
                     /codon_start=1
                     /product="vermilion"
                     /db_xref="PID:g2612996"
                     /translation="MSCPYAGSGNDHDDAAVPLSTEVGEIYGEYLMLDKLLDAQCMLS
                     KEDKRPVHDEHLFIITHQAYELWFKQIIFEFDSIRDMLDAEVIDETKTLEIVKRLNRV
                     VLILKLLVDQVPILETMTPLDFMDFRKYLAPASGFQSLQFRLIENKLGVLTEQRVRYN
                     QKYSDVPGGDEDALTAIRSSEQEPSLLELVQRWLERTPGLEESGFNFWEKFQQSVDKF
                     LEAQVHSAMEEPVERAKNYRLMDIEKRREVYRSIFDPAVHDALVRRGDRRFSHRALQG
                     AIMITFYRDEPRFSQPHQLLTLLMDIDSLITKWRYNHVIMVQRMIGSQQLGTGGSSGY
                     QYLRSTLSDRYKVFLDLFNLSTFLIPREAIPPLDESIRKKLINKNV"
BASE COUNT     1108 a    719 c    701 g   1038 t
ORIGIN

  Query Match           6.4%;  Score 19;  DB 18;  Length 3566;
  Best Local Similarity 100.0%;  Pred. No. 5.52e+00;
  Matches    19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1041 TTATTTAAATAAAAATATT 1059
         ||||||||||||||||||
Cp    125 TTATTTAAATAAAAATATT 107


RESULT   15
LOCUS       AF028835     3626 bp    DNA             INV       25-JUN-1998
DEFINITION  Drosophila pallidosa vermilion (v) gene, complete cds.
ACCESSION   AF028835
NID         g2612997
KEYWORDS    .
SOURCE      Drosophila pallidosa.
  ORGANISM  Drosophila pallidosa
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea;
            Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 3626)
  AUTHORS   Stephan,W., Xing,L., Kirby,D.A. and Braverman,J.M.
  TITLE     A test of the background selection hypothesis based on nucleotide
            data from Drosophila ananassae
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 95 (10), 5649-5654 (1998)
  MEDLINE   98245135
REFERENCE   2  (bases 1 to 3626)
  AUTHORS   Stephan,W., Xing,L., Kirby,D.A. and Braverman,J.
  TITLE     Direct Submission
```



```
  JOURNAL   Submitted (06-OCT-1997) Zoology, U. of Maryland, College Park, MD
            20742, USA
FEATURES             Location/Qualifiers
     source          1. .3626
                     /organism="Drosophila pallidosa"
                     /db_xref="taxon:69217"
     mRNA            join(<1916. .1941,2031. .2191,2247. .2380,2437. .3046,
                     3134. .3227,3294. .>3411)
                     /gene="v"
                     /product="vermilion"
     gene            <1916. .>3411
                     /gene="v"
     CDS             join(1916. .1941,2031. .2191,2247. .2380,2437. .3046,
                     3134. .3227,3294. .3411)
                     /gene="v"
                     /note="tryptophan-2,3-dioxygenase"
                     /codon_start=1
                     /product="vermilion"
                     /db_xref="PID:g2612998"
                     /translation="MSCPYAGSGNDHDDAAVPLSTEVGKFYGEYLMLDKLLDAQCMLS
                     KEDKRPVHDEHLFIITHQAYELWFKQIIFEFDSIRDMLDAEVIDETKTLEIVKRLNRV
                     VLILKLLVDQVPILETMTPLDFMDFRKYLAPASGFQSLQFRLIENKLGVLTEQRVRYN
                     QKYSDVPGGDEDALTAIRSSEQEPSLLELVQRWLERTPGLEESGFNFWEKFQQSVDKF
                     LEAQVHSAMEEPVERAKNYRLMDIEKRREVYRSIFDPAVHDALVRRGDRRFSHRALQG
                     AIMITFYRDEPRFSQPHQLLTLLMDIDSLITKWRYNHVIMVQRMIGSQQLGTGGSSGY
                     QYLRSTLSDRYKVFLDLFNLSTFLIPREAIPPLDESIRKKLINKSV"
BASE COUNT     1118 a    727 c    724 g   1057 t
ORIGIN

  Query Match           6.4%;  Score 19;  DB 18;  Length 3626;
  Best Local Similarity 100.0%;  Pred. No. 5.52e+00;
  Matches    19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1099 TTATTTAAATAAAAATATT 1117
         ||||||||||||||||||
Cp    125 TTATTTAAATAAAAATATT 107


RESULT   16
LOCUS       MUSGRIPGHS   4152 bp    mRNA            ROD       03-JUN-1992
DEFINITION  Mouse glucocortoid-regulated inflammatory prostaglandin G/H
            synthase (griPGHS) mRNA, complete cds.
ACCESSION   M88242
NID         g193637
KEYWORDS    glucocorticoid-regulated inflammatory prostaglandin G/H synthase.
SOURCE      Mus musculus (strain RIII) cDNA to mRNA.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;
            Murinae; Mus.
REFERENCE   1  (bases 1 to 4152)
```



```
  AUTHORS   O'Banion,M.K., Winn,V.D. and Young,D.A.
  TITLE     cDNA cloning and functional activity of a glucocorticoid-regulated
            inflammatory cyclooxygenase
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 89, 4888-4892 (1992)
  MEDLINE   92279232
FEATURES             Location/Qualifiers
     source          1. .4152
                     /organism="Mus musculus"
                     /strain="RIII"
                     /db_xref="taxon:10090"
     sig_peptide     1. .17
                     /partial
                     /gene="griPGHS"
     gene            1. .1915
                     /gene="griPGHS"
     CDS             101. .1915
                     /gene="griPGHS"
                     /codon_start=1
                     /product="glucocorticoid-regulated inflammatory
                     prostaglandin synthase"
                     /db_xref="PID:g193638"
                     /translation="MLFRAVLLCAALGLSQAANPCCSNPCQNRGECMSTGFDQYKCDC
                     TRTGFYGENCTTPEFLTRIKLLLKPTPNTVHYILTHFKGVWNIVNNIPFLRSLIMKYV
                     LTSRSYLIDSPPTYNVHYGYKSWEAFSNLSYYTRALPPVADDCPTPMGVKGNKELPDS
                     KEVLEKVLLRREFIPDPQGSNMMFAFFAQHFTHQFFKTDHKRGPGFTRGLGHGVDLNH
                     IYGETLDRQHKLRLFKDGKLKYQVIGGEVYPPTVKDTQVEMIYPPHIPENLQFAVGQE
                     VFGLVPGLMMYATIWLREHNNRVCDILKQEHPEWGDEQLFQTSRLILIGETIKIVIEDY
                     VQHLSGYHFKLKFDPELLFNQQFQYQNRIASEFNTLYHWHPLLPDTFNIEDQEYSFKQ
                     FLYNNSILLEHGLTQFVESFTRQIAGRVAGGRNVPIAVQAVAKASIDQSREMKYQSLN
                     EYRKRFSLKPYTSFEELTGEKEMAAELKALYSDIDVMELYPALLVEKPRPDAIFGETM
                     VELGAPFSLKGLMGNPICSPQYWKPSTFGGEVGFKIINTASIQSLICNNVKGCPFTSF
                     NVQDPQPTKTATINASASHSRLDDINPTVLIKRRSTEL"
     polyA_site      2052
                     /gene="griPGHS"
                     /note="alternative site"
     polyA_site      2256
                     /gene="griPGHS"
                     /note="alternative site"
BASE COUNT     1261 a    856 c    823 g   1212 t
ORIGIN

  Query Match           6.4%;  Score 19;  DB 28;  Length 4152;
  Best Local Similarity 100.0%;  Pred. No. 5.52e+00;
  Matches    19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   3327 TTAAAAATATTTTTATTTA 3345
         ||||||||||||||||||
Qy    102 TTAAAAATATTTTTATTTA 120


RESULT   17
```



```
LOCUS       U32740      11622 bp    DNA             BCT       29-MAY-1998
DEFINITION  Haemophilus influenzae Rd section 55 of 163 of the complete genome.
ACCESSION   U32740 L42023
NID         g1573572
KEYWORDS    .
SOURCE      Haemophilus influenzae Rd.
  ORGANISM  Haemophilus influenzae Rd
            Eubacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
            Haemophilus.
REFERENCE   1  (bases 1 to 11622)
  AUTHORS   Fleischmann,R.D., Adams,M.D., White,O., Clayton,R.A.,
            Kirkness,E.F., Kerlavage,A.R., Bult,C.J., Tomb,J., Dougherty,B.A.,
            Merrick,J.M., McKenney,K., Sutton,G.G., FitzHugh,W., Fields,C.A.,
            Gocayne,J.D., Scott,J.D., Shirley,R., Liu,L.I., Glodek,A.,
            Kelley,J.M., Weidman,J.F., Phillips,C.A., Spriggs,T., Hedblom,E.,
            Cotton,M.D., Utterback,T., Hanna,M.C., Nguyen,D.T., Saudek,D.M.,
            Brandon,R.C., Fine,L.D., Fritchman,J.L., Fuhrmann,J.L.,
            Geoghagen,N.S., Gnehm,C.L., McDonald,L.A., Small,K.V., Fraser,C.M.,
            Smith,H.O. and Venter,J.C.
  TITLE     Whole-genome random sequencing and assembly of Haemophilus
            influenzae Rd
  JOURNAL   Science 269 (5223), 496-512 (1995)
  MEDLINE   95350630
REFERENCE   2  (bases 1 to 11622)
  AUTHORS   Tatusov,R.L., Mushegian,A.R., Bork,P., Brown,N.P., Hayes,W.S.,
            Borodovsky,M., Rudd,K.E. and Koonin,E.V.
  TITLE     Metabolism and evolution of Haemophilus influenzae deduced from a
            whole-genome comparison with Escherichia coli
  JOURNAL   Curr. Biol. 6 (3), 279-291 (1996)
  MEDLINE   96398784
REFERENCE   3  (bases 1 to 11622)
  AUTHORS   White,O., Clayton,R.A., Kerlavage,A.R. and Fleischmann,R.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUL-1995) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
REFERENCE   4  (bases 1 to 11622)
  AUTHORS   White,O., Clayton,R.A., Kerlavage,A.R. and Fleischmann,R.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-SEP-1997) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
  REMARK    The H. influenzae sequence has been updated by R. Fleischmann. New
            database matches have been assigned, product names have been
            improved, and a number of frame shifts have been corrected. We
            gratefully acknowledge the work of Tatusov et. al. We have
            incorporated their annotation into the /notes fields of the
            corresponding H. influenzae genes
REFERENCE   5  (bases 1 to 11622)
  AUTHORS   White,O., Clayton,R.A., Kerlavage,A.R., Fleischmann,R.D.,
            Peterson,J., Hickey,E., Dodson,R. and Gwinn,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAY-1998) The Institute for Genomic Research, 9712
```

```
          Medical Center Dr, Rockville, MD 20850, USA
  REMARK  The whole genome was shifted by 588 nucleotides for a new start
COMMENT   On Sep 30, 1996 this sequence version replaced gi:1221260.
FEATURES            Location/Qualifiers
     source         1. .11622
                    /organism="Haemophilus influenzae Rd"
                    /db_xref="taxon:71421"
     gene           complement(373. .2346)
                    /gene="HI0583"
     CDS            complement(373. .2346)
                    /gene="HI0583"
                    /note="similar to SP:P08331 GB:M13464 PID:145584 GB:U00096
                    PID:1790658 percent identity: 62.40; identified by
                    sequence similarity; putative"
                    /codon_start=1
                    /transl_table=11
                    /product="2',3'-cyclic-nucleotide 2'-phosphodiesterase
                    (cpdB)"
                    /db_xref="PID:g1573573"
                    /translation="MMNRRHFIQISATSILALSANRPAMAKGKSDVDLRIVATTDVHS
                    FLTDFDYYKDAPTDKPGFTRAASLIRQARAEVKNSVLVDNGDLIQGNPIADYQAAQGY
                    KEGKSNPAIDCLNAMNYEVGTLGNHEFNYGLNYLADAIKQAKFPIVNSNVVKAGTEEP
                    YFTPYVIQEKSVVDNQGKTHKLKIGYIGFVPPQIMVWDKANLQGKVETRDIVKTAQKY
                    VPEMKKKGADIVVALAHTGPSDEPYQEGAENSAFYLADVPHIDAVIFGHSHRLFPNKE
                    FAKSPNADIVNGTVKGIPESMAGYWANNISVVDLGLTEHKGKWIVTSGKAVLRPIYDI
                    ETKKALAKNDPEITALLKPVHEATRKYVSQPIGKATDNMYSYLALLQDDPTIQIVNQA
                    QKAVVEKVAPSIAAMAGLPILSAGAPFKAGGRKNDPTGYTEVNKGKLTFRNAADLYLY
                    PRTLVVVKATGEQLEFWLECSAGMFKQIDPTSDKPQSLIDWEGPRTYNFDVIDGVNYE
                    YDLTKPARYDGECKLINPESHRVVNLTYQGKPVDPKAEFLIATNNYRAYGNKFPGTGD
                    KHIVYASPDESRQILADYIKATSEKEGSVNPNADKNWRFVPITGNDKLDVRFETSPSE
                    QAAKPIAEKAQYPMKQVGTDEIGFAVYQIDLSK"
     gene           complement(2440. .3711)
                    /gene="HI0584"
     CDS            complement(2440. .3711)
                    /gene="HI0584"
                    /note="similar to GB:U00096 SP:P77357 PID:1742205
                    PID:1742209 PID:1787599 percent identity: 40.34;
                    identified by sequence similarity; putative"
                    /codon_start=1
                    /transl_table=11
                    /product="hydrolase"
                    /db_xref="PID:g1573574"
                    /translation="MNLDLNQLVKWHREFHRFPEIGWSEFWTTSRIADYLEDLDCPEI
                    FLGKQIINPDFVRGRKQAVVDKGLANAKAYGANEKWLEKMEGYTGCVALPDSGKPGKT
                    IALEFDIDCVNTETRGPEHIPNKEGFASINDGFMHACGHDSHITIGLGVALWIAQNK
                    DKLTGKVKIVFQPAEEGVRGAAAIAQSGIIDDADYFASSHISFCANTGTVIANPRNFL
                    STTKIDIRYKGKPAHAGAAPHLGRNALLAAAHTVTQLHGIARHGKGMTRINVGVLKAG
                    EGRNVIPSSAELQLEVRGENKAINEYMTEQVMQIAKGISISFNVAYETEIVGEAVDMN
                    NDVELIKLIEEISLEQPQINNVNSDYAFNASEDATILGRRVQEHGGKAIYFILGADRT
                    AGHHEAEFDFDENQLLTGVNIYTSLVQKLLS"
     gene           complement(3783. .4622)
```



```
                    /gene="HI0585"
     CDS            complement(3783. .4622)
                    /gene="HI0585"
                    /note="hypothetical protein; identified by GeneMark;
                    putative"
                    /codon_start=1
                    /transl_table=11
                    /product="H. influenzae predicted coding region HI0585"
                    /db_xref="PID:g1573575"
                    /translation="MMSPGSSHSAMISEMSGLTITQVNLSHAPYNMIAGAIGAVVLTI
                    LALVPKDYGEQHRQAYLAEQKESEIKVIEGVNVLYALAPLIPLVILVIGGTSLQQVPG
                    LEWTKMGVPQAMLIGAITGIIVTRISPVKITEEFFNGMGNSYANVLGIIIAASVFVAG
                    LKSTGAVDAAISPLKESNEPVRWGATIGPPLMGLITGSGDAAAIAFNTAVTPHAVELG
                    YTHVNLGMAAAIAGAIGRTASPIAGVTIVCAGLAMVSPVEMVKRTAPGMILAVLPLAL
                    FML"
     gene           complement(4640. .5077)
                    /gene="HI0586"
     CDS            complement(4640. .5077)
                    /gene="HI0586"
                    /note="hypothetical protein; identified by GeneMark;
                    putative"
                    /codon_start=1
                    /transl_table=11
                    /product="H. influenzae predicted coding region HI0586"
                    /db_xref="PID:g1573576"
                    /translation="MELFKSIVAVIGIIATIYFLIKKAETRTVLIGVGLIMSILTLNP
                    MGALDAFAKSMTSGGLIMAICSSMGFAYVMKYTQCDTHLVHLLTKPLSGLKFPLIPIA
                    TIITFFINIAIPSAAGCAAAVGATLIPVLKSAGVRPATAGQLF"
     gene           complement(5144. .5824)
                    /gene="HI0587"
     CDS            complement(5144. .5824)
                    /gene="HI0587"
                    /note="similar to SP:P32666 PID:396356 GB:U00096
                    PID:1790452 percent identity: 41.40; identified by
                    sequence similarity; putative"
                    /codon_start=1
                    /transl_table=11
                    /product="peptidase E (pepE)"
                    /db_xref="PID:g1573577"
                    /translation="MPLKRIINMKNMLLLSSSKYKNTGYLEHTIPWLQNFLADYRGKT
                    IAFVPTAGVSRTFDETEKTVQNALSDLGMNIVSVHRGKQHRDIIEQADVIAIGGGNTF
                    CLLKQLYEHNLIDIIREEVNNGTPYFGWSAGANVAGSSINTTNDMPITTPPSFQALQL
                    FPHQINPHFISGKMQGHNGESREERLAEFLLVNPTALVTALPEGSALHIQNGMGNRIG
                    RKSYPVLQ"
     gene           5924. .7159
                    /gene="HI0588"
     CDS            5924. .7159
                    /gene="HI0588"
                    /note="similar to SP:P37113 GB:S67784 GB:X74289 PID:460895
                    SP:Q53389 percent identity: 35.94; identified by sequence
                    similarity; putative"
```



```
                    /codon_start=1
                    /transl_table=11
                    /product="N-carbamyl-L-amino acid amidohydrolase"
                    /db_xref="PID:g1573578"
                    /translation="MSINLNRVQNLIEKLAPISSVPNELTRLAPTEEDEKAHNMIIEL
                    CKEYDLSIRRDSIGNLFIRKAGKEDFLPAVAPGSHIDTVVNAGKFDGPLGSVAGLEIL
                    LQLCEQNIQTRYPLELIIFTCEESSRFNVPATLGSKVMCGIVNQEKLSSLRDKQGKGLS
                    EAMAEVGMNFNLVNQAKRDAKEFKCFFELHIEQGPRLENEGKTIGVVTGIAAPIRAIV
                    KIKGQADHSGATAMHYRHDALLGGSELSLAIERAAIQAGHSTVATVGNITAKPGVMNV
                    VPGYCELLVDIRGTHVQARDSVFELLQEEISKVSEKRGLLIELQLISKDNPIILPENM
                    VNQIAETAHSLGYSYEIMPSGAGHDAMHMATLCPTGMIFIPSHLGISHNPLEFTDWKD
                    IEAGIKVLQRVILEQAEVC"
     gene           7374. .7793
                    /gene="HI0589"
     CDS            7374. .7793
                    /gene="HI0589"
                    /note="similar to SP:P46187 PID:987645 PID:1045631
                    PID:1050877 GB:U00096 percent identity: 29.01; identified
                    by sequence similarity; putative"
                    /codon_start=1
                    /transl_table=11
                    /product="sigma factor regulatory protein, putative"
                    /db_xref="PID:g1573579"
                    /translation="MMSYDAETGIAKVKCQSQSTCGACSARETCGTESLSELNGKRGE
                    HIPTLETITPLRTDQMVEIGLEEKSMLPSALLMYIVQLPTLLVATLLSSYISENELIR
                    AILIFMLTALSFVMVKRYTRKLGQQTEFQSVLLRVLF"
     gene           complement(7895. .9202)
                    /gene="HI0590"
     CDS            complement(7895. .9202)
                    /gene="HI0590"
                    /note="similar to GB:M64495 SP:P24170 PID:147332 GB:U00096
                    PID:1651299 percent identity: 76.83; identified by
                    sequence similarity; putative"
                    /codon_start=1
                    /transl_table=11
                    /product="putrescine-ornithine antiporter (potE)"
                    /db_xref="PID:g1573580"
                    /translation="MSAKSNKIGVVQLTILTMVNMMGSGIIMLPTKLAEIGTISIVSW
                    LVTAVGSTALAYAFAQCGMFSKKSGGMGGYAETYSPGKAGNFMANYTYGVSLVIANTAI
                    AISAVGYGSELFGTILSPLSIALWTIPTLWLATVLNFGGARITGNISSPTIWGVIIPV
                    VGISIIGWKWFDGSMTVNSWNPHNVPTFEAIGVSISMTLWAFLGLESACANADAVENP
                    EKNVPIAVLGGTLSAAVIYIVSTNVIAGIVPNLELANSTAPFGLAFAHMFDETVGKVI
                    MGLMVMSCFGSLLGWQPTIAQVFKSSAEEGYFPAFFKLTSKDAPVVGHITITALQTL
                    LSLMTISPSLNKQFNVLVDLAVVTNVIPYLLSMAALAVLKAENVAPQKYKITVFVAP
                    IGSLYSIYALYAAGEQAMLYGSIVTFIGWTLYGFVSYKFDLKK"
     gene           complement(9307. .11469)
                    /gene="HI0591"
...
Note: remainder of annotations omitted.

  Query Match          6.4%;  Score 19;  DB 15;  Length 11622;
```



```
    Best Local Similarity 100.0%;  Pred. No. 5.52e+00;
    Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2421 TAAAAATATTTTTATTTAA 2439
          ||||||||||||||||||
Qy    103 TAAAAATATTTTTATTTAA 121


RESULT   18
LOCUS       CELF08D12   33268 bp    DNA              INV       05-DEC-1996
DEFINITION  Caenorhabditis elegans cosmid F08D12.
ACCESSION   U80840
NID         g1707182
KEYWORDS    .
SOURCE      Caenorhabditis elegans strain=Bristol N2.
  ORGANISM  Caenorhabditis elegans
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Nematoda;
            Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditidae;
            Caenorhabditis.
REFERENCE   1  (bases 1 to 33268)
  AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
  TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
  JOURNAL   Nature 368 (6466), 32-38 (1994)
  MEDLINE   94150718
REFERENCE   2  (bases 1 to 33268)
  AUTHORS   Le,T. and Waterston,R.
  TITLE     The sequence of C. elegans cosmid F08D12
  JOURNAL   Unpublished (1996)
REFERENCE   3  (bases 1 to 33268)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-DEC-1996)
COMMENT     Submitted by:
                    Genome Sequencing Center
                    Department of Genetics, Washington University,
                    St. Louis, MO  63110, USA, and
                    Sanger Centre, Hinxton Hall
                    Cambridge CB10 IRQ, England
                    e-mail: rw@nematode.wustl.edu and jes@sanger.ac.uk
```

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:

The 5' cosmid is F29A7, 200 bp overlap;3' end lies in a gap. Actual start of this cosmid is at base position 197 of CELF08D12; actual end is at 33268 of CELF08D12. This cosmid lies in an unanchored cluster, the orientation of which is unknown.

NOTES:

Coding sequences below are predicted from computer analysis, using the program Genefinder(P. Green and L. Hillier, ms in preparation).

FEATURES Location/Qualifiers

source 1. .33268
/organism="Caenorhabditis elegans"
/strain="Bristol N2"
/db_xref="taxon:6239"
/chromosome="II"
/clone="F08D12"

gene 256. .3879
/gene="F08D12.6"

CDS join(256. .335,380. .1286,2008. .2239,2288. .2499,2662. .2844, 2964. .3214,3291. .3879)
/gene="F08D12.6"
/codon_start=1
/evidence=not_experimental
/db_xref="PID:g1707188"
/translation="MATSFPILRLPNKIIRNVVENLSFYNQLTFSLCSKATKSCVTPL RHKIRKINITVDHPLKIHIFSSNSVMTYLSFYKYNFDLWNVQSCGSIIATEFDTNFFG TESVREIKGPGLRDCIDHIMEIFNHSKIFEVHINGVMGIQNLKRVCEAMKGLTARKIF LKNWTDEIVQNVICSIQHKGISFNKKLISNSSSLNHEILVRNFFSFSASFPEHNFNLN DLLIMNSTHLRISSQLAFEDLNIFLKHLKTGLNQTLIGAQFLKIPQGPAWDDVKVLM NGIPHSRYRNSKLVSQDDNYQSGDCCMINQDDERRISLTWKWTQHDGSIEIAISCDSV TVPETMLENCINNDQHKKIITFSLPPERHLMPNHKIHCLNLGSVSYELPGRGKNLQLD DVLLMNCSKLCLFPSQPNRQLETVELHRTNEGIWKVEDSMKGISYSIWNKNPGNNHKL ZKRIKPTQNDETMMHLDWKLDSIRCSTLCVPLTCPPPPRARNVMVPDAGELPCGRLVE HHLPFADEKLGAVELKPVLFALEDQAFQVVQYADLPEIFNILPICRFAGNFIIQQVAD LPICRRCSIPTIYRFADLPENCIFDKLSACQFADLPISPIFNSGNLPICRKFQFRHKL QLKRNIKYIALDASSGVSIRLQLPDREQINLDVKTDRFDRCNLNFWEIANVTYRQPTR GYQQARYEQEFCWKKNGFGLRDWIDHTMEIFNLPVIDQFYLSGTVLSLADTRQIGEAL NGLKVTSMDFWCYTEDRVYNIRRNLETVGLIFAHPDSEHFSKLSTQKILSENQTSVLH RSPDFALMAEMRYFVDEASTACITLNDLLVSNCTKLIITDSTLKSEELNIFLKHWING SIQTCEFLKLETSVRQFNNMETVLKGVAHKTASKVRDGAPGARDMFIIPRIDLPNLSL TTIPTDGRYSVLTIDARCG"

gene complement(9283. .10291)
/gene="F08D12.11"

CDS complement(join(9283. .10156,10212. .10291))
/gene="F08D12.11"
/codon_start=1
/evidence=not_experimental
/db_xref="PID:g1707193"
/translation="MAACFPLLRLSSKDLRNVLQNMGILNLFSPSLCSTKAHVKLLNS TIQSIELHLSTGIGIAVWSPTQDRIELVFEKDQCELAKVVNTNALGHLKTRSFWEKTG FGLRDWIEHVMEIFNHSAIDMFRIVGSRHAESYLKRIHKTTDGLEVGTMEFVYWTEER VHPMRRTFQPTDVSLTFPDEENFTPKSTQRILLENQHFLMYRWADSEPRFTLNDLLAI NCSQLFIPESTLTAKDLNIFLKHWIIGFIGNLNMVQLKKHDGRMWSLEAVLNDIDYTI SAELAEDIQTVVIPRYRLSELFLTYTTPPNMGSILTIYRCG"

gene 12080. .12623
/gene="F08D12.5"

CDS join(12080. .12265,12312. .12459,12514. .12623)
/gene="F08D12.5"
/codon_start=1
/evidence=not_experimental
/db_xref="PID:g1707187"
/translation="MIPYSHGHRLLAMLNNAHLQRSVPNTDLSQPITSCPASTVTLPT CEVGWKLFPRYDTTLCLTLGINRKERDGYLLQFDQISSTVACRWDGSKFIDGILSGLE SDEERAYASDRAYQIMLQNMNETLFEFPSSERLGWGVWSGVPNNT"

gene 13505. .14569
/gene="F08D12.4"

CDS join(13505. .13607,13717. .13760,13817. .13942,13988. .14135, 14189. .14301,14346. .14504,14555. .14569)
/gene="F08D12.4"
/codon_start=1
/evidence=not_experimental
/db_xref="PID:g1707186"
/translation="MTLLRGTVRTRTSCQLVSAIGTQNCTVWEGYEVSEITPGNSNTS YITWKRSVPNTNLSQPITSCPASTVKFPTCKVGWKLFPRYNTTWCLTLGINRRERDGY LLQFDQSFSTVACRWDGSEFIDGRISGLESNEERAYATDRAYQIMKQNMRETLFVFPS NVRLGWGVWVDGVKDVAGQVKYSDTTLNLKLGYKWDIGQPNWVSRKQACVMLMVPVNQ SSRSNGSLRTAEYLRSI"

gene 14842. .16079
/gene="F08D12.3"

CDS join(14842. .14890,14943. .15059,15108. .15218,15263. .15385, 15455. .15565,15610. .15735,15793. .15897,15952. .16079)
/gene="F08D12.3"
/note="weak similarity to fibrillin 1 precusor in one region and drosophila fulin-like protease 2 precursor (SP:pH94375) in different region; coded for by C. elegans cDNA yk72a1.3; coded for by C. elegans cDNA yk72a1.5"

PICRLAENSSNLPICRKSSIPTIFRVARNVQFKQFAVLPEIFNNLPICRFAGNFNCAN LPTSRNFSVFSGKFGKLSSCGICRFSGNFQWCIPQSTLKSCLCILKKLQKLHQIRVIR LYVYPVDLPLLRHNTTFQSFTLKFIERMVQSIAVVLKASRHSCEHINQVVIVEVGAQD GYSLLGAFLANDLEEAEPA"

gene complement(3583. .4404)
/gene="F08D12.7"

CDS complement(join(3583. .3912,3971. .4144,4201. .4404))
/gene="F08D12.7"
/codon_start=1
/evidence=not_experimental
/db_xref="PID:g1707192"
/translation="MLKPTLTFYLTLLYTLSPVSSQNSTEPCDSDKCLLMVAELGPII SKHEHVKPRAETLKKMNDLCTDIKTCFKAATCKKGAHTTESMRDACDGLELMTGNVQP CMTKFYLAVYNEKTNCTGERDYLTDDLPARRVSTTSGRFCPFQVIGKECSEETVAILS SNFNVDNLINVLTVPGGFQDNCNRLHHAFNKLQCEALESGIVAKERETNWVDVQSND TNLVEPLQLFEDAEARP"

gene complement(5081. .6175)
/gene="F08D12.8"

CDS complement(join(5081. .5999,6096. .6175))
/gene="F08D12.8"
/codon_start=1
/evidence=not_experimental
/db_xref="PID:g1707191"
/translation="MATSFPVLCLPNKNIRDVLQNMSTCNLIAFSLCSKATKSHVMHL KNDIQQISIKAFLSIDINILSHNQEVTSLVFFKKDYGRWNFFSCESVTATETLTPLEP ESDRVWVERGFKLRDWIDHVMEIFNCSRIMEIYISRIGHEEHYISLSNAVKELKTETI FFSFCSDTVVQSVVANVQSNDISFNLSFNIEPSMQKHKVLTRHFERVSFSKDDTREAL TLNDLLITNVSQLNISAHHLSFEELNIFIRHLISGSNRRLKIVGIVCAEATRNMQILL KGISHSVYSAELKIVARNENDYQLGDTIYVYQNCGRRLAMQTYLFEGVEKIVMRVY"

gene complement(6461. .7527)
/gene="F08D12.9"

CDS complement(join(6461. .7400,7445. .7527))
/gene="F08D12.9"
/codon_start=1
/evidence=not_experimental
/db_xref="PID:g1707190"
/translation="MAARFPLLLRLPEKNLRIVLEQMGPLDLLAFSLCSKITKAHTTQ LKRRIDNLKLHASIVVNISFRLTNQEEIRLDVKTDRFDRWNLNFCEIANVTYTVGYQQ QFCWKKKGFGLRDWINHIMETFNYPVIDRLRLHGYMPPLANSRQISKVLNGLKVKNMD FWNYTEDGVPNIRSNLEIVALTLAPPDSEHFSELSTRRILSENQTNMIHHSMPFMAEM GVDRVLYSCITVNDLLVNNCTQLTISRSILTSKDMNIPLKHWINGSIKTLEFLELQTY VQTQWNMKILLKDIDHKIPCRVENGVPRGAKMFRIPRIDSSNLFMTYIVHDDRKYVLK IDARCG"

gene complement(7788. .8899)
/gene="F08D12.10"

CDS complement(join(7788. .8763,8814. .8899))
/gene="F08D12.10"
/codon_start=1
/evidence=not_experimental
/db_xref="PID:g1707189"
/translation="MATRFPLLLFCLPEKNLRIVLEQMRFLDLLAFSLCSKITKAHTT

/codon_start=1
/db_xref="PID:g1707183"
/translation="MSRTFLTIFFCSAVAPGATSSASTCKSPTGPAIGGACAAGQVVI QDTCCWEDDVYKLPAESCKNAFGPAIMGYCSDGQVVVGDICCEAQDVYDSSTATCQEG KSVGPAGAGGSCPQGASITKGSVCCWNEGIQDAPASACKSSASPAILGICPSGQVLIG

...

Note: remainder of annotations omitted.

Query Match 6.4%; Score 19; DB 18; Length 33268;
Best Local Similarity 100.0%; Pred. No. 5.52e+00;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 19043 TCAGTTTAACAACCAAAAA 19061
|||||||||||||||||||
Cp 294 TCAGTTTAACAACCAAAAA 276

RESULT 19
LOCUS CER07H5 33802 bp DNA INV 21-SEP-1998
DEFINITION Caenorhabditis elegans cosmid R07H5, complete sequence.
ACCESSION Z81107
NID g1628081
KEYWORDS HTG.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 33802)
AUTHORS Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M., Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J., Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A., Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M., Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N., Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B., O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A., Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E., Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M., Vaughan,K., Waterston,R., Watson,A., Weinstock,L., Wilkinson-Sproat,J. and Wohldman,P.
TITLE 2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans
JOURNAL Nature 368 (6466), 32-38 (1994)
MEDLINE 94150718
REFERENCE 2 (bases 1 to 33802)
AUTHORS Lennard,N.
TITLE Direct Submission
JOURNAL Submitted (06-NOV-1996) Nematode Sequencing Project, Sanger Centre, Hinxton, Cambridge CB10 1RQ, England and Department of Genetics, Washington University, St. Louis, MO 63110, USA. E-mail: jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT Coding sequences below are predicted from computer analysis, using

predictions from Genefinder (P. Green, U. Washington), and other available information.
For a graphical representation of this sequence and its analysis see:-
http://webace.sanger.ac.uk/cgi-bin/display?db=wormace&class=Sequence &object=R07H5
Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.
IMPORTANT: This sequence is not the entire insert of clone R07H5. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.
The true left end of clone R07H5 is at 1 in this sequence. The true right end of clone R07H5 is at 6112 in sequence Z68335.
The true left end of clone C29F4 is at 33698 in this sequence. The true right end of clone C08F8 is at 9591 in this sequence. The start of this sequence (1. .107) overlaps with the end of sequence Z73103.
The end of this sequence (33698. .33802) overlaps with the start of sequence Z68335.

FEATURES Location/Qualifiers

source 1. .33802
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="IV"
/clone="R07H5"

gene 1996. .2870
/gene="R07H5.1"

CDS join(1996. .2252,2300. .2582,2634. .2870)
/gene="R07H5.1"
/note="predicted using Genefinder"
/codon_start=1
/db_xref="PID:e275844"
/db_xref="PID:g1628086"
/db_xref="SPTREMBL:Q93930"
/translation="MTDPSSSPIRPDMVEAARKFMLTPKVKETPPEEQRQFLLGKGVS
EAEILEARASIPPEQLRSQIGTENMIPIGGHPQMMIAPKQNGIVSFAQSAVVLGSISY
AGYRFVRSYVLPRFFDIPDPATEEIRQLQSQVDDLQNSIKFIMDSVSQTTQQLATQQS
EISRALYSVANRDADLNRVESGISTIKSLLLSQHNFAPIVTPSVTSSIPSWQQSATPT
SATSITTTTTSEAASGYATPPANFKESSLADELADIENIE"

gene 3072. .4774
/gene="R07H5.3"

CDS join(3072. .3252,3833. .3989,4038. .4193,4239. .4302,

4351. .4488,4539. .4652,4697. .4774)
/gene="R07H5.3"
/note="predicted using Genefinder"
/codon_start=1
/db_xref="PID:e276200"
/db_xref="PID:g1628087"
/db_xref="SPTREMBL:Q93931"
/translation="MVEKPDINFQNKNVLHEINICYLGIMNKSLIALGAQTARRFLSS
SSKNTIGGGDANRADTGVLDGYHINPLEESDMTDRSRISMLSTEMLEAKQIGVRGLSC
YGPRLLDGTFLYGPIALFPKTALSWRVPTPEDITPRSLALFAALEPKIDILVLGVGDK
KNIDKVRASVAPFLREHKIGLEIMDTEDAIATFNFLNAEGRTVGAALYPPDDMVVTDK
EYGRALALLKGWDTVEENPLLTGLSDTINGAEDLVKRLWSGDEKSWQSARQKVLESPS
ERKERMQQEMNDKEKKRIE"

gene 5907. .7059
/gene="R07H5.4"

CDS join(5907. .5966,6016. .6129,6613. .7059)
/gene="R07H5.4"
/note="cDNA EST EMBL:T02072 comes from this gene"
/codon_start=1
/db_xref="PID:e275846"
/db_xref="PID:g1628089"
/db_xref="SPTREMBL:Q93932"
/translation="MKLHLLTVLLSLVLLSVNAKPSKEKDFQFSDGNLGTPMEKYFEQ
MGFEQKERNEQIINKPSSEDESSTATIPNTDHHQLTWIPNDAYLFIGKTRYYFRENWI
KLPPIGMKVTEVDSIVQMPPEMNHQEVPELRRETEKKEGYVDPKEFEKIHDTIKFSFTM
DTFTTYPHKYTTEILLVTAVIVIICIAVGLHHYKKKVEEKDQLHFA"

gene complement(7741. .8488)
/gene="R07H5.7"

CDS complement(join(7741. .8180,8317. .8488))
/gene="R07H5.7"
/codon_start=1
/db_xref="PID:e283343"
/db_xref="PID:g1694739"
/db_xref="SPTREMBL:P90944"
/translation="MDFIHLGQILTRDSEGYWKLIIRMQLTRFIDHILELFHLSGLHE
LSFCECNYDYSTNQPKRPQSVIVSNFTFIKIWEYIPLKINDLLASNSSTICMGYQVPP
KDICTFLKHWIHGSNEQLKQLIMMFPLSFNKNDFIPIALNGIKYQLAPVERKLVFIQD
TTKWYDKEVEVEGGLDIQRIDGTNATITLPEEPAWSTILYVWD"

gene complement(8567. .15431)
/gene="R07H5.6"

CDS complement(join(8567. .8677,8884. .9306,9375. .9510,
9554. .9767,14748. .15015,15147. .15431))
/gene="R07H5.6"
/note="Similarity to C.elegans mut-2 mutator strain transposon Tc5 transposase"
/codon_start=1
/db_xref="PID:e281777"
/db_xref="PID:g1673485"
/db_xref="SPTREMBL:Q93933"
/translation="MPFRADLIVGGLVEEPGHEYNPSFMSREEQKILERVRRILRDGK
THELIVYSEDEEEIVGGIVEDENWTADDDGPTTYALPDKVNFSGNLVDVETVEKAIEF

MMDTKHNKFRPMSSVQHRYRPIKSEQDMKKLRKFARNRQFSPSIFSVHNYLFVLEKVN
CSRASSFAKLSELLRAKVFETMEANDWLAKWKKSNGLVSRHITTFVTKASFANKEKTE
EAAKKYVDDMKLELQTLDPNTAFNCDQSGFTKEQYCKRTLAPRGVKRVERLVQSKDAL
SHSYTILPMLSVTGKLAPKLYIVLQEKGGKFPKKGHCSPSNLIVRANSSHIMTKDLMR
DWVESAVCDPAMPKETVLLLDSWPAWKNERDVQDAALYGHTVHVRTIPPGATSFIQPC
DIYLFCPMKNFVKKVQSHKKFFQTGESRPGLNFAPNLVPFSAPPWGKIFLRDYIFHFK
SRKTLCCGIEYEYAICLCGFGAVNINLP"

gene complement(11613. .12632)
/gene="R07H5.5"

CDS complement(11613. .12632)
/gene="R07H5.5"
/note="similar to transposase"
/codon_start=1
/db_xref="PID:e281776"
/db_xref="PID:g1673484"
/db_xref="SPTREMBL:Q27282"
/translation="MTIIKLERRDVRLLLLYEFRLGHSAMEAERNICGAMGEGALSYN
TAKSWFQKFKNGDFSLEEIERSGRPVELNEEDLVKLVEEEPRLSLREMEEKLECCHST
IARHLGRLGFTSKLGTWVPHELSASQKLTRVNVCTQLLTFRRKFDWLNNLVTGDEKWV
LYVNHSRKRQWLPIGEKGIPTPKPDLHPKKIMICVWWGVQGPVHWELLPTNKTITADY
YCAQLDRVAEKTNGKYEKLYFLHDWARPHVAKKTFQKLQDLGWTVLPHPPYSPDLAPT
DYHLFLSLSDYMRDKQFDDEEHLKTELSTFPSSRSPDFFSRGIMMLPSKWQQVVDTNG
EYLCE"

gene 26381. .28160
/gene="R07H5.9"

CDS join(26381. .26416,26466. .26513,26554. .26636,26684. .26769,
26815. .26863,26908. .26965,27342. .27403,27449. .27629,
27691. .27758,28070. .28160)
/gene="R07H5.9"
/codon_start=1
/db_xref="PID:e275845"
/db_xref="PID:g1628088"
/db_xref="SPTREMBL:Q93935"
/translation="MNAAVRCNPKVQQDQVNQNANNLMVLEDLLPKCLEMIGDKAKSS
ELHVAFDVPTNQTTFQYSIVWEDSSPDADEKYRVSCTEANKIMLALKEKNIKKTSKLM
GYKITNEKSRTNSLREKASETLNNGLPPPPDLRSRSQQCEVSESRSSSVTASSSAPTS
TTCYDSEPTKSDTSKKSSKKSRKSSRSSKVPHRKPKNVGSLDENGCVNVKVVVIERPI
IRDKLSRGSVVLVLDLIGFVEKGKKIPIQIDMDEE"

gene complement(28380. .29654)
/gene="R07H5.8"

CDS complement(join(28380. .28583,28636. .28962,29023. .29128,
29190. .29506,29580. .29654))
/gene="R07H5.8"
/note="Similarity to Human adenosine kinase (SW:P55263); cDNA EST EMBL:Z14851 comes from this gene; cDNA EST EMBL:M89406 comes from this gene; cDNA EST EMBL:M89090 comes from this gene; cDNA EST EMBL:M89195 comes from this gene; cDNA EST EMBL:Z14338 comes from this gene; cDNA EST EMBL:Z14670 comes from this gene; cDNA EST EMBL:T00169 comes from this gene. cDNA EST EMBL:D71377 comes from this gene; cDNA EST EMBL:D74057 comes from this gene; cDNA EST

yk404c6.3 comes from this gene; cDNA EST yk404c6.5 comes from this gene; cDNA EST yk336h7.3 comes from this gene; cDNA EST yk336h7.5 comes from this gene; cDNA EST yk237d7.5 comes from this gene; cDNA EST yk333h8.5 comes from this gene; cDNA EST yk341d5.5 comes from this gene; cDNA EST yk342e12.5 comes from this gene; cDNA EST yk402f9.5 comes from this gene; cDNA EST yk417b2.5 comes from this gene; cDNA EST yk470c4.5 comes from this gene; cDNA EST yk483g11.5 comes from this gene"
/codon_start=1
/db_xref="PID:e275842"
/db_xref="PID:g1628083"
/db_xref="SPTREMBL:Q93934"
/translation="MSPLPENTLIGMCNPLLDIQTTVEKAFLDKWGLKENDAILCDDK
...

Note: remainder of annotations omitted.

Query Match 6.4%; Score 19; DB 18; Length 33802;
Best Local Similarity 100.0%; Pred. No. 5.52e+00;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 4206 AAAAATATTTTTATTTAAA 4224
||||||||||||||||||||
Qy 104 AAAAATATTTTTATTTAAA 122

RESULT 20
LOCUS AC003987 37000 bp DNA PRI 14-JAN-1998
DEFINITION Human PAC clone DJ0866N18 from 7q31, complete sequence.
ACCESSION AC003987
NID g2772541
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 37000)
AUTHORS Bauer,C, Langston,Y and Gillam,B.
TITLE The sequence of H. sapiens PAC clone DJ0866N18
JOURNAL Unpublished (1998)
REFERENCE 2 (bases 1 to 37000)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (14-JAN-1998) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT SUBMITTED BY:
Genome Sequencing Center
Department of Genetics
Washington University
St. Louis MO 63108, USA
http://genome.wustl.edu/gsc

```
            mailto:sapiens@watson.wustl.edu

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded or sequenced with an alternate
            chemistry; an attempt was made to resolve all sequencing problems,
            such as compressions and repeats; all regions were covered by
            sequence from more than one subclone; and the assembly was
            confirmed by restriction digest.

            MAPPING INFORMATION:
            The sequence of this clone was established as part of a mapping and
            sequencing collaboration between the NHGRI Chromosome 7 Mapping
            Project (Eric D. Green, Director), John D. McPherson in the
            Department of Genetics (Washington University), and the Washington
            University Genome Sequencing Center.  For additional information
            about the map position of this sequence, see
            http://www.nhgri.nih.gov/DIR/GTB/CHR7 or send
            mailto:egreen@nhgri.nih.gov

            SOURCE INFORMATION:
            This clone was derived from human PAC library RPCI-4, prepared by
            Pieter de Jong and coworkers at Roswell Park Cancer Institute,
            using the method described by Ioannou et al., Nature Genetics
            6:84-9 (1994).  The library is from one male donor.  For further
            details, see http://bacpac.med.buffalo.edu/
            The clone is available from Genome Systems, Inc.
            (http://www.genomesystems.com).
            VECTOR:  pCYPAC2
            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is RG114A06, 200 base overlap. The
            actual start of this clone is at base position 124318 of RG114A06;
            actual end is at 37000 of DJ0866N18.

            This clone contains STS sWSS377 (NID:g388836).
FEATURES             Location/Qualifiers
     source          1. .37000
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /clone="DJ0866N18"
                     /clone_lib="RPCI-4"
                     /map="7q31"
     repeat_region   210. .360
                     /rpt_family="MER1_type"
     misc_feature    5160. .5412
                     /note="match to EST  (NID:g504646)"
```



```
     repeat_region   6951. .7224
                     /rpt_family="Alu"
     repeat_region   7483. .7720
                     /rpt_family="MaLR"
     repeat_region   7725. .8025
                     /rpt_family="Alu"
     repeat_region   8071. .8182
                     /rpt_family="MaLR"
     repeat_region   8855. .9254
                     /rpt_family="MaLR"
     repeat_region   9315. .9610
                     /rpt_family="Alu"
     repeat_region   9757. .12090
                     /rpt_family="L1"
     repeat_region   12252. .12673
                     /rpt_family="L1"
     repeat_region   14298. .14344
                     /rpt_family="L2"
     repeat_region   14834. .14868
                     /rpt_family="Retroviral"
     repeat_region   14878. .14923
                     /rpt_family="(CA)n"
     repeat_region   14998. .15265
                     /rpt_family="Retroviral"
     misc_feature    15939. .16365
                     /note="match to EST W86345 (NID:g1398837) zh54c09.r1"
     misc_feature    complement(16263. .16674)
                     /note="match to EST W86346 (NID:g1398838) zh54c09.s1"
     misc_feature    complement(18336. .18383)
                     /note="match to EST W86345 (NID:g1398837) zh54c09.r1"
     repeat_region   19327. .19370
                     /rpt_family="L1"
     repeat_region   19371. .19679
                     /rpt_family="Alu"
     repeat_region   19680. .21050
                     /rpt_family="L1"
     repeat_region   21055. .21408
                     /rpt_family="MaLR"
     repeat_region   21409. .21750
                     /rpt_family="L1"
     repeat_region   21858. .22159
                     /rpt_family="Alu"
     repeat_region   22191. .22410
                     /rpt_family="MER1_type"
     repeat_region   22664. .23061
                     /rpt_family="MaLR"
     repeat_region   23222. .23834
                     /rpt_family="L1"
     repeat_region   23818. .24062
                     /rpt_family="L1"
     repeat_region   25549. .25810
```



```
                     /rpt_family="Alu"
     repeat_region   25816. .25859
                     /rpt_family="Alu"
     repeat_region   27822. .28153
                     /rpt_family="MER2_type"
     repeat_region   28231. .28509
                     /rpt_family="Alu"
     repeat_region   28616. .28767
                     /rpt_family="MER1_type"
     repeat_region   28786. .28835
                     /rpt_family="L2"
     repeat_region   30060. .30368
                     /rpt_family="Alu"
     repeat_region   31169. .31262
                     /rpt_family="MIR"
     repeat_region   32125. .32234
                     /rpt_family="MER1_type"
     repeat_region   33338. .33918
                     /rpt_family="MER4-group"
     repeat_region   33922. .34244
                     /rpt_family="MER2_type"
     repeat_region   34445. .34511
                     /rpt_family="(CA)n"
BASE COUNT    10092 a   8098 c   7751 g  11059 t
ORIGIN

  Query Match           6.4%;  Score 19;  DB 26;  Length 37000;
  Best Local Similarity 100.0%;  Pred. No. 5.52e+00;
  Matches    19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  29427 ACTTTATTTATAACTTTTA 29445
          ||||||||||||||||||
Qy    125 ACTTTATTTATAACTTTTA 143


RESULT   21
LOCUS       CEC25D7     42361 bp    DNA             INV       21-SEP-1998
DEFINITION  Caenorhabditis elegans cosmid C25D7, complete sequence.
ACCESSION   Z81039
NID         g1627604
KEYWORDS    HTG.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 42361)
  AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
```



```
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
  TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
  JOURNAL   Nature 368 (6466), 32-38 (1994)
  MEDLINE   94150718
REFERENCE   2  (bases 1 to 42361)
  AUTHORS   Ainscough,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-NOV-1996) Nematode Sequencing Project, Sanger Centre,
            Hinxton, Cambridge CB10 1RQ, England and Department of Genetics,
            Washington University, St. Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     Coding sequences below are predicted from computer analysis, using
            predictions from Genefinder (P. Green, U. Washington), and other
            available information.
            For a graphical representation of this sequence and its analysis
            see:-
            http://webace.sanger.ac.uk/cgi-
            bin/display?db=wormace&class=Sequence &object=C25D7
            Current sequence finishing criteria for the C. elegans genome
            sequencing consortium are that all bases are either sequenced
            unambiguously on both strands, or on a single strand with both a
            dye primer and dye terminator reaction, from distinct subclones.
            Exceptions are indicated by an explicit note.
            IMPORTANT:  This sequence is NOT necessarily the entire insert of
            the specified clone.  It may be shorter because we only sequence
            overlapping sections once, or longer because we arrange for a small
            overlap between neighbouring submissions.
            This sequence is the entire insert of clone C25D7. The true left
            end of clone C01G10 is at 35450 in this sequence. The true right
            end of clone F14H8 is at 19646 in this sequence. The start of this
            sequence (1. .104) overlaps with the end of sequence Z81061.
            The end of this sequence (35450. .42361) overlaps with the start of
            sequence Z81030.
FEATURES             Location/Qualifiers
     source          1. .42361
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /chromosome="V"
                     /clone="C25D7"
     gene            complement(3754. .4270)
                     /gene="C25D7.1"
     CDS             complement(join(3754. .4107,4172. .4270))
                     /gene="C25D7.1"
                     /note="predicted using Genefinder; cDNA EST EMBL:C13397
                     comes from this gene; cDNA EST EMBL:C11360 comes from this
```

gene; cDNA EST yk279f9.3 comes from this gene; cDNA EST yk279f9.5 comes from this gene"
/codon_start=1
/db_xref="PID:e276084"
/db_xref="PID:g1627608"
/translation="MEKEKSKTESLKKDEIDEANSESSKVPLAIDPEEAKLPNAGGKS EHMVVNFTSKRMAIKVRCGNALFRVEPTHMIIEPNKCRQLTINRMPGPIQKDKAIVQY LQIENDVQDPKTAFKAADSAGTKIPHLKIKLVAGASGGRQMSREVVDE"

gene complement(5248. .6288)
/gene="C25D7.2"

CDS complement(join(5248. .5472,5534. .5785,5835. .6134, 6205. .6288))
/gene="C25D7.2"
/note="cDNA EST yk364a11.3 comes from this gene; cDNA EST yk364a11.5 comes from this gene"
/codon_start=1
/db_xref="PID:e330171"
/db_xref="PID:g2274870"
/db_xref="SPTREMBL:O17595"
/translation="MISLVSLIATASATSTVLAMGSSKDRQAASATKSSRASKSQKSA KSTRSGKSSKSGHLSRGKPSKSKKSKSKSVRPLAVPGARGAPKSESKKLRRDKDKSSK SERSRRSSKSKKCDKSAKKCDLNKAKNLCPPVSVVSDASIKSNSGEKSEKSKALKLVP NQSNYTPPEGQGTASVKTSTLRASPNKLPPATTGGVQTVSIANNTKSRKAFKVKTSDN LLYRVNPVFGVFVEPGDKLSIDVLRHNGVEKTDHMIVLTSNASSEQNCAKGVPESDQPR ELTVIPLVYN"

gene complement(6810. .15244)
/gene="sdc-3"

CDS complement(join(6810. .7038,7124. .7401,7602. .7823, 7867. .7931,7981. .8082,8135. .8583,8909. .9824,9870. .10214, 10265. .13198,13521. .13661,13839. .14151,14206. .14360, 14408. .14503,15037. .15244))
/gene="sdc-3"
/note="similar to Zinc finger, C2H2 type; cDNA EST EMBL:D33689 comes from this gene; cDNA EST EMBL:D36575 comes from this gene; cDNA EST yk356g7.3 comes from this gene; cDNA EST yk214e7.3 comes from this gene; cDNA EST yk214e7.5 comes from this gene; cDNA EST yk297h11.5 comes from this gene; cDNA EST yk462a11.5 comes from this gene"
/codon_start=1
/db_xref="PID:e330172"
/db_xref="PID:g2274871"
/db_xref="SPTREMBL:O17596"
/translation="MSDEPPIVLPTEEDGPPSNCVSIAPVNDKSIPVELRIRKAIRGK VTFTQELVVAARRLYLDGELERPDLNYHVAADFYDPPMIPVPDHRNWEYNPLYPREWP RLIDRDLVQKIKMSNTYKKNQARRKRRRLLRIFHRKRGGEFLNYLWHSKENDVHYKRY MAAYQFPLPFEKCIRDVVREELKDPETSFSKPPSVVHHSKIRPATPDLLKLAKYLGKD RFCPIVTPFKFRSQFRLRYYKKSPPKRSNILEKWLKFKKVRDQGVFNIRDAAREVKKL LQRKNVYHMTDPQCEQEFWDVYRLQMQWLEEDNIEENITMLRENLSKVTDEEFNRSVY IMRPENRASIHDDVTFTTDAEINLKPIVLELFRNNMPLEDIRQVKECIDDMINLTVCQ ELIPQVATIETVRDYSGVEQHRLKYIRPENELEIMPQLATVLVKNENTHDQEITSPMF ALLRTEQYLWSIPLRPFLRMYTDWRKTKLNFKKFRKFSPELVSKAPIDLFKVFLTTNM

DKKGVATKDTDCFDYRTASEIKKLYHEEFIRVRPIIEHQRFHRCYTSVITEEPLPNPF YWVHRGQWEKVYECNWEQRVKPLAGRFSWKTKHPNICDLPRLEKDDKQEMEMLAKMRE KIREYNKEKEKHPFYVSSSLPEFASTSDFVKFPISNTGIKYTAPRRNQPKVQRSLVTY TRFRLAELEDKDFSFPPEEKMLEKNRLLEKLNFEERLQLDALMSKDVKYKPRMIPEKGF LQLQDEKLILKRLCEKEDEWSGKQKRRFPDLYEQATQEAERLGDPDASTSGTSNKELG LLPVWKHDDFRDLKKINVLGSKEYPISSPYLIKRLHYDIHLEIVRDKLKSWIYRTELW EKMKPVYERLKRDKEKREKEWHEMRAAQRAQKEKEEEDQRRMDMGIEHGQESLLKSIT PEVEQQILLPISRENSFEEMLPAPQEATYDEDLDQDIDIHFDGAESVCTGIRLDSEDRE LVDSPDQVENEPERVEAVTENIRQEMSDAEMRDLMRKKREQYSMLFPNVPQPFQPLN RPDVTAPGAGIVPPYDEAPDFDLSAYLEEDEVRIPTLSYVEINSLLKDQDLVEEQEIV EEEETPLVVEEPDIPAAPVLSAEEKASRISREMSALRKLPDAAFNYASDRDGVVVEEV ETIEEIQEIYNDKLEKMIAARLGSIAQGANIERLSEDQLLGDQGSEDISFEEIQVDLL LESGVEVQVNQSVTISRSSTSFESLLVEDPEEHPEQLPVSASEKANNQIVPEVEVEGS VVPVTNQQEENVTSEGPTLQEGSSIPSSSHITTVDELLGTESPGPEATETPVAEESPK KKSGKTTRGRPKKVKENLKKRIQPRRGQKEEAAHEPEVVEEQEQVEPEVGPEVVHEPV PAPAAQLETEPIEQQIEEPDKVFEPIIEALPLFETSPVPAPEGNIPSRAHSSDDDVQV ISSETDPNGPINLVEQVQNDKLTAYQYSTEELLGETGELDEAGAPASPSEIVVHDEVLQ DEVLQPNPKSSKKRGRRRKKTPPHIAKARKVFTSISKTEEIELAPTPTQQSRKRMANV SSEEATATRRQKRAKVEEPNDSDVSRVLTPEPEDLHETERPGHVGEEGFETPSLRTGR ESTASSVKTSRSKRLPLSKNNPVPRMRIQSQAGTNASPTPARRTGVADSVSPDGASE KLKQLPKSVQDIFEVFDVERSAERGESATVTNLEGPVKIEIEDENWWAPPAITETSRK SKKRLRAEKQQQILDDIKLELNGEPPQKRECEILMERVQVKIEEGVENADCRIVQANF NCKSVEEHETLPLKIIRVPEENEKPVHKFFMTQIIWKEINEAFMTDPPEKFMLLVRILP SDRNISGQQIYKISMTVTDRIKGFDDDFIKLLTQFPKKLSQEHKPLVDYNKLAHALREK ASLHLNNHKITPFTPHGVISNITEMKEKIIGRCEIGMLTDGDRDVLNSTENMLLGAYM KSVLRMTAQSQVSWAHPEYIKRRLEMIYYGWRMPLGSGGFFRVALAINRKDVKPMSQQ FRDFFIETLKDVNENYAKAVELVQMDEDKLMEAMIEKAGLSSIDLLALDEDVQGQVSE SHKHKCVQCSIRNQSVYFSSYSLLELHGKLHQNLHELAPEDADDCQDCYETLTSSFEV IVHRINHHHSRNCPFADD"

gene complement(18527. .20451)
/gene="C25D7.4"

CDS complement(join(18527. .18736,18779. .19421,19772. .19855, 20411. .20451))
/gene="C25D7.4"
/codon_start=1
/db_xref="PID:e330173"
/db_xref="PID:g2274872"
/db_xref="SPTREMBL:O17597"
/translation="MDYENGVGEGGEQWSKTASPSPSELPIELVGEVLNKLDPIDRLI AWKVSSKFQKILMVMEPGFKEIKVDFIERNVHLILDGNVLQYKPTKTGCIVKRNNHFS SLKNHDYIETAIRDLRLIFKRPKLELHLFSVIIEILYKKEERDRLYMFLTELMRRVEP INTKTISFTRMHCEEMKSILSLLNANNIEKIELNFHHGILLENYLAQLEQWKNAKELC ASAVFAGFYTSIENLFNFRKFEINLETFTVADAVKIKNILLNSANFNYCKIEYSMQFV AHRFEVARLFQPSNTNGNLDYFAYKCSGGNYIIFLDDYHLEIQRENDES"

gene 21896. .24171
/gene="C25D7.5"

CDS join(21896. .22064,22107. .22210,22274. .22436,22482. .22648, 23331. .23540,23854. .23919,23965. .24171)
/gene="C25D7.5"
/codon_start=1
/db_xref="PID:e330174"

/db_xref="PID:g2274873"
/db_xref="SPTREMBL:O17598"
/translation="MSPCQPDGSQDYVCCGTPELLVRLVGNQMKQSYPENWNFYPRQS HYDYSVIIRSSKEQLESDAAFKNQVILTTPKTTTTTVTTTRKRQPVKKSNVVLQMLTI PSLSRGNYSNSLKLHTNSVLVRDGACVPVVDTGLPSQKKQISKNLASYGAPAKKIKFV VVTSSQPQFSGNLNLFPFSQFIMADATMYKDNIVFMKRFEKHSILELCSPNSLIQSTP GPTPNSITVIIRNVDLMGTIAIAGALFPNGNDLNVFDRNSVYDIDKLIESRNQALKGI KTRNIRETNVVLFKKNEEIHVRIELKPMEFPDTNRCKAVQANTIPLVGDHRLGCTGIQ YRQEKAAKRGKARMSTGVQTVSTKNQ"

gene complement(24413. .27047)
/gene="C25D7.6"

CDS complement(join(24413. .24531,24579. .25231,25285. .26649, 26701. .26793,26839. .27047))
/gene="C25D7.6"
/note="predicted using Genefinder; Similarity to Xenopus DNA replication licensing factor MCM3 (SW:P49739); cDNA EST EMBL:M75844 comes from this gene; cDNA EST EMBL:M89318 comes from this gene; cDNA EST EMBL:D27738 comes from this gene; cDNA EST EMBL:D27737 comes from this gene; cDNA EST EMBL:T00522 comes from this gene; cDNA EST EMBL:T01528 comes from this gene; cDNA EST EMBL:T02369 comes from this gene; cDNA EST EMBL:D33517 comes from this gene; cDNA EST EMBL:D34462 comes from this gene; cDNA EST EMBL:D37530 comes from this gene; cDNA EST EMBL:D72865 comes from this gene; cDNA EST EMBL:D75815 comes from this gene; cDNA EST EMBL:D64342 comes from this gene; cDNA EST EMBL:D67309 comes from this gene; cDNA EST EMBL:C12435 comes from this gene; cDNA EST EMBL:C10642 comes from this gene; cDNA EST EMBL:C10909 comes from this gene; cDNA EST yk503f9.3 comes from this gene; cDNA EST yk495g11.3 comes from this gene; cDNA EST yk484c7.3 comes from this gene; cDNA EST yk484c7.5 comes from this gene; cDNA EST yk477c3.3 comes

...

Note: remainder of annotations omitted.

Query Match 6.4%; Score 19; DB 18; Length 42361;
Best Local Similarity 100.0%; Pred. No. 5.52e+00;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 32196 ATATTTTTATTTAAATAAC 32214
||||||||||||||||||||
Qy 108 ATATTTTTATTTAAATAAC 126

RESULT 22
LOCUS AC002516 48158 bp DNA INV 28-AUG-1997
DEFINITION Drosophila melanogaster (P1 DS01068 (D37)) DNA sequence, complete sequence.
ACCESSION AC002516 L77001 L77002 L49202 L77003 AC000601 L79870 L79871 L79872 L77004 L79873 L79874 L49203 L48471
NID g2347072

KEYWORDS HTG.
SOURCE Drosophila melanogaster (Subclones in pOT2 from P1 clone DS01068 (D37)) DNA.
ORGANISM Drosophila melanogaster
Eukaryotae; mitochondrial eukaryotes; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 48158)
AUTHORS Martin,C.H., Alcivare,D.A., Arcaina,T.T., Bondoc,M.M., Chiang,A., Critz,P.A., Davis,C.A., Doyle,C.M., Ericsson,C.L., Farfan,D.E., Gunning,K.M., Houston,K.A., Hummasti,S.R., Jaklevic,M.A., Kadner,K.E., Karra,K., Kim,K., Kim,S.F., Ko,C.L., Lewis,K.D., Li,M., Lindquist,K.J., Lomotan,M.A., Lustre,V.M., Machrus,M.U., Mayeda,C.A., Mazda,P., Miguel,T.M., Miller,C.A., Mok,M.S., Pacleb,J.M., Patel,S.G., Salva,A., Santos,R.F., Snir,E., Subramanian,S., Wan,K.H., Whitelaw,K.R., Yee,A., Yeh,R.T., Yu,C., Zieran,L.L. and Palazzolo,M.J.
TITLE Sequencing of Alcohol dehydrogenase region; Drosophila chromosome 2L, region 34F2-5
JOURNAL Unpublished (1997)
REFERENCE 2 (bases 1 to 48158)
AUTHORS Martin,C.H., Alcivare,D.A., Arcaina,T.T., Bondoc,M.M., Chiang,A., Critz,P.A., Davis,C.A., Doyle,C.M., Ericsson,C.L., Farfan,D.E., Gunning,K.M., Houston,K.A., Hummasti,S.R., Jaklevic,M.A., Kadner,K.E., Karra,K., Kim,K., Kim,S.F., Ko,C.L., Lewis,K.D., Li,M., Lindquist,K.J., Lomotan,M.A., Lustre,V.M., Machrus,M.U., Mayeda,C.A., Mazda,P., Miguel,T.M., Miller,C.A., Mok,M.S., Pacleb,J.M., Patel,S.G., Salva,A., Santos,R.F., Snir,E., Subramanian,S., Wan,K.H., Whitelaw,K.R., Yee,A., Yeh,R.T., Yu,C., Zieran,L.L. and Palazzolo,M.J.
TITLE Direct Submission
JOURNAL Submitted (28-AUG-1997) Berkeley Drosophila Genome Project, MS 74-157, Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley,.CA 94720, US
COMMENT Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 74-157
Berkeley, CA 94720
For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (http://www-hgc.lbl.gov/sequence-archive.html) or send email to drosophila@genome.lbl.gov.
Library location: 12-12.
This P1 was assembled from the following subclones: 3_g5 (L77001), 3_e6 (L77002), 2_b7 (L49202), 9_a7 (L77003), 3_f7 (AC000601), 3_f4, 7_b2, 7_f10 (L79870), 3_e1 (L79871), 8_e4 (L79872), 7_h2 (L77004), 7_c8 (L79873), 5_a4 (L79874), 4_e2 (L49203), 3_c6 (L48471).
FEATURES Location/Qualifiers
source 1. .48158
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"

```
                     /chromosome="2"
                     /map="35A1-35A4"
                     /clone="Pl DS01068 (D37)"
  BASE COUNT    13294 a  10754 c  10782 g  13328 t
  ORIGIN

    Query Match           6.4%;  Score 19;  DB 18;  Length 48158;
    Best Local Similarity 100.0%;  Pred. No. 5.52e+00;
    Matches    19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

  Db  40528 TATTTTTATTTAAATAACT 40546
            ||||||||||||||||||
  Qy    109 TATTTTTATTTAAATAACT 127


  RESULT   23
  LOCUS       AC004606    99512 bp    DNA             PRI       01-MAY-1998
  DEFINITION  Homo sapiens chromosome 17, clone hRPC.1043_H_15, complete
              sequence.
  ACCESSION   AC004606
  NID         g3097870
  KEYWORDS    HTG.
  SOURCE      human.
    ORGANISM  Homo sapiens
              Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
              Primates; Catarrhini; Hominidae; Homo.
  REFERENCE   1  (bases 1 to 99512)
    AUTHORS   Birren,B., Fasman,K., Linton,L., Nusbaum,C. and Lander,E.
    TITLE     Homo sapiens chromosome 17, clone hRPC.1043_H_15
    JOURNAL   Unpublished
  REFERENCE   2  (bases 1 to 99512)
    AUTHORS   Birren,B., Fasman,K., McKernan,K., Nusbaum,C., Richardson,P.,
              Lander,E., Allen,N., Baker,J., Baldwin,J., Barna,N., Beckerly,R.,
              Benn,J., Boatin,C., Boutwell,C., Brown,A., Byrne,S., Cantu,C.,
              Castle,A., Cerny,J., Cooke,P., Daly,M.J., Depayre,E., Devon,K.,
              Dewar,K., Donelan,L., DuRette,B., Etemadi,S., Ferreira,P.,
              FitzHugh,W., Forrest,C., Funke,R., Gage,D., Gardyna,S.,
              Gensheimer,S., Geraigery,K., Gilmartin,T., Grant,G., Gray,D.,
              Hagos,B., Harris,K., Horton,L., Howland,J.C., Hui,L., Jacotot,L.,
              Kann,L., Linton,L., Macdonald,P., Marquis,N., McEwan,P., McGurk,A.,
              Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
              Nachman,A., Nahf,R., Naylor,J., O'Connor,T., Pavlin,B.,
              Peterson,K., Riley,R., Roberts,D., Rollins,G., Rossello,R., Roy,A.,
              Shyam,R., Stange-Thomann,N., Stilwell,J., Stone,C., Strickland,C.,
              Subramanian,A., Sydney,K., Tang,L., Torruella-Miller,I.,
              Vassiliev,H., Vo,A., Wagner,A., Wang,B., Wheeler,J., Wu,Y.,
              Ye,W.J., Zhao,J. and Zody,M.
    TITLE     Direct Submission
    JOURNAL   Submitted (23-APR-1998) Whitehead Institute/MIT Center for Genome
              Research, 320 Charles Street, Cambridge, MA 02141, USA
  REFERENCE   3  (bases 1 to 99512)
```



```
    AUTHORS   Birren,B., Fasman,K., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
              Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boatin,C.,
              Boutwell,C., Brown,A., Byrne,S., Cantu,C., Castle,A., Cerny,J.,
              Cooke,P., Depayre,E., Devon,K., Dewar,K., Donelan,L., DuRette,B.,
              Etemadi,S., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R.,
              Gage,D., Gardyna,S., Gensheimer,S., Geraigery,K., Gilmartin,T.,
              Grant,G., Hagos,B., Harris,K., Horton,L., Howland,J.C., Hui,L.,
              Jacotot,L., Kann,L., Macdonald,P., Marquis,N., McEwan,P.,
              McGurk,A., McKernan,K., Meldrim,J., Molla,M., Morris,W., Morrow,J.,
              Mychaleckyj,J., Nachman,A., Nahf,R., Naylor,J., O'Connor,T.,
              Pavlin,B., Peterson,K., Riley,R., Roberts,D., Rossello,R., Roy,A.,
              Shyam,R., Stange-Thomann,N., Stilwell,J., Stone,C., Strickland,C.,
              Subramanian,A., Torruella-Miller,I., Vassiliev,H., Vo,A.,
              Wagner,A., Wang,B., Wheeler,J., Wu,Y., Ye,W.J., Zhao,J. and Zody,M.
    TITLE     Direct Submission
    JOURNAL   Submitted (01-MAY-1998) Whitehead Institute/MIT Center for Genome
              Research, 320 Charles Street, Cambridge, MA 02141, USA
  COMMENT     On May 1, 1998 this sequence version replaced gi:3095003.
              All repeats were identified using RepeatMasker: Smit, A.F.A. &
              Green, P. (1996-1997)
              http://ftp.genome.washington.edu/RM/RepeatMasker.html.
  FEATURES             Location/Qualifiers
       source          1. .99512
                       /organism="Homo sapiens"
                       /db_xref="taxon:9606"
                       /clone="hRPC.1043_H_15"
                       /clone_lib="Peter de Jong/ human PAC library"
                       /map="17"
                       /chromosome="17"
       repeat_region   complement(1. .146)
                       /rpt_family="L1MC/D"
       repeat_region   complement(172. .1051)
                       /rpt_family="L1MC/D"
       repeat_region   1409. .2034
                       /rpt_family="L2"
       repeat_region   4624. .4674
                       /rpt_family="AT_rich"
       repeat_region   4763. .4814
                       /rpt_family="AT_rich"
       repeat_region   5921. .5959
                       /rpt_family="AT_rich"
       repeat_region   complement(6332. .7051)
                       /rpt_family="L1MA7"
       repeat_region   complement(7069. .7372)
                       /rpt_family="AluSc"
       repeat_region   complement(7400. .7689)
                       /rpt_family="L1MA7"
       repeat_region   7689. .8107
                       /rpt_family="L1"
       repeat_region   8108. .8301
                       /rpt_family="THE1C"
```



```
       repeat_region   8304. .8559
                       /rpt_family="L1PA2"
       repeat_region   complement(8572. .10975)
                       /rpt_family="L1MA4A"
       repeat_region   complement(11007. .11074)
                       /rpt_family="(TA)n"
       repeat_region   complement(11076. .11162)
                       /rpt_family="L1MA4A"
       repeat_region   complement(11172. .11276)
                       /rpt_family="AluSg/x"
       repeat_region   11282. .11677
                       /rpt_family="THE1C"
       repeat_region   complement(11678. .11766)
                       /rpt_family="AluY"
       repeat_region   complement(11797. .11843)
                       /rpt_family="(CA)n"
       repeat_region   complement(11845. .12028)
                       /rpt_family="L1MA4A"
       repeat_region   12027. .12729
                       /rpt_family="L1"
       repeat_region   12715. .13454
                       /rpt_family="L1MA8"
       repeat_region   13452. .13666
                       /rpt_family="L1"
       repeat_region   13690. .14077
                       /rpt_family="MSTA"
       repeat_region   complement(14086. .14412)
                       /rpt_family="L1ME1"
       repeat_region   complement(14391. .14636)
                       /rpt_family="L1MC/D"
       repeat_region   complement(14632. .14760)
                       /rpt_family="L1M4"
       repeat_region   15025. .15291
                       /rpt_family="L1PA8"
       repeat_region   15292. .15334
                       /rpt_family="AT_rich"
       repeat_region   15477. .15568
                       /rpt_family="L1M4"
       repeat_region   16441. .17498
                       /rpt_family="L1ME3A"
       repeat_region   complement(18781. .18916)
                       /rpt_family="L1PA12"
       repeat_region   18907. .19237
                       /rpt_family="L1PA12"
       repeat_region   complement(19315. .19496)
                       /rpt_family="MIR"
       repeat_region   21444. .21504
                       /rpt_family="L2"
       repeat_region   complement(21505. .21870)
                       /rpt_family="MLT1A1"
       repeat_region   21942. .22030
```



```
                       /rpt_family="MIR"
       repeat_region   complement(22031. .22121)
                       /rpt_family="L1PA6"
       repeat_region   22119. .23393
                       /rpt_family="L1PA6"
       repeat_region   complement(23440. .23801)
                       /rpt_family="L1PB1"
       repeat_region   complement(23808. .23868)
                       /rpt_family="(TA)n"
       repeat_region   complement(23882. .24177)
                       /rpt_family="AluSp"
       repeat_region   complement(24187. .24823)
                       /rpt_family="L1PB1"
       repeat_region   complement(24826. .25342)
                       /rpt_family="L1"
       repeat_region   25355. .25488
                       /rpt_family="MIR"
       repeat_region   25948. .26024
                       /rpt_family="MLT1D"
       repeat_region   complement(27114. .27413)
                       /rpt_family="AluSc"
       repeat_region   complement(27414. .27606)
                       /rpt_family="MLT2A"
       repeat_region   complement(27905. .27992)
                       /rpt_family="L1ME1"
       repeat_region   28057. .28462
                       /rpt_family="MSTA"
       repeat_region   complement(28466. .28816)
                       /rpt_family="MLT1A2"
       repeat_region   complement(28827. .29361)
                       /rpt_family="L1ME1"
       repeat_region   complement(29387. .29618)
                       /rpt_family="L1MB8"
       repeat_region   30005. .30032
                       /rpt_family="(GA)n"
       repeat_region   30643. .30889
                       /rpt_family="L1M4"
       repeat_region   30890. .31792
                       /rpt_family="L1MA3"
       repeat_region   31804. .31870
                       /rpt_family="AT_rich"
       repeat_region   31917. .31999
                       /rpt_family="MADE1"
       repeat_region   32023. .32853
                       /rpt_family="L1M4"
       repeat_region   32858. .33192
                       /rpt_family="AluJb"
       repeat_region   33416. .33967
                       /rpt_family="L1M4"
       repeat_region   complement(33984. .34248)
                       /rpt_family="L1PA16"
```

```
     repeat_region   34249. .34839
                     /rpt_family="L1PA16"
     repeat_region   34823. .35126
                     /rpt_family="L1M4"
     repeat_region   35132. .35378
                     /rpt_family="L1M4"
     repeat_region   35366. .35500
                     /rpt_family="L1M4"
     repeat_region   complement(35500. .35932)
                     /rpt_family="L1P4"
     repeat_region   36641. .36761
                     /rpt_family="LTR16C"
...
Note: remainder of annotations omitted.

  Query Match           6.4%;  Score 19;  DB 26;  Length 99512;
  Best Local Similarity 100.0%;  Pred. No. 5.52e+00;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  97697 AGTTATTTAAATAAAAATA 97715
          |||||||||||||||||||
Cp    127 AGTTATTTAAATAAAAATA 109


RESULT   24
LOCUS       HS422H11   112984 bp    DNA              PRI       16-MAY-1998
DEFINITION  Homo sapiens DNA sequence from PAC 422H11 on chromosome
            6p21.1-22.2. Contains the gene coding for two isoforms of a known
            Serine Kinase. Contains ESTs, STSs and a GSS, complete sequence.
ACCESSION   Z99128
NID         g3135974
KEYWORDS    HTG; Serine Kinase.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 112984)
  AUTHORS   Phillips,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAY-1998) sanger.ac.uk/HGP/Chr6/) Sanger Centre,
            Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquires:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On May 18, 1998 this sequence version replaced gi:2578108.
            IMPORTANT: This sequence is the entire insert of clone 422H11.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variations annotated may not be found in the sequence submission
            corresponding to the overlapping clone as we submit sequences with
            only a small overlap as described above.
            This sequence was generated from part of bacterial clone contigs of
```



```
            human chromosome 6, constructed in collaboration by the Sanger
            Centre chromosome 6 mapping group and Ioannis Ragoussis. Further
            information can be found at http://www.sanger.ac.uk/HGP/Chr6/ This
            sequence has been finished according to sequence map criteria as
            follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            The true left end of clone 422H11 is at 1 in this sequence. The
            true right end of clone 422H11 is at 112984.
            422H11 is from the library RPCI3 constructed at the Roswell Park
            Cancer Institute by the group of Pieter de Jong.
            For further details see http://bacpac.med.buffalo.edu/.
FEATURES            Location/Qualifiers
     source         1. .112984
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="6"
                    /map="p21.1-22.2"
                    /clone="422H11"
                    /clone_lib="RPCI3"
     repeat_region  1. .236
                    /note="AluJb repeat: matches 59. .294 of consensus;
                    incomplete repeat"
     repeat_region  332. .469
                    /note="AluSg repeat: matches 163. .300 of consensus;
                    incomplete repeat"
     repeat_region  717. .854
                    /note="MLT1F repeat: matches 534. .389 of consensus"
     repeat_region  966. .1856
                    /note="L1PA2 repeat: matches 891. .1 of consensus"
     repeat_region  1707. .2550
                    /note="L1 repeat: matches 5390. .4549 of consensus"
     repeat_region  2545. .2805
                    /note="MLT1F repeat: matches 283. .11 of consensus"
     repeat_region  2851. .3153
                    /note="AluJb repeat: matches 301. .2 of consensus"
     repeat_region  3701. .4002
                    /note="AluSx repeat: matches 302. .1 of consensus"
     repeat_region  4138. .4460
                    /note="AluSp repeat: matches 296. .1 of consensus"
     repeat_region  4708. .4997
                    /note="AluJo repeat: matches 293. .1 of consensus"
     repeat_region  5875. .6162
                    /note="AluSc repeat: matches 297. .3 of consensus"
     mRNA           complement(join(6642. .9096,11945. .12037,16143. .16212,
                    30869. .30976,42621. .42718,43090. .43509,43889. .44102,
                    44542. .44567,46171. .46336,47841. .47947,60332. .60419,
                    61632. .61719,62423. .62531,64503. .64621,94076. .>94136))
                    /gene="dJ422H11.1"
```



```
                    /note="match: cDNAs U09564 AB012290 AJ224115 AB006036
                    U88666; match: ESTs H52845 H84273 D59613 D53745 D53987
                    N55433 T19220 R97577 AA868625 AA830216 C17208 R27665
                    N59394 AA665432 AA379216 AA664660 D59631 H21014 H52846
                    D79129 D51765 AA526955 T89751 N71643 AA053254 AA630604
                    AA568359 AA053643 AA393213 W27036 N76806 R62548 AA379797
                    H20972 D51804 D19849 D52125 D59461 AA416882 T77446 N76402
                    AA311740 AA490202 AA323127 W01670 F07122 AA899108 AA629288
                    AA309978 AA383402 R21674 AA053643 N77083 AA284106 R78141
                    R78142 R29335 T77232 AA921804 D56469 F05956 AA377246
                    AA331949 T77232 AA309978 AA888014 AA826550 T77232 AA888014
                    AA309978 T77232 AA309978AA332381 AA333218 AA291188
                    AA332381 AA333218 AA888014 AA291188 F05125 H15206 Z45803
                    AA332381 AA333218 AA291188 F05124 AA619632 AA535594
                    AA535883 AA535236 AA382461 AA382461 AA060080 AA211241
                    AA492926"
                    /evidence=not_experimental
                    /product="dJ422H11.1.1 (Serine Kinase) (isoform 1)"
     polyA_site     complement(6642)
                    /gene="dJ422H11.1"
     gene           complement(6642. .94136)
                    /gene="dJ422H11.1"
     misc_feature   complement(6642. .8911)
                    /gene="dJ422H11.1"
                    /note="match: STSs G06774 G06295"
     mRNA           complement(join(6642. .9096,12314. .12406,16143. .16212,
                    30869. .30976,42621. .42718,43090. .43509,43889. .44102,
                    44542. .44567,46171. .46336,47841. .47947,60332. .60419,
                    61632. .61719,62423. .62531,64503. .64621,94076. .>94136))
                    /gene="dJ422H11.1"
                    /note="match: cDNAs U88666 AB006036"
                    /evidence=not_experimental
                    /product="dJ422H11.1.2 (Serine Kinase) (isoform 2)"
     polyA_signal   complement(6661. .6666)
                    /gene="dJ422H11.1"
     CDS            complement(join(8912. .9096,11945. .12037,16143. .16212,
                    30869. .30976,42621. .42718,43090. .43509,43889. .44102,
                    44542. .44567,46171. .46336,47841. .47947,60332. .60419,
                    61632. .61719,62423. .62531,64503. .64621,94076. .>94136))
                    /gene="dJ422H11.1"
                    /note="match: proteins Q12890 P36616 Q03656 CE02949
                    CE05897"
                    /codon_start=3
                    /evidence=not_experimental
                    /product="dJ422H11.1.1 (Serine Kinase) (isoform 1)"
                    /db_xref="PID:e1293208"
                    /db_xref="PID:g3135975"
                    /translation="LALQARKKRTKAKKDKAQRKSETQHRGSAPHSESDLPEQEEEIL
                    GSDDDEQEDPNDYCKGGYHLVKIGDLFNGRYHVIRKLGWGHFSTVWLSWDIQGKKFVA
                    MKVVKSAEHYTETALDEIRLLKSVRNSDPNDPNREMVVQLLDDFKISGVNGTHICMVF
                    EVLGHHLLKWIIKSNYQGLPLPCVKKIIQQVLQGLDYLHTKCRIIHTDIKPENILLSV
```



```
                    NEQYIRRLAAEATEWQRSGAPPPSGSAVSTAPQPKPADKMSKNKKKKLKKKQKRQAEL
                    LEKRMQEIEEMEKESGPGQKRPNKQEESESPVERPLKENPPNKMTQEKLEESSTIGQD
                    QTLMERDTEGGAAEINCNGVIEVINYTQNSNNETLRHKEDLHNANDCDVQNLNQESSF
                    LSSQNGDSSTSQETDSCTPITSEVSDTMVCQSSSTVGQSFSEQHISQLQESIRAEIPC
                    EDEQEQEHNGPLDNKGKSTAGNFLVNPLEPKNAEKLKVKIADLGNACWHKHFTEDIQT
                    RQYRSLEVLIGSGYNTPADIWSTACMAFELATGDYLFEPHSGEETYRDEDHIVHIIEL
                    LGKVPRKLIVAGKYSKEFFTKKGDLKHITKLKPWGLFEVLVEKYEWSQEEAAGFTDFL
                    LPMLELIPEKRATAAECLRHPWLNS"
     CDS            complement(join(8912. .9096,12314. .12406,16143. .16212,
                    30869. .30976,42621. .42718,43090. .43509,43889. .44102,
                    44542. .44567,46171. .46336,47841. .47947,60332. .60419,
                    61632. .61719,62423. .62531,64503. .64621,94076. .>94136))
                    /gene="dJ422H11.1"
                    /note="match: proteins P78362 Q13583 P36616 Q03656"
                    /codon_start=3
                    /evidence=not_experimental
                    /product="dJ422H11.1.2 (Serine Kinase) (isoform 2)"
                    /db_xref="PID:e1293209"
                    /db_xref="PID:g3135976"
                    /translation="LALQARKKRTKAKKDKAQRKSETQHRGSAPHSESDLPEQEEEIL
                    GSDDDEQEDPNDYCKGGYHLVKIGDLFNGRYHVIRKLGWGHFSTVWLSWDIQGKKFVA
                    MKVVKSAEHYTETALDEIRLLKSVRNSDPNDPNREMVVQLLDDFKISGVNGTHICMVF
                    EVLGHHLLKWIIKSNYQGLPLPCVKKIIQQVLQGLDYLHTKCRIIHTDIKPENILLSV
                    NEQYIRRLAAEATEWQRSGAPPPSGSAVSTAPQPKPADKMSKNKKKKLKKKQKRQAEL
                    LEKRMQEIEEMEKESGPGQKRPNKQEESESPVERPLKENPPNKMTQEKLEESSTIGQD
                    QTLMERDTEGGAAEINCNGVIEVINYTQNSNNETLRHKEDLHNANDCDVQNLNQESSF
                    LSSQNGDSSTSQETDSCTPITSEVSDTMVCQSSSTVGQSFSEQHISQLQESIRAEIPC
                    EDEQEQEHNGPLDNKGKSTAGNFLVNPLEPKNAEKLKVKIADLGNACWHKHFTEDIQT
                    RQYRSLEVLIGSGYNTPADIWSTACMAFELATGDYLFEPHSGEETYRDEDHIVHIIEL
                    IGRIPRRFSLSGKYSQDFFSHRGDLKHITKLKPWGLFEVLVEKYEWSQEEAAGFTDFL
                    LPMLELIPEKRATAAECLRHPWLNS"
     repeat_region  9636. .9968
                    /note="MER1B repeat: matches 1. .337 of consensus"
     repeat_region  10079. .10382
                    /note="AluJb repeat: matches 1. .302 of consensus"
     repeat_region  11046. .11162
                    /note="MIR repeat: matches 168. .33 of consensus"
     repeat_region  11170. .11464
                    /note="AluJb repeat: matches 296. .2 of consensus"
     repeat_region  11477. .11774
                    /note="AluSq repeat: matches 300. .2 of consensus"
     repeat_region  12725. .12915
                    /note="AluJo repeat: matches 199. .9 of consensus;
                    incomplete repeat"
     repeat_region  13293. .13585
                    /note="AluSx repeat: matches 296. .1 of consensus"
     repeat_region  13586. .13712
                    /note="AluJb repeat: matches 129. .3 of consensus;
                    incomplete repeat"
     repeat_region  15160. .15458
                    /note="AluSc repeat: matches 1. .299 of consensus"
```

```
     repeat_region   16577. .16875
                     /note="AluY repeat: matches 299. .1 of consensus"
     repeat_region   16929. .16956
                     /note="14 copies of 2 mer 89 % conserved"
     repeat_region   17068. .17356
                     /note="AluSc repeat: matches 288. .1 of consensus"
     repeat_region   17614. .17910
                     /note="AluY repeat: matches 299. .1 of consensus"
     repeat_region   18485. .18555
                     /note="AluJ repeat: matches 118. .183 of consensus;
                     incomplete repeat"
     repeat_region   18554. .19043
                     /note="L1MC1 repeat: matches 534. .1079 of consensus"
     repeat_region   19659. .19750
                     /note="L1MB3 repeat: matches 721. .811 of consensus"
     repeat_region   19751. .19883
                     /note="FLAM_C repeat: matches 1. .133 of consensus"
     repeat_region   19917. .20180
                     /note="AluJb repeat: matches 1. .299 of consensus"
     repeat_region   20228. .20483
                     /note="L1 repeat: matches 3307. .3588 of consensus"
...
Note: remainder of annotations omitted.

  Query Match           6.4%;  Score 19;  DB 26;  Length 112984;
  Best Local Similarity 100.0%;  Pred. No. 5.52e+00;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  65812 TAATGCATTAAAAATATTT 65830
          ||||||||||||||||||| 
Qy     95 TAATGCATTAAAAATATTT 113


RESULT   25
LOCUS       HS516C23   118021 bp    DNA             PRI       02-JUN-1998
DEFINITION  Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone
            516C23, complete sequence.
ACCESSION   Z93021
NID         g3183886
KEYWORDS    HTG; HTGS_PHASE3.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 118021)
  AUTHORS   Milne,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1998) Wellcome Trust Genome Campus, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquires:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jun 4, 1998 this sequence version replaced gi:2462415.
```



```
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and the release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc. Order of segments is not known; 800 n's separate
            segments. Unfinished sequence: dJ516C23 Contig_ID: 00970 acc=Z93021
            Length: 118021 bp.
FEATURES             Location/Qualifiers
     source          1. .118021
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="516C23"
BASE COUNT    37057 a   20462 c   21743 g   38759 t
ORIGIN

  Query Match           6.4%;  Score 19;  DB 26;  Length 118021;
  Best Local Similarity 100.0%;  Pred. No. 5.52e+00;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  48299 ATATTTTTATTTAAATAAC 48317
          ||||||||||||||||||| 
Qy    108 ATATTTTTATTTAAATAAC 126


RESULT   26
LOCUS       HSAC002070 165197 bp    DNA             PRI       12-MAY-1997
DEFINITION  Human BAC clone 7E17 from 12q, complete sequence.
ACCESSION   AC002070
NID         g2078475
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 165197)
  AUTHORS   Connell,M.
  TITLE     The sequence of H. sapiens BAC clone 7E17
  JOURNAL   Unpublished (1997)
REFERENCE   2  (bases 1 to 165197)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAY-1997)
COMMENT     SUBMITTED BY:
            Genome Sequencing Center
            Department of Genetics
            Washington University
            St. Louis MO 63108, USA
            http://genome.wustl.edu/gsc
```



```
            mailto:sapiens@watson.wustl.edu

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded or sequenced with an alternate
            chemistry; an attempt was made to resolve all sequencing problems,
            such as compressions and repeats; all regions were covered by
            sequence from more than one subclone; and the assembly was
            confirmed by restriction digest.

            MAPPING INFORMATION:
            This clone was originally isolated in the laboratory of Professor
            Graeme Bell, Howard Hughes Medical Institute and Departments of
            Biochemistry and Molecular Biology, and Medicine, The University of
            Chicago, Chicago, IL, USA.  The clone was provided by the
            laboratory of Dr. Roger Cox at The Wellcome Trust Centre For Human
            Genetics, Oxford, UK.  Some contig information was also obtained
            from Yamagata et al., Nature 384:455-8 (1996).

            SOURCE INFORMATION:
            This clone is from the first release of the human BAC library.  The
            library contains cloned DNA from a human male fibroblast cell line
            978SK.  See:  Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7
            (1992); Kim et al., Genomics 34:213-8 (1996).  The clone is
            available from Research Genetics, Inc. (http://www.resgen.com).
            VECTOR:  pBeloBAC11
            Selection:  chloramphenicol

            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the right is H_127H14, 200 bp overlap.
            Actual start of this clone is at base position 1 of H_7E17; actual
            end is at 73841 of H_127H14.

            This clone contains STS WI-5762 (NID:g721907).
FEATURES             Location/Qualifiers
     source          1. .165197
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /clone="7E17"
                     /clone_lib="CITB-978SK-B"
                     /map="12q"
     repeat_region   1568. .1628
                     /rpt_family="L1"
     repeat_region   complement(1652. .1831)
                     /rpt_family="ALU"
     repeat_region   complement(1856. .2147)
```



```
                     /rpt_family="ALU"
     repeat_region   3229. .3246
                     /rpt_family="L1"
     repeat_region   complement(3523. .3721)
                     /rpt_family="MER"
     repeat_region   4054. .4346
                     /rpt_family="ALU"
     repeat_region   4500. .4791
                     /rpt_family="ALU"
     repeat_region   6356. .6408
                     /rpt_family="THE"
     repeat_region   6477. .6716
                     /rpt_family="THE"
     repeat_region   7019. .7326
                     /rpt_family="ALU"
     repeat_region   7819. .7964
                     /rpt_family="ALU"
     repeat_region   complement(8286. .8586)
                     /rpt_family="ALU"
     repeat_region   complement(11616. .11907)
                     /rpt_family="ALU"
     repeat_region   complement(12488. .13138)
                     /rpt_family="L1"
     repeat_region   12845. .12865
                     /rpt_family="L1"
     repeat_region   13101. .13212
                     /rpt_family="L1"
     repeat_region   complement(13432. .13531)
                     /rpt_family="L1"
     repeat_region   13752. .13868
                     /rpt_family="ALU"
     repeat_region   16917. .16951
                     /rpt_family="L1"
     misc_feature    16952. .16999
                     /note="similar to human EST C14248 (NID:g1568955)"
     misc_feature    16952. .16996
                     /note="similar to human EST R52841 (NID:g814743)
                     yh01a04.r1"
     misc_feature    16952. .16998
                     /note="similar to human EST C14217 (NID:g1568924)"
     repeat_region   17060. .17348
                     /rpt_family="ALU"
     repeat_region   complement(17643. .17931)
                     /rpt_family="ALU"
     repeat_region   complement(20639. .21148)
                     /rpt_family="L1"
     repeat_region   20965. .21148
                     /rpt_family="L1"
     repeat_region   complement(21204. .21286)
                     /rpt_family="L1"
     repeat_region   complement(21334. .21354)
```

```
                     /rpt_family="L1"
     misc_feature    21420. .21466
                     /note="similar to human EST H90627 (NID:g1081057)
                     yv02h02.r1"
     misc_feature    21420. .21466
                     /note="similar to human EST W03790 (NID:g1275787)
                     za61g04.r1"
     repeat_region   23461. .23749
                     /rpt_family="ALU"
     repeat_region   23825. .23851
                     /rpt_family="L1"
     repeat_region   25282. .25504
                     /rpt_family="ALU"
     repeat_region   complement(26333. .26623)
                     /rpt_family="ALU"
     repeat_region   complement(26879. .27142)
                     /rpt_family="ALU"
     repeat_region   complement(28834. .29122)
                     /rpt_family="ALU"
     repeat_region   complement(29595. .29885)
                     /rpt_family="ALU"
     repeat_region   complement(30338. .30363)
                     /rpt_family="L1"
     repeat_region   complement(30364. .30654)
                     /rpt_family="ALU"
     repeat_region   complement(30740. .30777)
                     /rpt_family="L1"
     repeat_region   complement(32850. .32919)
                     /rpt_family="L1"
     repeat_region   33727. .34011
                     /rpt_family="ALU"
     repeat_region   complement(34586. .34635)
                     /rpt_family="L1"
     repeat_region   complement(36160. .36185)
                     /rpt_family="L1"
     repeat_region   36778. .36797
                     /rpt_family="L1"
     repeat_region   complement(37368. .37650)
                     /rpt_family="ALU"
     repeat_region   39600. .39897
                     /rpt_family="ALU"
     repeat_region   41914. .42212
                     /rpt_family="ALU"
     repeat_region   complement(43352. .43374)
                     /rpt_family="L1"
     repeat_region   44601. .44891
                     /rpt_family="ALU"
     repeat_region   complement(45031. .45282)
                     /rpt_family="L1"
     repeat_region   45348. .45647
                     /rpt_family="ALU"
```



```
     repeat_region   46430. .46618
                     /rpt_family="L1"
     repeat_region   complement(46437. .46618)
                     /rpt_family="L1"
     repeat_region   complement(46651. .46768)
                     /rpt_family="ALU"
     repeat_region   46913. .47207
                     /rpt_family="ALU"
     repeat_region   47228. .47516
                     /rpt_family="ALU"
     repeat_region   48350. .48493
                     /rpt_family="ALU"
     repeat_region   52711. .52741
                     /rpt_family="MER"
     repeat_region   52783. .53192
                     /rpt_family="MER"
     repeat_region   complement(53211. .53515)
                     /rpt_family="ALU"
     misc_feature    53677. .53732
                     /note="similar to human EST AA015842 (NID:g1476890)
                     ze30f01.r1"
     misc_feature    53677. .53732
                     /note="similar to human EST AA044938 (NID:g1523141)
...
Note: remainder of annotations omitted.

  Query Match           6.4%;  Score 19;  DB 26;  Length 165197;
  Best Local Similarity 100.0%;  Pred. No. 5.52e+00;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  44582 TTAAAAATATTTTTATTTA 44600
          ||||||||||||||||||| |
Qy    102 TTAAAAATATTTTTATTTA 120


RESULT   27
LOCUS       AC002094    167101 bp    DNA             PRI       09-SEP-1997
DEFINITION  Genomic sequence from Human 17, complete sequence.
ACCESSION   AC002094
NID         g2155224
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 167101)
  AUTHORS   Hawkins,T.L., Birren,B.W., Fasman,K.H., Nussbaum,C. and Lander,E.S.
  TITLE     Genomic sequence from Human 17
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 167101)
  AUTHORS   Lane,M., Hawkins,T.L., Reeve,M.P., Christoffersen,A., Birren,B.W.,
```



```
            Fasman,K.H., Lander,E.S., McKernan,K., Munro,C., Richardson,P.,
            Barna,N., Chang,A., Cooke,P., Daly,M.J., Forrest,C., Fripp,W.J.,
            Gage,D., Geraigery,K., Hagos,B., Jacotot,L., MacKenzie,J.,
            Marquis,N., McDermott,J., Moloney,N., Morrow,J., Nachman,A.,
            Naylor,J., O'Connor,T., Peterson,K., Rollins,G., Spencer,J.,
            Stilwell,J., Stone,C., Strickland,C., Sydney,K., Wilmer,F. and
            Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-1997) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 167101)
  AUTHORS   Lane,M., Hawkins,T.L., Reeve,M.P., Christoffersen,A., Birren,B.W.,
            Fasman,K.H., Lander,E.S., McKernan,K., Munro,C., Richardson,P.,
            Barna,N., Chang,A., Cooke,P., Daly,M.J., Forrest,C., Fripp,W.J.,
            Gage,D., Geraigery,K., Hagos,B., Jacotot,L., MacKenzie,J.,
            Marquis,N., McDermott,J., Moloney,N., Morrow,J., Nachman,A.,
            Naylor,J., O'Connor,T., Peterson,K., Rollins,G., Spencer,J.,
            Stilwell,J., Stone,C., Strickland,C., Sydney,K., Wilmer,F. and
            Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAY-1997) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 167101)
  AUTHORS   Hawkins,T.L., Birren,B.W., Fasman,K.H., Nussbaum,C., Lander,E.S.,
            McKernan,K., Munro,C., Richardson,P., Barna,N., Chang,A., Cooke,P.,
            Daly,M.J., Devon,K., Dewar,K., Forrest,C., Gage,D., Geraigery,K.,
            Hagos,B., Harris,K., Huang,J., Hui,L., Jacotot,L., Kirby,A.,
            Lane,M., MacKenzie,J., Marquis,N., McDermott,J., Molla,M.,
            Morrow,J., Nachman,A., Naylor,J., Nusbaum,C., O'Connor,T.,
            Olotu,A., Peterson,K., Reeve,M.P., Roberts,D., Rollins,G.,
            Stilwell,J., Stone,C., Strickland,C., Sydney,K., Tang,L.,
            Wilmer,F., Zemtseva,I. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-1997) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   5  (bases 1 to 167101)
  AUTHORS   Hawkins,T.L., Birren,B.W., Fasman,K.H., Nussbaum,C., Lander,E.S.,
            McKernan,K., Munro,C., Richardson,P., Barna,N., Chang,A., Cooke,P.,
            Daly,M.J., Devon,K., Dewar,K., Forrest,C., Gage,D., Geraigery,K.,
            Hagos,B., Harris,K., Huang,J., Hui,L., Jacotot,L., Kirby,A.,
            Lane,M., MacKenzie,J., Marquis,N., McDermott,J., Molla,M.,
            Morrow,J., Nachman,A., Naylor,J., Nusbaum,C., O'Connor,T.,
            Olotu,A., Peterson,K., Reeve,M.P., Roberts,D., Rollins,G.,
            Stilwell,J., Stone,C., Strickland,C., Sydney,K., Tang,L.,
            Wilmer,F., Zemtseva,I. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-SEP-1997) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jun 3, 1997 this sequence version replaced gi:2085762.
            The Staden databases, finishing information, and all
            chromatographic files used in the assembly of this clone are
```



```
            available from our anonymous ftp site.

            All repeats were identified using RepeatMasker: Smit, A.F.A. &
            Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html.
FEATURES            Location/Qualifiers
     source         1. .167101
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="96E2"
                    /clone_lib="Research Genetics/Cal Tech CITB978SK-B (plates
                    1-194)"
                    /chromosome="17"
     repeat_region  complement(89. .288)
                    /rpt_family="MIR"
     repeat_region  883. .1179
                    /rpt_family="AluJo"
     repeat_region  3490. .3627
                    /rpt_family="MIR"
     repeat_region  3699. .3823
                    /rpt_family="MER5A"
     repeat_region  complement(5338. .5797)
                    /rpt_family="MER21B"
     repeat_region  5818. .5900
                    /rpt_family="AluJ/FLAM"
     repeat_region  complement(5901. .5927)
                    /rpt_family="AT_rich"
     repeat_region  complement(6404. .6506)
                    /rpt_family="MER21B"
     repeat_region  6721. .7036
                    /rpt_family="LINE2"
     repeat_region  7232. .7351
                    /rpt_family="LINE2"
     repeat_region  complement(7492. .7804)
                    /rpt_family="AluJb"
     repeat_region  7807. .7950
                    /rpt_family="MIR"
     repeat_region  7952. .8008
                    /rpt_family="MIR"
     repeat_region  complement(8362. .8424)
                    /rpt_family="MIR"
     repeat_region  8659. .8963
                    /rpt_family="AluSc"
     repeat_region  9048. .9191
                    /rpt_family="FLAM_C"
     repeat_region  complement(9332. .9366)
                    /rpt_family="POLY_A"
     repeat_region  complement(9380. .9463)
                    /rpt_family="LINE2"
     repeat_region  9774. .10005
                    /rpt_family="MER63A"
```

```
     repeat_region   10065. .10198
                     /rpt_family="MIR"
     repeat_region   10366. .11078
                     /rpt_family="MER4D"
     repeat_region   complement(11199. .11470)
                     /rpt_family="AluJb"
     repeat_region   11696. .11717
                     /rpt_family="AT_rich"
     repeat_region   complement(11718. .11895)
                     /rpt_family="MIR"
     repeat_region   12071. .12305
                     /rpt_family="AluY"
     repeat_region   13135. .13186
                     /rpt_family="LINE2"
     repeat_region   complement(13292. .13400)
                     /rpt_family="LINE2"
     repeat_region   13552. .13842
                     /rpt_family="AluSx"
     repeat_region   13864. .13960
                     /rpt_family="LINE2"
     repeat_region   14037. .14319
                     /rpt_family="AluJb"
     repeat_region   complement(14320. .14340)
                     /rpt_family="AT_rich"
     repeat_region   14667. .14746
                     /rpt_family="L1ME1"
     repeat_region   complement(14751. .15062)
                     /rpt_family="AluJb"
     repeat_region   15221. .15261
                     /rpt_family="(TAAG)n"
     repeat_region   17335. .17437
                     /rpt_family="LINE2"
     repeat_region   complement(17683. .17982)
                     /rpt_family="AluSx"
     repeat_region   18234. .18346
                     /rpt_family="FLAM_A"
     repeat_region   complement(18437. .18501)
                     /rpt_family="MIR"
     repeat_region   18638. .19147
                     /rpt_family="LINE2"
     repeat_region   20110. .20366
                     /rpt_family="MIR"
     repeat_region   complement(20452. .20809)
                     /rpt_family="LINE2"
     repeat_region   20854. .20985
                     /rpt_family="FLAM_C"
     repeat_region   20986. .21029
                     /rpt_family="POLY_A"
     repeat_region   21042. .21219
                     /rpt_family="L1ME"
     repeat_region   21268. .21416
```



```
                     /rpt_family="LINE2"
     repeat_region   21465. .21765
                     /rpt_family="AluY"
     repeat_region   22204. .22418
                     /rpt_family="AluJo"
     repeat_region   complement(22826. .22964)
                     /rpt_family="LINE2"
     repeat_region   23077. .23368
                     /rpt_family="AluSx"
     repeat_region   complement(23518. .23766)
                     /rpt_family="AluJo"
     repeat_region   complement(23770. .23903)
                     /rpt_family="MIR"
     repeat_region   23985. .24205
                     /rpt_family="AluSx"
     repeat_region   24305. .24435
                     /rpt_family="FLAM_C"
     repeat_region   complement(24482. .24797)
                     /rpt_family="AluSx"
     repeat_region   complement(25241. .25384)
                     /rpt_family="FLAM_C"
     repeat_region   complement(25460. .25758)
                     /rpt_family="AluSx"
     repeat_region   complement(26082. .26379)
                     /rpt_family="AluJb"
     repeat_region   complement(26524. .26764)
                     /rpt_family="AluSg/x"
     repeat_region   complement(26773. .27088)
                     /rpt_family="AluSx"
     repeat_region   28367. .28425
                     /rpt_family="MIR"
     repeat_region   complement(28767. .28798)
...
Note: remainder of annotations omitted.

  Query Match             6.4%;  Score 19;  DB 26;  Length 167101;
  Best Local Similarity 100.0%;  Pred. No. 5.52e+00;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db 161009 TAAATAAAAATATTTTTAA 161027
          ||||||||||||||||||
Cp    120 TAAATAAAAATATTTTTAA 102


RESULT   28
LOCUS       AC003668   170342 bp    DNA             HTG       14-MAR-1998
DEFINITION  *** SEQUENCING IN PROGRESS *** Homo sapiens; HTGS phase 1, 11
            unordered pieces.
ACCESSION   AC003668
NID         g2960511
KEYWORDS    HTG; HTGS_PHASE1.
```



```
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 170342)
  AUTHORS   Muzny,D., Arenson,A.D., Brundage,E., Carvelli,K., Chen,E., Di,W.,
            Ding,Y., Dugan,S., Durbin,J., Forcum,J., Ganesh,R., Garcia,C.,
            Goodman,M., Gorrell,J.H., Haywood,M., Jackson,L., Kampal,R.,
            Karpathy,S., Leal,B., Li,Y., Liu,W., Logan,O., Lu,J., Ly,T.,
            Martinez,C., Oswal,G., Peres,L., Rashid,N.D., Rowland,K.,
            Savage,L., Scherer,S.S., Shen,H., Timms,K.M., Todd,J., Vo,Q.,
            Worley,K.C., Yu,W., Chinault,C., Nelson,D. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 170342)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-1997) Molecular and Human Genetics, Baylor
            College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Mar 14, 1998 this sequence version replaced gi:2943851.
            ***                                                        ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                        ***
            * This sequence is unfinished. It consists of 11 contigs for
            * which the order is not known; their order in this record is
            * arbitrary. In some cases, the exact lengths of the gaps
            * between the contigs are also unknown; these gaps are  presented
            * as runs of N as a convenience only. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
            *       1   105950: contig of 105950 bp in length
            *  105951   105970: gap of unknown length
            *  105971   158358: contig of 52388 bp in length
            *  158359   158378: gap of unknown length
            *  158379   163530: contig of 5152 bp in length
            *  163531   163550: gap of unknown length
            *  163551   165005: contig of 1455 bp in length
            *  165006   165025: gap of unknown length
            *  165026   166000: contig of 975 bp in length
            *  166001   166020: gap of unknown length
            *  166021   166856: contig of 836 bp in length
            *  166857   166876: gap of unknown length
            *  166877   167697: contig of 821 bp in length
            *  167698   167717: gap of unknown length
            *  167718   168532: contig of 815 bp in length
            *  168533   168552: gap of unknown length
            *  168553   169290: contig of 738 bp in length
            *  169291   169310: gap of unknown length
            *  169311   170009: contig of 699 bp in length
            *  170010   170029: gap of unknown length
            *  170030   170342: contig of 313 bp in length.
```



```
FEATURES             Location/Qualifiers
     source          1. .170342
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT    52941 a  34345 c  33059 g  49787 t    210 others
ORIGIN

  Query Match             6.4%;  Score 19;  DB 17;  Length 170342;
  Best Local Similarity 100.0%;  Pred. No. 5.52e+00;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  25541 TTAAAAATATTTTTATTTA 25559
          ||||||||||||||||||
Qy    102 TTAAAAATATTTTTATTTA 120


RESULT   29
LOCUS       AC005477   185364 bp    DNA             HTG       14-AUG-1998
DEFINITION  *** SEQUENCING IN PROGRESS *** Homo sapiens clone DJ0283M22; HTGS
            phase 1, 25 unordered pieces.
ACCESSION   AC005477
NID         g3419866
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 185364)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 185364)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-AUG-1998) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     ***                                                        ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                        ***
            * This sequence is unfinished. It consists of 25 contigs for
            * which the order is not known; their order in this record is
            * arbitrary. In some cases, the exact lengths of the gaps
            * between the contigs are also unknown; these gaps are  presented
            * as runs of N as a convenience only. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
            *       1     2563: contig of 2563 bp in length
            *    2564     2581: gap of unknown length
            *    2582     5723: contig of 3142 bp in length
            *    5724     5741: gap of unknown length
```

```
              *     5742     8662: contig of 2921 bp in length
              *     8663     8680: gap of unknown length
              *     8681    11049: contig of 2369 bp in length
              *    11050    11067: gap of unknown length
              *    11068    13256: contig of 2189 bp in length
              *    13257    13274: gap of unknown length
              *    13275    16916: contig of 3642 bp in length
              *    16917    16934: gap of unknown length
              *    16935    19835: contig of 2901 bp in length
              *    19836    19853: gap of unknown length
              *    19854    23375: contig of 3522 bp in length
              *    23376    23393: gap of unknown length
              *    23394    28148: contig of 4755 bp in length
              *    28149    28166: gap of unknown length
              *    28167    32368: contig of 4202 bp in length
              *    32369    32386: gap of unknown length
              *    32387    38366: contig of 5980 bp in length
              *    38367    38384: gap of unknown length
              *    38385    44426: contig of 6042 bp in length
              *    44427    44444: gap of unknown length
              *    44445    52104: contig of 7660 bp in length
              *    52105    52122: gap of unknown length
              *    52123    57974: contig of 5852 bp in length
              *    57975    57992: gap of unknown length
              *    57993    65398: contig of 7406 bp in length
              *    65399    65416: gap of unknown length
              *    65417    72968: contig of 7552 bp in length
              *    72969    72986: gap of unknown length
              *    72987    79895: contig of 6909 bp in length
              *    79896    79913: gap of unknown length
              *    79914    88340: contig of 8427 bp in length
              *    88341    88358: gap of unknown length
              *    88359    97757: contig of 9399 bp in length
              *    97758    97775: gap of unknown length
              *    97776   108898: contig of 11123 bp in length
              *   108899   108916: gap of unknown length
              *   108917   118822: contig of 9906 bp in length
              *   118823   118840: gap of unknown length
              *   118841   132929: contig of 14089 bp in length
              *   132930   132947: gap of unknown length
              *   132948   148143: contig of 15196 bp in length
              *   148144   148161: gap of unknown length
              *   148162   165354: contig of 17193 bp in length
              *   165355   165372: gap of unknown length
              *   165373   185364: contig of 19992 bp in length.
FEATURES             Location/Qualifiers
     source          1. .185364
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="DJ0283M22"
BASE COUNT    55960 a  37832 c  37787 g  53353 t    432 others
```



```
ORIGIN

  Query Match           6.4%;  Score 19;  DB 17;  Length 185364;
  Best Local Similarity 100.0%;  Pred. No. 5.52e+00;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  13205 TAAAAATATTTTTAATGCA 13223
          ||||||||||||||||||
Cp    116 TAAAAATATTTTTAATGCA 98


RESULT   30
LOCUS       AC005037   190943 bp    DNA              HTG       12-JUL-1998
DEFINITION  *** SEQUENCING IN PROGRESS *** Homo sapiens clone NH0469M07; HTGS
            phase 1, 7 unordered pieces.
ACCESSION   AC005037
NID         g3309089
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 190943)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 190943)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-1998) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Jul 12, 1998 this sequence version replaced gi:3212954.
            ***                                                        ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                        ***
            * This sequence is unfinished. It consists of 7 contigs for
            * which the order is not known; their order in this record is
            * arbitrary. In some cases, the exact lengths of the gaps
            * between the contigs are also unknown; these gaps are  presented
            * as runs of N as a convenience only. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
            *        1     3875: contig of 3875 bp in length
            *     3876     3893: gap of unknown length
            *     3894    17937: contig of 14044 bp in length
            *    17938    17955: gap of unknown length
            *    17956    30958: contig of 13003 bp in length
            *    30959    30976: gap of unknown length
            *    30977    48923: contig of 17947 bp in length
            *    48924    48941: gap of unknown length
```



```
              *    48942    77647: contig of 28706 bp in length
              *    77648    77665: gap of unknown length
              *    77666   119989: contig of 42324 bp in length
              *   119990   120007: gap of unknown length
              *   120008   190943: contig of 70936 bp in length.
FEATURES             Location/Qualifiers
     source          1. .190943
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="NH0469M07"
BASE COUNT    55783 a  39619 c  38286 g  57147 t    108 others
ORIGIN

  Query Match           6.4%;  Score 19;  DB 17;  Length 190943;
  Best Local Similarity 100.0%;  Pred. No. 5.52e+00;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db 118712 AAAAATATTTTTATTTAAA 118730
          ||||||||||||||||||||
Qy    104 AAAAATATTTTTATTTAAA 122


RESULT   31
LOCUS       CEY46G5    302988 bp    DNA              HTG       08-OCT-1998
DEFINITION  Caenorhabditis elegans DNA *** SEQUENCING IN PROGRESS *** from
            clone Y46G5; HTGS phase 1.
ACCESSION   Z92852
NID         g3724184
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 302988)
  AUTHORS   Wallis,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-OCT-1998) Nematode Sequencing Project, Sanger Centre,
            Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics,
            Washington University, St. Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     On Oct 9, 1998 this sequence version replaced gi:3702093.
            Order of segments is not known; 800 n's separate segments.
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and the release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc.
            ***                                                        ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                        ***
```



```
            * This sequence is unfinished. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
FEATURES             Location/Qualifiers
     source          1. .302988
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /chromosome="II"
                     /clone="Y46G5"
BASE COUNT    92609 a  55023 c  53868 g  92680 t   8808 others
ORIGIN

  Query Match           6.4%;  Score 19;  DB 17;  Length 302988;
  Best Local Similarity 100.0%;  Pred. No. 5.52e+00;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db 217249 TAAAAATATTTTTATTTAA 217267
          ||||||||||||||||||||
Qy    103 TAAAAATATTTTTATTTAA 121


RESULT   32
LOCUS       G15468       250 bp    DNA              STS       04-JAN-1996
DEFINITION  human STS SHGC-16896.
ACCESSION   G15468
NID         g1132231
KEYWORDS    STS sequence; primer; sequence tagged site.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Choanata;
            Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates;
            Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 250)
  AUTHORS   Myers,R.M.
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Richard M. Myers
            Stanford Human Genome Center (SHGC)
            Stanford University School of Medicine
            Department of Genetics, M-344, Stanford, CA 94305, USA
            Tel: 4157259687
            Fax: 4157259689
            Email: myers@shgc.stanford.edu

            Primer A: AAAGTTCAAAAACTGGCCGA
            Primer B: AGCCCCCACCTAAGAGGTAA
            STS size: 131
            PCR Profile:
                    Initial incubation: 94 degrees C for 90 seconds
```

```
                    Denaturation:       94 degrees C for 15 seconds
                    Annealing:          62 degrees C for 23 seconds
                    Polymerization:     72 degrees C for 30 seconds
                    PCR Cycles:         30
                    Thermal Cycler:     Perkin Elmer 9600
            Protocol:
                    Template:           25 ng
                    Primer:             each 1 uM
                    dNTPs:              each 200 uM
                    Taq Polymerase:     0.05 units/ul
                    Total Vol:          10 ul

            Buffer:
                    MgCl2:              2.5 mM
                    KCl:                50  mM
                    Tris-HCl:           20  mM
                    pH:                 8.3

            Prepared with primer pairs derived from T35708--Merck/UniEST.
FEATURES            Location/Qualifiers
     source         1. .250
                    /organism="Homo sapiens"
     STS            21. .151
     primer_bind    21. .40
     primer_bind    complement(132. .151)
BASE COUNT      86 a     30 c     53 g     77 t      4 others
ORIGIN

  Query Match           6.1%;  Score 18;  DB 30;  Length 250;
  Best Local Similarity 100.0%;  Pred. No. 2.36e+01;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    101 AAAGTTATTTAAATAAAA 118
          ||||||||||||||||||
Cp    129 AAAGTTATTTAAATAAAA 112


RESULT   33
LOCUS       BTMSINRBP     351 bp    DNA             MAM       27-JUN-1994
DEFINITION  B.taurus microsatellite DNA INRA150.
ACCESSION   X74182
NID         g508143
KEYWORDS    microsatellite.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Artiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 351)
  AUTHORS   Vaiman,D., Mercier,D., Moazami-Goudarzi,K., Eggen,A. and
            Leveziel,H.
```



```
  TITLE     Somatic cell mapping and characterization of new bovine
            microsatellites
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 351)
  AUTHORS   Vaiman,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUL-1993) D. Vaiman, Inra, Laboratoire de Genetique
            Biochimique, et de Cytogenetique, INRA CRJ Domaine de Vilvert,
            78352 Jouy en Josas Cedex, FRANCE
FEATURES            Location/Qualifiers
     source         1. .351
                    /organism="Bos taurus"
                    /db_xref="taxon:9913"
     satellite      1. .351
                    /note="microsatellite INRA150"
BASE COUNT      135 a     64 c     41 g    111 t
ORIGIN

  Query Match           6.1%;  Score 18;  DB 19;  Length 351;
  Best Local Similarity 100.0%;  Pred. No. 2.36e+01;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    141 TTTAAATAAAAATATTTT 158
          ||||||||||||||||||
Cp    122 TTTAAATAAAAATATTTT 105


RESULT   34
LOCUS       NAJRG16S      405 bp    DNA             VRT       11-FEB-1993
DEFINITION  Naja naja naja 16S ribosomal RNA gene, partial.
ACCESSION   L10674
NID         g213377
KEYWORDS    16S ribosomal RNA.
SOURCE      Naja naja (sub_species naja) DNA.
  ORGANISM  Naja naja
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Lepidosauria; Squamata; Scleroglossa; Serpentes;
            Colubroidea; Elapidae; Elapinae; Naja.
REFERENCE   1  (bases 1 to 405)
  AUTHORS   Knight,A.
  TITLE     The phylogenetic relationship of Colubrinae, Elapidae and Viperidae
            and the evolution of front-fanged venom systems in snakes
  JOURNAL   Unpublished (1993)
FEATURES            Location/Qualifiers
     source         1. .405
                    /organism="Naja naja"
                    /sub_species="naja"
                    /db_xref="taxon:35670"
     rRNA           <1. .>405
                    /gene="16S rRNA"
                    /product="16S ribosomal RNA"
```



```
     gene           1. .405
                    /gene="16S rRNA"
BASE COUNT      136 a     94 c     75 g    100 t
ORIGIN

  Query Match           6.1%;  Score 18;  DB 20;  Length 405;
  Best Local Similarity 100.0%;  Pred. No. 2.36e+01;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     31 TATCTGTCTCTTGTAGTA 48
          ||||||||||||||||||
Qy    230 TATCTGTCTCTTGTAGTA 247


RESULT   35
LOCUS       MTNNRNA16     412 bp    DNA             VRT       23-FEB-1995
DEFINITION  N.naja mitochondrial gene for 16S ribosomal RNA.
ACCESSION   Z46482
NID         g683635
KEYWORDS    16S ribosomal RNA; 16S rRNA gene.
SOURCE      Naja naja.
  ORGANISM  Mitochondrion Naja naja
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Lepidosauria; Squamata; Scleroglossa; Serpentes;
            Colubroidea; Elapidae; Elapinae; Naja.
REFERENCE   1  (bases 1 to 412)
  AUTHORS   Heise,P.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-NOV-1994) Heise P. J., Pennsylvania State University,
            Department of Biology, 208 Erwin W. Mueller Laboratory, University
            Park, Pennsylvania, USA, 16802
REFERENCE   2  (bases 1 to 412)
  AUTHORS   Heise,P.J., Maxson,L.R., Dowling,H.G. and Hedges,S.B.
  TITLE     Higher-level snake phylogeny inferred from mitochondrial DNA
            sequences of 12S rRNA and 16S rRNA genes
  JOURNAL   Mol. Biol. Evol. 12 (2), 259-265 (1995)
  MEDLINE   95214537
FEATURES            Location/Qualifiers
     source         1. .412
                    /organism="Naja naja"
                    /mitochondrion
                    /db_xref="taxon:35670"
                    /tissue_type="liver"
     rRNA           <1. .>412
                    /gene="16S rRNA gene"
                    /product="16S ribosomal RNA"
     gene           1. .412
                    /gene="16S rRNA gene"
BASE COUNT      139 a     98 c     76 g     97 t      2 others
ORIGIN
```



```
  Query Match           6.1%;  Score 18;  DB 20;  Length 412;
  Best Local Similarity 100.0%;  Pred. No. 2.36e+01;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     44 TATCTGTCTCTTGTAGTA 61
          ||||||||||||||||||
Qy    230 TATCTGTCTCTTGTAGTA 247


RESULT   36
LOCUS       G21159        422 bp    DNA             STS       31-MAY-1996
DEFINITION  human STS WI-14062.
ACCESSION   G21159
NID         g1341485
KEYWORDS    STS sequence; primer; sequence tagged site.
SOURCE      human STSs derived from sequences in dbEST and the Unigene
            collection.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 422)
  AUTHORS   Hudson,T.
  TITLE     Whitehead Institute/MIT Center for Genome Research; Physically
            Mapped STSs
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Thomas Hudson
            Whitehead Institute/MIT Center for Genome Research
            Whitehead Institute for Biomedical Research
            9 Cambridge Center, Cambridge MA 02142 USA
            Tel: 617 252 1900
            Fax: 617 252 1902
            Email: thudson@genome.wi.mit.edu

            Primer A: ATCTCACAACTGTAATTCGAACTTG
            Primer B: CCCAGGTTATTAAAGGATTCACA
            STS size: 110
            PCR Profile:
                  Presoak:
                  Denaturation:
                  Annealing: 56 degrees C
                  Polymerization:
                  PCR Cycles: 35
                  Thermal Cycler:
            Protocol:
                Template: 10 ng
                Primer: each 5 pM
                dNTPs: each 4 nM
                Taq Polymerase: 0.025 units/ul
                Total Vol: 20 ul
```

```
            Buffer:
                MgCl2: 1.5 mM
                KCl: 50 mM
                Tris-HCL: 10 mM
                pH: 9.3

            Derived from dbEST (genbank accession R43863).
FEATURES             Location/Qualifiers
     source          1. .422
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="107.5 cR from top of Chr7 linkage group"
     STS             53. .162
     primer_bind     53. .77
     primer_bind     complement(140. .162)
BASE COUNT      133 a      67 c      61 g     153 t       8 others
ORIGIN

  Query Match           6.1%;  Score 18;  DB 30;  Length 422;
  Best Local Similarity 100.0%;  Pred. No. 2.36e+01;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      31 ATATTTTTATTTAAATAA 48
           ||||||||||||||||||
Qy     108 ATATTTTTATTTAAATAA 125


RESULT   37
LOCUS       PTAJ5292      504 bp    DNA             PLN       02-APR-1998
DEFINITION  Populus tremula x Populus tremuloides hybrid, transgenic line
            Esch5:35S-rolC-#16, junction right T-DNA border/genomic sequence.
ACCESSION   AJ005292
NID         g3021559
KEYWORDS    junction sequence; t-dna.
SOURCE      Populus tremula x Populus tremuloides.
  ORGANISM  Populus tremula x Populus tremuloides
            Eukaryota; Viridiplantae; Charophyta/Embryophyta group;
            Embryophyta; Tracheophyta; seed plants; Magnoliophyta;
            eudicotyledons; Rosidae; Violales; Salicaceae; Populus.
REFERENCE   1  (bases 1 to 504)
  AUTHORS   Pladung,M.
  TITLE     Gene stability in transgenic aspen-Populus. I. Flanking DNA
            sequences and T-DNA structure
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 504)
  AUTHORS   Pladung,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-1998) Pladung M., Institut fuer Forstgenetik und
            Forstpflanzenzuechtung, Bundesforschungsanstalt fuer Forst- und
            Holzwirtschaft, Sieker Landstrasse 2, 22927 Grosshansdorf, FRG
FEATURES             Location/Qualifiers
```



```
     source          1. .504
                     /organism="Populus tremula x Populus tremuloides"
                     /cultivar="Esch5"
                     /isolate="Esch5:35-rolC-#16"
                     /db_xref="taxon:47664"
                     /sex="Female"
                     /tissue_type="leaf"
     source          1. .91
                     /note="T-DNA of plasmid pPCV002-CaMV-35S-rolC"
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
     source          1. .504
                     /organism="unidentified"
                     /db_xref="taxon:32644"
     misc_feature    89. .91
                     /note="right T-DNA border"
BASE COUNT      156 a      93 c      56 g     199 t
ORIGIN

  Query Match           6.1%;  Score 18;  DB 23;  Length 504;
  Best Local Similarity 100.0%;  Pred. No. 2.36e+01;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     330 TTTAAATAAAAATATTTT 347
           ||||||||||||||||||
Cp     122 TTTAAATAAAAATATTTT 105


RESULT   38
LOCUS       G24158        504 bp    DNA             STS       31-MAY-1996
DEFINITION  human STS WI-13539.
ACCESSION   G24158
NID         g1344484
KEYWORDS    STS sequence; primer; sequence tagged site.
SOURCE      human STSs derived from sequences in dbEST and the Unigene
            collection.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 504)
  AUTHORS   Hudson,T.
  TITLE     Whitehead Institute/MIT Center for Genome Research; Physically
            Mapped STSs
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Thomas Hudson
            Whitehead Institute/MIT Center for Genome Research
            Whitehead Institute for Biomedical Research
            9 Cambridge Center, Cambridge MA 02142 USA
            Tel: 617 252 1900
            Fax: 617 252 1902
```



```
            Email: thudson@genome.wi.mit.edu

            Primer A: TTTTTTTAGGTAGGAGAAAACTGC
            Primer B: GTGAATATTCCCATTTTCTTCCA
            STS size: 150
            PCR Profile:
                 Presoak:
                 Denaturation:
                 Annealing: 56 degrees C
                 Polymerization:
                 PCR Cycles: 35
                 Thermal Cycler:
            Protocol:
                Template: 10 ng
                Primer: each 5 pM
                dNTPs: each 4 nM
                Taq Polymerase: 0.025 units/ul
                Total Vol: 20 ul

            Buffer:
                MgCl2: 1.5 mM
                KCl: 50 mM
                Tris-HCL: 10 mM
                pH: 9.3

            Derived from dbEST (genbank accession R41639).
FEATURES             Location/Qualifiers
     source          1. .504
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="781.7 cR from top of Chr1 linkage group"
     STS             1. .150
     primer_bind     1. .24
     primer_bind     complement(128. .150)
BASE COUNT      134 a      74 c     119 g     162 t      15 others
ORIGIN

  Query Match           6.1%;  Score 18;  DB 30;  Length 504;
  Best Local Similarity 100.0%;  Pred. No. 2.36e+01;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     218 TGTTTCTAAAAGTTATAA 235
           ||||||||||||||||||
Cp     149 TGTTTCTAAAAGTTATAA 132


RESULT   39
LOCUS       TETHIS22MA    905 bp    DNA             INV       15-JUN-1990
DEFINITION  T.thermophila macronuclear histone H2B-2 gene, complete cds.
ACCESSION   M31333
NID         g161785
```



```
KEYWORDS    histone.
SOURCE      T.thermophila macronuclear DNA.
  ORGANISM  Tetrahymena thermophila
            Eukaryotae; mitochondrial eukaryotes; Alveolata; Ciliophora;
            Oligohymenophorea; Hymenostomatida; Tetrahymenina; Tetrahymena.
REFERENCE   1  (bases 1 to 905)
  AUTHORS   Nomoto,M., Imai,N., Saiga,H., Matsui,T. and Mita,T.
  TITLE     Characterization of two types of histone H2B genes from macronuclei
            of Tetrahymena thermophila
  JOURNAL   Nucleic Acids Res. 15, 5681-5697 (1987)
  MEDLINE   87289037
FEATURES             Location/Qualifiers
     source          1. .905
                     /organism="Tetrahymena thermophila"
                     /db_xref="taxon:5911"
     mRNA            189. .809
                     /note="H2B-2 mRNA (minor alt., 3' end put.); putative"
     mRNA            192. .809
                     /note="H2B-2 mRNA (major alt., 3' end put.); putative"
     mRNA            203. .809
                     /note="H2B-2 mRNA (minor alt., 3' end put.); putative"
     CDS             231. .599
                     /note="histone H2B-2"
                     /codon_start=1
                     /transl_table=6
                     /db_xref="PID:g161786"
                     /translation="MAPKKAPAATTEKKVKKAPTTEKKNKKKRSETPAIYIFKVLKQV
                     HPDVGISKKAMNIMNSFINDSFERIALESSKLVRFNKRRTLSSREVQTAVKLLLPGEL
                     ARHAISEGTKAVTKPSSSSN"
BASE COUNT      326 a     160 c     138 g     281 t
ORIGIN

  Query Match           6.1%;  Score 18;  DB 18;  Length 905;
  Best Local Similarity 100.0%;  Pred. No. 2.36e+01;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      36 GGCAAATTGAATGAAGTA 53
           ||||||||||||||||||
Qy     208 GGCAAATTGAATGAAGTA 225


RESULT   40
LOCUS       TTH2B2        905 bp    DNA             INV       22-OCT-1996
DEFINITION  Tetrahymena thermophila gene for histone H2B-2 from macronucleus.
ACCESSION   X05544 M31333 Y00376
NID         g10813
KEYWORDS    histone; histone H2B-2.
SOURCE      Tetrahymena thermophila.
  ORGANISM  Tetrahymena thermophila
            Eukaryotae; mitochondrial eukaryotes; Alveolata; Ciliophora;
            Oligohymenophorea; Hymenostomatida; Tetrahymenina; Tetrahymena.
```

```
REFERENCE   1  (bases 1 to 905)
  AUTHORS   Nomoto,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUL-1987) Minoru Nomoto, Department of Molecular
            Biology, School of Medicine, University of Occupational and
            Environmental Health, 1-1 Iseigaoka, Yahatanishi-Ku, Kitakyushu
            807, Japan
REFERENCE   2  (bases 1 to 905)
  AUTHORS   Nomoto,M., Imai,N., Saiga,H., Matsui,T. and Mita,T.
  TITLE     Characterization of two types of histone H2B genes from macronuclei
            of Tetrahymena thermophila
  JOURNAL   Nucleic Acids Res. 15 (14), 5681-5697 (1987)
  MEDLINE   87289037
COMMENT     see also x05543 for H2B-1.
FEATURES             Location/Qualifiers
     source          1. .905
                     /organism="Tetrahymena thermophila"
                     /db_xref="taxon:5911"
                     /clone="pTH2B-2"
     misc_feature    85. .101
                     /note="conserved sequence block"
     misc_feature    111. .116
                     /note="conserved sequence block"
     misc_feature    139. .144
                     /note="conserved sequence block"
     misc_feature    192
                     /note="put. transcription start site"
     CDS             231. .599
                     /note="(AA 1 - 122)"
                     /codon_start=1
                     /transl_except=(pos:501. .503,aa:Gln)
                     /transl_table=6
                     /product="Histone H2B-2"
                     /db_xref="PID:g578564"
                     /db_xref="SWISS-PROT:P08994"
                     /translation="MAPKKAPAATTEKKVKKAPTTEKKNKKKRSETPAIYIPKVLKQV
                     HPDVGISKKAMNIMNSFINDSFERIALESSKLVRFNKRRTLSSREVQTAVKLLLPGEL
                     ARHAISEGTKAVTKFSSSSN"
BASE COUNT      326 a    160 c    138 g    281 t
ORIGIN

  Query Match           6.1%;  Score 18;  DB 18;  Length 905;
  Best Local Similarity 100.0%;  Pred. No. 2.36e+01;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     36 GGCAAATTGAATGAAGTA 53
          ||||||||||||||||||
Qy    208 GGCAAATTGAATGAAGTA 225


RESULT   41
```



```
LOCUS       HSSCPXT16    1052 bp    DNA             PRI       25-JAN-1997
DEFINITION  Human sterol carrier protein-X/sterol carrier protein-2
            (SCP-X/SCP-2) gene, exon 16, and complete cds.
ACCESSION   U11313
NID         g532077
KEYWORDS    .
SEGMENT     16 of 16
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1052)
  AUTHORS   Ohba,T., Rennert,H., Pfeifer,S.M., He,Z., Yamamoto,R., Holt,J.A.,
            Billheimer,J.T. and Strauss,J.F. 3rd.
  TITLE     The structure of the human sterol carrier protein X/sterol carrier
            protein 2 gene (SCP2)
  JOURNAL   Genomics 24 (2), 370-374 (1994)
  MEDLINE   95213031
REFERENCE   2  (bases 1 to 1052)
  AUTHORS   Strauss,J.F. III.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUN-1994) Jerome F. Strauss III, Department of
            Obstetrics and Gynecology, Division of Reproductive Biology,
            University of Pennsylvania Medical Center, 778 Clinical Research
            Building, 422 Curie Boulevard, Philadelphia, PA 19104-6142, USA
FEATURES             Location/Qualifiers
     source          1. .1052
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /tissue_type="liver"
                     /map="1p32"
     gene            join(U11297:1. .3666,U11299:1. .108,U11300:1. .122,
                     U11301:1. .182,U11302:1. .115,U11303:1. .177,U11304:1. .114,
                     U11305:1. .137,U11306:1. .201,U11307:1. .198,U11308:1. .158,
                     U11309:1. .1725,U11310:1. .153,U11311:1. .180,U11312:1. .130,
                     1. .1034)
                     /partial
                     /gene="SCP-X/SCP-2"
     mRNA            join(U11297:3452. .3641,U11299:26. .83,U11300:26. .97,
                     U11301:26. .157,U11302:26. .90,U11303:26. .152,U11304:26. .89,
                     U11305:26. .112,U11306:26. .176,U11307:26. .173,
                     U11308:26. .133,U11309:1547. .1700,U11310:26. .128,
                     U11311:26. .155,U11312:26. .105,26. .1034)
                     /gene="SCP-X/SCP-2"
                     /product="sterol carrier protein-X/sterol carrier
                     protein-2"
     CDS             join(U11297:3573. .3641,U11299:26. .83,U11300:26. .97,
                     U11301:26. .157,U11302:26. .90,U11303:26. .152,U11304:26. .89,
                     U11305:26. .112,U11306:26. .176,U11307:26. .173,
                     U11308:26. .133,U11309:1547. .1700,U11310:26. .128,
```



```
                     U11311:26. .155,U11312:26. .105,26. .121)
                     /gene="SCP-X/SCP-2"
                     /codon_start=1
                     /product="sterol carrier protein-X/sterol carrier
                     protein-2"
                     /db_xref="PID:g1773239"
                     /translation="MSSSPWEPAALRRVFVVGVGMTKFVKPGAENSRDYPDLAEEAGK
                     KALADAQIPTSAVDQACVGYVFGDSTCGQRAIYHSLGMTGIPIINVNNNCATGSTALF
                     MARQLIQGGVAECVLALGFEKMSKGSLGIKFSDRTIPTDKHVDLLINKYGLSAHPVAP
                     QMFGYAGKEHMEKYGTKIEHFAKIGWKNHKHSVNNPYSQPQDETSLDEVMASKEVFDF
                     LTILQCCPTSDGAAAAILASEAFVQKYGLQSKAVEILAQEMMTDLPSSFEEKSIIKMV
                     GFDMSKEAARKCYEKSGLTPNDIDVIELHDCFSTNELLTYEALGLCPEGQGATLVDRG
                     DNTYGGKWVINPSGGLISKGHPLGATGLAQCAELCWQLRGEAGKRQVPGAKVALQHNL
                     GIGGAVVVTLYKMGPPEAASSPRTHQIEAVPTSSASDGFKANLVFKEIEKKLEEEGEQ
                     FVKKIGGIFAFKVKDGPGGKEATWVVDVKNGKGSVLPNSDKKADCTITMADSDFLALM
                     TGKMNPQSAFPQGKLKITGNMGLAMKLQNLQLQPGNAKL"
     exon            26. .1034
                     /gene="SCP-X/SCP-2"
                     /note="nucleotides 122-1027 are not translated"
                     /number=16
BASE COUNT      352 a    157 c    164 g    379 t
ORIGIN

  Query Match           6.1%;  Score 18;  DB 25;  Length 1052;
  Best Local Similarity 100.0%;  Pred. No. 2.36e+01;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    713 AAATAAAAATATTTTTAA 730
          ||||||||||||||||||
Cp    119 AAATAAAAATATTTTTAA 102


RESULT   42
LOCUS       HUMSTEAA     1219 bp    mRNA            PRI       06-DEC-1993
DEFINITION  Human sterol carrier protein-2 (SCP-2) mRNA, complete cds.
ACCESSION   M55421
NID         g432978
KEYWORDS    sterol carrier protein-2.
SOURCE      Human female liver.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1219)
  AUTHORS   Yamamoto,R., Kallen,C.B., Babalola,G.O., Rennert,H.,
            Billheimer,J.T. and Strauss,J.F.III.
  TITLE     Cloning and expression of a cDNA encoding human sterol carrier
            protein 2
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 88, 463-467 (1991)
  MEDLINE   91110550
REFERENCE   2  (bases 1 to 1219)
  AUTHORS   Vesa,J., Hellsten,E., Branoski,B.L., Emanuel,B.S., Billheimer,J.T.,
```



```
            Mead,S., Cowell,J.K., Strauss,J.F.III. and Peltonen,L.
  TITLE     Assignment of sterol carrier protein X/sterol carrier protein 2 to
            1p32 and exclusion as the causative gene for infantile neuronal
            ceroid lipofusionosis
  JOURNAL   Unpublished
COMMENT     On Dec 6, 1993 this sequence version replaced gi:432972.
FEATURES             Location/Qualifiers
     source          1. .1219
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="liver"
                     /map="1p32"
     gene            84. .515
                     /gene="SCP-2"
     CDS             84. .515
                     /gene="SCP-2"
                     /codon_start=1
                     /product="sterol carrier protein-2"
                     /db_xref="PID:g432973"
                     /translation="MGFPEAASSFRTHQIEAVPTSSASDGFKANLVFKEIEKKLEEEG
                     EQFVKKIGGIFAFKVKDGPGGKEATWVVDVKNGKGSVLPNSDKKADCTITMADSDFLA
                     LMTGKMNPQSAFPQGKLKITGNMGLAMKLQNLQLQPGNAKL"
     sig_peptide     84. .143
                     /gene="SCP-2"
     mat_peptide     144. .512
                     /gene="SCP-2"
     polyA_signal    782. .787
                     /evidence=not_experimental
     polyA_signal    1062. .1067
                     /evidence=not_experimental
     polyA_signal    1107. .1112
                     /evidence=not_experimental
     polyA_signal    1190. .1195
                     /evidence=experimental
     polyA_site      1219
                     /evidence=experimental
BASE COUNT      402 a    195 c    255 g    367 t
ORIGIN

  Query Match           6.1%;  Score 18;  DB 25;  Length 1219;
  Best Local Similarity 100.0%;  Pred. No. 2.36e+01;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1106 AAATAAAAATATTTTTAA 1123
          ||||||||||||||||||
Cp    119 AAATAAAAATATTTTTAA 102


RESULT   43
LOCUS       S52450       1229 bp    mRNA            PRI       08-MAR-1993
DEFINITION  sterol carrier protein 2 [human, liver, mRNA, 1229 nt].
```

```
ACCESSION   S52450
NID         g263550
KEYWORDS    .
SOURCE      human liver.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1229)
  AUTHORS   Yamamoto,R.
  TITLE     [Localization of human sterol carrier protein 2 gene and cDNA
            expression in COS-7 cell]
  JOURNAL   Hokkaido Igaku Zasshi 67 (6), 839-848 (1992)
  MEDLINE   93131254
  REMARK    GenBank staff at the National Library of Medicine created this
            entry [NCBI gibbsq 122504] from the original journal article.
            This sequence comes from Fig. 1.
            Map location: 1.
FEATURES             Location/Qualifiers
     source          1. .1229
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     gene            84. .515
                     /gene="sterol carrier protein 2, SCP2"
     CDS             84. .515
                     /gene="sterol carrier protein 2, SCP2"
                     /note="This sequence comes from Fig. 1; SCP2"
                     /codon_start=1
                     /product="sterol carrier protein 2"
                     /db_xref="PID:g263551"
                     /translation="MGFPEAASSFRTHQIEAVPTSSASDGFKANLVFKEIEKKLEEEG
                     EQFVKKIGGIFAFKVKDGPGGKEDTWVVDVKNGQGSVLPNSDKKADCTITMAASDFLA
                     LMTGKMNPQSAFFQGPLKITGNMGLAMKLQNLQLQPGNAKL"
BASE COUNT      409 a    200 c    255 g    365 t
ORIGIN

  Query Match           6.1%;  Score 18;  DB 25;  Length 1229;
  Best Local Similarity 100.0%;  Pred. No. 2.36e+01;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1107 AAATAAAAATATTTTTAA 1124
           ||||||||||||||||||
Cp    119 AAATAAAAATATTTTTAA 102


RESULT   44
LOCUS       SLU73444     1260 bp    DNA              BCT       11-MAR-1997
DEFINITION  Staphylococcus lugdunensis SLUSH A (slushA), SLUSH B (slushB), and
            hemolysin (slushC) genes, complete cds.
ACCESSION   U73444
NID         g1778750
KEYWORDS    .
```



```
SOURCE      Staphylococcus lugdunensis.
  ORGANISM  Staphylococcus lugdunensis
            Eubacteria; Firmicutes; Low G+C gram-positive bacteria;
            Bacillaceae; Staphylococcus.
REFERENCE   1  (bases 1 to 1260)
  AUTHORS   Donvito,B., Etienne,J., Denoroy,L., Greenland,T., Benito,Y. and
            Vandenesch,F.
  TITLE     Synergistic hemolytic activity of Staphylococcus lugdunensis is
            mediated by three peptides encoded by a non-agr genetic locus
  JOURNAL   Infect. Immun. 65 (1), 95-100 (1997)
  MEDLINE   97130021
REFERENCE   2  (bases 1 to 1260)
  AUTHORS   Donvito,B., Etienne,J., Denoroy,L., Greenland,T., Benito,Y. and
            Vandenesch,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-OCT-1996) Microbiology, Facult Laennec, Rue G.
            Paradin, Lyon Cedex 03 69372, France
FEATURES             Location/Qualifiers
     source          1. .1260
                     /organism="Staphylococcus lugdunensis"
                     /strain="307"
                     /db_xref="taxon:28035"
     gene            570. .701
                     /gene="slushA"
     CDS             570. .701
                     /gene="slushA"
                     /function="mediates haemolytic activity"
                     /codon_start=1
                     /transl_table=11
                     /product="SLUSH A"
                     /db_xref="PID:g1778751"
                     /translation="MSGIVDAITKAVQAGLDKDWATMATSIADAIAKGVDFIAGFFN"
     gene            751. .882
                     /gene="slushB"
     CDS             751. .882
                     /gene="slushB"
                     /function="mediates haemolytic activity"
                     /codon_start=1
                     /transl_table=11
                     /product="SLUSH B"
                     /db_xref="PID:g1778752"
                     /translation="MSGIIEAITKAVQAGLDKDWATMGTSIAEALAKGIDAISGLFG"
     gene            944. .1075
                     /gene="slushC"
     CDS             944. .1075
                     /gene="slushC"
                     /codon_start=1
                     /transl_table=11
                     /product="hemolysin"
                     /db_xref="PID:g1778753"
                     /translation="MDGIFEAISKAVQAGLDKDWATMGTSIAEALAKGVDFIIGLFH"
```



```
BASE COUNT      428 a    189 c    219 g    424 t
ORIGIN

  Query Match           6.1%;  Score 18;  DB 15;  Length 1260;
  Best Local Similarity 100.0%;  Pred. No. 2.36e+01;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    911 AAATAAAAATATTTTTAA 928
          ||||||||||||||||||
Cp    119 AAATAAAAATATTTTTAA 102


RESULT   45
LOCUS       AF070566     1530 bp    mRNA             PRI       05-AUG-1998
DEFINITION  Homo sapiens clones 24611 and 24705 mRNA sequence.
ACCESSION   AF070566
NID         g3387934
KEYWORDS    FLI_CDNA.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1530)
  AUTHORS   Andersson,B., Wentland,M.A., Ricafrente,J.Y., Liu,W. and Gibbs,R.A.
  TITLE     A 'double adaptor' method for improved shotgun library construction
  JOURNAL   Anal. Biochem. 236 (1), 107-113 (1996)
  MEDLINE   96207227
REFERENCE   2  (bases 1 to 1530)
  AUTHORS   Yu,W., Andersson,B., Worley,K.C., Muzny,D.M., Ding,Y., Liu,W.,
            Ricafrente,J.Y., Wentland,M.A., Lennon,G. and Gibbs,R.A.
  TITLE     Large-scale concatenation cDNA sequencing
  JOURNAL   Genome Res. 7 (4), 353-358 (1997)
  MEDLINE   97264341
REFERENCE   3  (bases 1 to 1530)
  AUTHORS   Yu,W. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JUN-1998) Molecular and Human Genetics, Baylor
            College of Medicine, One Baylor Plaza S930, Houston, TX 77030, USA
FEATURES             Location/Qualifiers
     source          1. .1530
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /sex="female"
                     /dev_stage="infant"
                     /tissue_type="brain"
                     /clone="I.M.A.G.E. Consortium clone ID 24611; 24705"
                     /clone_lib="1NIB"
BASE COUNT      465 a    260 c    304 g    501 t
ORIGIN

  Query Match           6.1%;  Score 18;  DB 26;  Length 1530;
```



```
  Best Local Similarity 100.0%;  Pred. No. 2.36e+01;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    857 CATGTAGTATTGTTTAAA 874
          ||||||||||||||||||
Qy    148 CATGTAGTATTGTTTAAA 165


RESULT   46
LOCUS       HSU85994     1678 bp    mRNA             PRI       01-JUL-1997
DEFINITION  Human clone IMAGE:30008 unknown protein mRNA, partial cds.
ACCESSION   U85994
NID         g1835747
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 1678)
  AUTHORS   Keen,T.J.
  TITLE     Positional candidates for the RP9 retinitis pigmentosa gene
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1678)
  AUTHORS   Keen,T.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JAN-1997) Department of Molecular Genetics, Institute
            of Ophthalmology, University College London, Bath St., London EC1V
            9EL, UK
FEATURES             Location/Qualifiers
     source          1. .1678
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7p14-15"
                     /tissue_type="brain"
                     /dev_stage="infant"
                     /clone_lib="Soares 1NIB"
                     /clone="IMAGE:30008"
     CDS             <1. .836
                     /codon_start=3
                     /product="unknown"
                     /db_xref="PID:g1835748"
                     /translation="LITNELILRLQEYFEKQGVKDFACSFSGSIPLQEYFELIDHHFE
                     LRINGERLEELLSERAVQFRAIQRRLLARFKDKTPAPLQHLDTLLDGTYKQVIALADA
                     VEENQGNLFQSFTRLKSATHLVILLIALWQKLSADQVAILEAAPLPLQEDTQELGWEE
                     TVDAAISHLLKTCLSKSSKEQALNLNSQLNIPKDTSQLKKHITLLCDRLSKGGRLCLS
                     TDAAAPQTMVMPGGCTTIPESDLEERSVEQDSTELFTNHRHLTAETPRPEVSPLQGVS
                     E"
     polyA_signal    1655. .1660
BASE COUNT      520 a    331 c    346 g    481 t
```

```
ORIGIN

  Query Match           6.1%;  Score 18;  DB 26;  Length 1678;
  Best Local Similarity 100.0%;  Pred. No. 2.36e+01;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1648 TTATTTAAATAAAAATAT 1665
          ||||||||||||||||||
Cp    125 TTATTTAAATAAAAATAT 108


RESULT   47
LOCUS       HSU85997     1892 bp    DNA             PRI       10-FEB-1997
DEFINITION  Human unknown protein gene, partial cds.
ACCESSION   U85997
NID         g1835632
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1892)
  AUTHORS   Keen,T.J.
  TITLE     Positional candidates for the RP9 retinitis pigmentosa gene
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1892)
  AUTHORS   Keen,T.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JAN-1997) Department of Molecular Genetics, Institute
            of Ophthalmology, University College London, Bath St., London EC1V
            9EL, UK
FEATURES             Location/Qualifiers
     source          1. .1892
                     /organism="Homo sapiens"
                     /note="sequence demonstrates that IMAGE:30008 is derived
                     from the same gene as IMAGE:22181 and IMAGE:74593"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7p14-p15"
     intron          <1. .382
     CDS             join(<383. .493,713. .744)
                     /codon_start=3
                     /product="unknown"
                     /db_xref="PID:g1835633"
                     /translation="GCTTIPESDLEERSVEQDSTELPTNHRHLTAETPRPEVSPLQGV
                     SE"
     intron          494. .712
                     /note="Retained in IMAGE:74593"
     polyA_signal    1119. .1125
                     /note="Used in IMAGE:22181 and IMAGE:74593"
     polyA_signal    1563. .1568
```



```
                     /note="Used in IMAGE:30008"
BASE COUNT      574 a    321 c    353 g    644 t
ORIGIN

  Query Match           6.1%;  Score 18;  DB 25;  Length 1892;
  Best Local Similarity 100.0%;  Pred. No. 2.36e+01;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1556 TTATTTAAATAAAAATAT 1573
          ||||||||||||||||||
Cp    125 TTATTTAAATAAAAATAT 108


RESULT   48
LOCUS       DMU77310     2424 bp    mRNA            INV       18-NOV-1996
DEFINITION  Drosophila melanogaster porcupine mRNA, complete cds.
ACCESSION   U77310
NID         g1673626
KEYWORDS    .
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Arthropoda;
            Tracheata; Hexapoda; Insecta; Pterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 2424)
  AUTHORS   Kadowaki,T., Wilder,E., Klingensmith,J., Zachery,K. and Perrimon,N.
  TITLE     The segment polarity gene porcupine encodes a putative
            multi-transmembrane protein involved in wingless processing
  JOURNAL   Genes Dev. (1996) In press
REFERENCE   2  (bases 1 to 2424)
  AUTHORS   Kadowaki,T., Wilder,E., Klingensmith,J., Zachery,K. and Perrimon,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-NOV-1996) Genetics, Harvard Medical School, 200
            Longwood Avenue, Boston, MA 02115, USA
FEATURES             Location/Qualifiers
     source          1. .2424
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="I"
                     /map="17A5-7"
     CDS             165. .1742
                     /codon_start=1
                     /product="porcupine"
                     /db_xref="PID:g1673627"
                     /translation="MDYQYPEEESDYIDLDEEEEDDDVVTAGSLDHRPGQPNGEEDYY
                     PGGDDVEEELVVDGHGVLELAGRLLESLQSCVQPSVLQVMQYVAPMLLLCLLCRLLCL
                     LYSQRRRLTSIAPLHLPHFACGLYIILQITVGYRLLLLLLLAAVGYLLLQLLRLGRRGA
                     QVLAVLTVGSQPLYELLIWRRRSDWPQLRGIQMVVNMKLISLGFDLTASGQLQARIPG
                     PFAYLGYIYSPATCALGPWVSFGCYMDCLVPRNSWLVSLRRLLPNVVICVLAVTVSNC
                     VAPALSDFFGDSSHFLVMYWDALSVRSSHYFVGMMAQALLVASDQRLDGATKESDMLG
                     PLISQPWRIEWPRSISSLVRSWNIPMHEWLKRYIYAPCKPTASTSRGRILVVVLSTYL
```



```
                     VSSLLHGMDLRIYLVLISLAFLAEGESLLRRQLASLLNACITANLCPGKERCRYSHCP
                     SKRRLNSLSTWLVRLTNLAFTALAIFHLAYLGVVLLGDDLEVGEDGDSFLWHWQQAGY
                     LSHYIGLGTFVLYLFIS"
BASE COUNT      460 a    636 c    647 g    681 t
ORIGIN

  Query Match           6.1%;  Score 18;  DB 18;  Length 2424;
  Best Local Similarity 100.0%;  Pred. No. 2.36e+01;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2372 TAAAAATATTTTTATTTA 2389
          ||||||||||||||||||
Qy    103 TAAAAATATTTTTATTTA 120


RESULT   49
LOCUS       HUMSCP2A     2572 bp    mRNA            PRI       06-DEC-1993
DEFINITION  Human sterol carrier protein X/sterol carrier protein 2 mRNA,
            complete cds.
ACCESSION   M75883
NID         g432974
KEYWORDS    sterol carrier protein-2, sterol carrier protein X.
SOURCE      Human liver cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2572)
  AUTHORS   He,Z., Yamamoto,R., Furth,E.A., Schantz,L.J., Naylor,S.L.,
            George,H., Billheimer,J.T. and Strauss,J.F.III.
  TITLE     cDNAs encoding members of a family of proteins related to human
            sterol carrier protein 2 and assignment of the gene to human
            chromosome 1p21-pter
  JOURNAL   DNA Cell Biol. 10, 559-569 (1991)
  MEDLINE   92029618
REFERENCE   2  (bases 1 to 2572)
  AUTHORS   Vesa,J., Hellsten,E., Branoski,B.L., Emanuel,B.S., Billheimer,J.T.,
            Mead,S., Cowell,J.K., Strauss,J.F.III. and Peltonen,L.
  TITLE     Assignment of sterol carrier protein X/sterol carrier protein 2 to
            1p32 and exclusion as the causative gene for infantile neuronal
            cesoid lipofusionosis
  JOURNAL   Unpublished
COMMENT     On Dec 6, 1993 this sequence version replaced gi:410029.
FEATURES             Location/Qualifiers
     source          1. .2572
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="liver"
                     /clone_lib="Clontech HL-1001b;HL-1115b"
                     /map="1p32"
     gene            22. .1665
                     /gene="SCP-X/SCP-2"
```



```
     CDS             22. .1665
                     /gene="SCP-X/SCP-2"
                     /codon_start=1
                     /product="sterol carrier protein-X/sterol carrier
                     protein-2"
                     /db_xref="PID:g432975"
                     /translation="MSSSPWEPATLRRVPVVGVGMTKFVKPGAENSRDYPDLAEEAGK
                     KALADAQIPYSAVDQACVGYVFGDSTCGQRAIYHSLGMTGIPIINVNNNCATGSTALF
                     MARQLIQGGVAECVLALGFEKMSKGSLGIKFSDRTIPTDKHVDLLINKYGLSAHPVAP
                     QMFGYAGKEHMEKYGTKIEHPAKIGWKNHKHSVNNPYSQFQDEYSLDEVMASKEVFDF
                     LTILQCCPTSDGAAAAILASEAFVQKYGLQSKAVEILAQEMMTDLPSSFEEKSIIKMV
                     GFDMSKEAAARKCYEKSGLTPNDIDVIELHDCPSTNELLTYEALGLCPEGQGATLVDRG
                     DNTYGGKWVINPSGGLISKGHPLGATGLAQCAELCWQLRGEAGKRQVPGARVALQHNL
                     GIGGAVVVTLTKMGPPEAASSPRTHQIEAVPTSSASDGFKANLVFKEIEKKLEEEGEQ
                     FVKKIGGIFAFKVKDGPGKEATWVVDVKNGKGSVLPNSDKKADCTITMADSDFLALM
                     TGKMNPQSAFFQGKLKITGNMGLAMKLQNLQLQPGNAKL"
     polyA_signal    1932. .1937
                     /evidence=not_experimental
     polyA_signal    2213. .2218
                     /evidence=not_experimental
     polyA_signal    2258. .2263
                     /evidence=not_experimental
     polyA_signal    2341. .2346
                     /evidence=experimental
     polyA_signal    2550. .2555
                     /evidence=experimental
     polyA_site      2572
                     /evidence=experimental
BASE COUNT      794 a    439 c    570 g    769 t
ORIGIN

  Query Match           6.1%;  Score 18;  DB 25;  Length 2572;
  Best Local Similarity 100.0%;  Pred. No. 2.36e+01;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2257 AAATAAAAATATTTTTAA 2274
          ||||||||||||||||||
Cp    119 AAATAAAAATATTTTTAA 102


RESULT   50
LOCUS       HSL81395     2613 bp    DNA             PRI       09-APR-1997
DEFINITION  Homo sapiens (subclone 2_d6 from P1 H39) DNA sequence, complete
            sequence.
ACCESSION   L81395
NID         g1930407
KEYWORDS    HTG.
SOURCE      Homo sapiens (Subclones in pOT2 from P1 clone H39) DNA.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
```

```
            Homo.
REFERENCE   1  (bases 1 to 2613)
  AUTHORS   Martin,C.H., Arcaina,T.T., Bondoc,M.M., Chiang,A., Critz,P.A.,
            Davis,C.A., Doyle,C.M., Ericsson,C.L., Farfan,D.E., Gunning,K.M.,
            Houston,K.A., Jaklevic,M.A., Kadner,K.E., Kim,K., Kim,S.F.,
            Ko,C.L., Lewis,K.D., Li,M., Lindquist,K.J., Lomotan,M.A.,
            Lustre,V.M., Machrus,M.U., Mayeda,C.A., Miguel,T.M., Miller,C.A.,
            Mok,M.S., Pacleb,J.M., Patel,S.G., Santos,R.F., Subramanian,S.,
            Wan,K.H., Whitelaw,K.R., Yee,A., Yeh,R.T., Yu,C. and Palazzolo,M.J.
  TITLE     Sequencing of human chromosome 5q
  JOURNAL   Unpublished (1996)
REFERENCE   2  (bases 1 to 2613)
  AUTHORS   Martin,C.H., Arcaina,T.T., Bondoc,M.M., Chiang,A., Critz,P.A.,
            Davis,C.A., Doyle,C.M., Ericsson,C.L., Farfan,D.E., Gunning,K.M.,
            Houston,K.A., Jaklevic,M.A., Kadner,K.E., Kim,K., Kim,S.F.,
            Ko,C.L., Lewis,K.D., Li,M., Lindquist,K.J., Lomotan,M.A.,
            Lustre,V.M., Machrus,M.U., Mayeda,C.A., Miguel,T.M., Miller,C.A.,
            Mok,M.S., Pacleb,J.M., Patel,S.G., Santos,R.F., Subramanian,S.,
            Wan,K.H., Whitelaw,K.R., Yee,A., Yeh,R.T., Yu,C. and Palazzolo,M.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-APR-1997)
COMMENT     Sequence submitted by:
            Human Genome Center
            Lawrence Berkeley National Laboratory, MS 74-157
            Berkeley, CA 94720
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Website (http://www-hgc.lbl.gov/sequence-archive.html) or
            send email to human@genome.lbl.gov.
FEATURES             Location/Qualifiers
     source          1. .2613
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="P1 H39 (5236)"
                     /sub_clone="2_d6"
BASE COUNT      896 a    468 c    505 g    744 t
ORIGIN

  Query Match           6.1%;  Score 18;  DB 26;  Length 2613;
  Best Local Similarity 100.0%;  Pred. No. 2.36e+01;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1699 TTAAATAAAAATATTTTT 1716
          ||||||||||||||||||
Cp    121 TTAAATAAAAATATTTTT 104


Search completed: Tue Dec 22 19:35:52 1998
Job time : 733 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Tue Dec 22 19:57:17 1998; MasPar time 237.38 Seconds
1108.948 Million cell updates/sec

Tabular output not generated.

Title: >US-08-530-112A-709
Description: (1-147) from US08530112A.seq
Perfect Score: 147
N.A. Sequence: 1 GATCCCCAGGGTTTCTTTGT.....TTAAAATAATAATGTACAAA 147
Comp: CTAGGGGTCCCAAAGAAACA.....AATTTTATTATTACATGTTT

Scoring table: TABLE jmetric
Gap 60

Nmatch STD : Dbase 0; Query 0

Searched: 2275026 seqs, 895388244 bases x 2

Post-processing: Minimum Match 0%
Listing first 1000 summaries

Database: embl-est55
1:em_est1 2:em_gss1 3:em_gss2 4:em_gss3
Database: genbank-est107
5:gb_est1 6:gb_est10 7:gb_est11 8:gb_est12 9:gb_est13
10:gb_est14 11:gb_est15 12:gb_est16 13:gb_est17
14:gb_est18 15:gb_est19 16:gb_est2 17:gb_est20
18:gb_est21 19:gb_est3 20:gb_est4 21:gb_est5 22:gb_est6
23:gb_est7 24:gb_est8 25:gb_est9 26:gb_gss1 27:gb_gss2
28:gb_gss3 29:gb_gss4

Statistics: Mean 7.871; Variance 1.262; scale 6.237

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
	1	147	100.0	147	7	HUMGS00783	Human HL60 3'directed	0.00e+00
c	2	147	100.0	349	15	AI018794	ov32g09.x1 Soares_test	0.00e+00
c	3	147	100.0	478	17	AI017469	ou23c08.x1 Soares_NFL_	0.00e+00
	4	146	99.3	385	24	AA303166	EST13228 Thymus tumor	0.00e+00
c	5	85	57.8	399	14	AA909023	ol11f10.s1 Soares_NFL_	1.69e-154
	6	83	56.5	390	16	H41523	ym90g09.r1 Homo sapien	1.41e-149
c	7	82	55.8	313	15	AA972448	op41b02.s1 Soares_NFL_	4.02e-147
c	8	73	49.7	273	9	AA557377	nl81f07.s1 NCI_CGAP_Br	3.64e-125
c	9	73	49.7	318	14	AA923630	nq98c09.s1 NCI_CGAP_Co	3.64e-125
c	10	72	49.0	232	13	AA862961	og98f02.s1 NCI_CGAP_Ki	9.60e-123
c	11	32	21.8	287	16	R59962	yh11g12.s1 Homo sapien	2.92e-31
	12	20	13.6	212	7	HUM293B06B	Human aorta cDNA 5'-en	1.75e-08
	13	20	13.6	310	23	AA122818	mr03b11.r1 Soares mous	1.75e-08
c	14	20	13.6	360	14	AA924990	UI-R-A1-eh-b-12-0-UI.s	1.75e-08
c	15	20	13.6	369	6	AA399462	zt53c08.s1 Soares ovar	1.75e-08
c	16	20	13.6	380	12	AA721239	ns72c05.s1 NCI_CGAP_GC	1.75e-08
	17	20	13.6	382	10	AA415181	vc98e08.s1 Knowles Sol	1.75e-08
	18	20	13.6	432	23	AA119227	mp77b12.r1 Soares 2NbM	1.75e-08
	19	20	13.6	447	21	W52272	zc46d11.r1 Soares sene	1.75e-08
	20	20	13.6	457	21	AA048254	mj27e04.r1 Soares mous	1.75e-08
	21	20	13.6	476	15	AI050617	ub38b11.r1 Soares 2NbM	1.75e-08
	22	20	13.6	491	21	AA048095	mj23d09.r1 Soares mous	1.75e-08
	23	20	13.6	493	23	AA185438	mt61c06.r1 Soares 2NbM	1.75e-08
c	24	19	12.9	335	17	AI083687	oy71g10.x1 NCI_CGAP_CL	7.84e-07
	25	19	12.9	716	27	B95458	CIT-HSP-2166816.TF CIT	7.84e-07
c	26	18	12.2	337	15	AA895256	os18e06.s1 NCI_CGAP_Ki	3.09e-05
	27	18	12.2	363	27	AQ027251	CIT-HSP-2314L21.TF CIT	3.09e-05
	28	18	12.2	400	6	CELK028B9R	C.elegans cDNA clone y	3.09e-05
	29	18	12.2	470	10	AA595195	no32f09.s1 NCI_CGAP_Pr	3.09e-05
c	30	18	12.2	622	26	B00235	cSRL-104g4-u cSRL flow	3.09e-05
c	31	17	11.6	68	15	AA999800	os41c07.s1 NCI_CGAP_Br	1.06e-03
c	32	17	11.6	150	13	AA224351	zr12f12.s1 Stratagene	1.06e-03
	33	17	11.6	191	16	M79056	EST01204 Homo sapiens	1.06e-03
	34	17	11.6	197	19	N54751	yz16g11.r1 Homo sapien	1.06e-03
	35	17	11.6	212	15	AU006635	Schizosaccharomyces po	1.06e-03
	36	17	11.6	245	19	HSC27P062	H. sapiens partial cDN	1.06e-03
c	37	17	11.6	259	6	AA100296	zl81g06.s1 Stratagene	1.06e-03
c	38	17	11.6	279	19	N66308	yz37h08.s1 Homo sapien	1.06e-03
	39	17	11.6	306	16	H25898	yl55b10.r1 Homo sapien	1.06e-03
c	40	17	11.6	325	16	T33015	EST56321 Homo sapiens	1.06e-03
c	41	17	11.6	340	5	T09911	0461m3 Plasmodium falc	1.06e-03
c	42	17	11.6	346	21	T16288	NIB1062 Normalized inf	1.06e-03

	43	17	11.6	351	16	H16570	ym26b04.s1 Homo sapien	1.06e-03
c	44	17	11.6	367	15	AA928637	om75e03.s1 NCI_CGAP_GC	1.06e-03
	45	17	11.6	375	23	AA219483	zq17e10.r1 Stratagene	1.06e-03
c	46	17	11.6	377	13	AA810850	oa76f06.s1 NCI_CGAP_GC	1.06e-03
	47	17	11.6	407	9	AA595015	nj81a01.s1 NCI_CGAP_Pr	1.06e-03
c	48	17	11.6	410	11	AA716223	sg60e07.s1 Soares feta	1.06e-03
c	49	17	11.6	417	27	B81278	CIT-HSP-2023F15.TF3B CI	1.06e-03
c	50	17	11.6	418	22	AA128639	zo04c03.s1 Stratagene	1.06e-03
	51	17	11.6	419	8	AA255963	zs27g12.r1 NCI_CGAP_GC	1.06e-03
c	52	17	11.6	419	18	AI143376	qa51c10.s1 Soares_NhHM	1.06e-03
c	53	17	11.6	426	26	B96988	T33O1TR TAMU Arabidops	1.06e-03
c	54	17	11.6	428	5	R43906	yg26h05.s1 Homo sapien	1.06e-03
	55	17	11.6	429	9	AA531554	nj64d04.s1 NCI_CGAP_Pr	1.06e-03
c	56	17	11.6	437	6	AA004937	zh87c11.s1 Soares feta	1.06e-03
	57	17	11.6	438	8	AA562189	vl22d10.r1 Stratagene	1.06e-03
c	58	17	11.6	438	28	AQ175789	HS_3215_A2_E01_MR CIT	1.06e-03
c	59	17	11.6	440	18	AI185526	qe52g11.x1 Soares_feta	1.06e-03
c	60	17	11.6	442	19	N64554	yx50d09.s1 Homo sapien	1.06e-03
c	61	17	11.6	447	18	AI185525	qe52g10.x1 Soares_feta	1.06e-03
c	62	17	11.6	447	8	AA262171	zs25c10.r1 NCI_CGAP_GC	1.06e-03
c	63	17	11.6	448	21	W48576	zc42f07.r1 Soares sene	1.06e-03
c	64	17	11.6	453	14	AA073134	oh60g07.s1 NCI_CGAP_Ki	1.06e-03
c	65	17	11.6	459	22	AA173052	zp05g07.r1 Stratagene	1.06e-03
	66	17	11.6	465	29	AQ227792	HS_2012_B2_H11_MR CIT	1.06e-03
c	67	17	11.6	468	22	W94906	ze14c02.s1 Soares feta	1.06e-03
c	68	17	11.6	470	15	AA995865	oa83f09.s1 NCI_CGAP_GC	1.06e-03
c	69	17	11.6	470	20	T15752	IB1836 Infant brain, B	1.06e-03
c	70	17	11.6	470	10	AA632039	np62g01.s1 NCI_CGAP_Br	1.06e-03
c	71	17	11.6	475	17	AA993662	ot85h03.s1 Soares_tota	1.06e-03
c	72	17	11.6	480	28	AQ208711	HS_3242_A2_G09_T7 CIT	1.06e-03
c	73	17	11.6	481	9	AA548784	nj13h07.s1 NCI_CGAP_Pr	1.06e-03
c	74	17	11.6	482	10	AA595228	nl84f10.s1 NCI_CGAP_Br	1.06e-03
c	75	17	11.6	483	16	H23117	ym51g12.s1 Homo sapien	1.06e-03
	76	17	11.6	501	12	AA748765	ny10e04.s1 NCI_CGAP_GC	1.06e-03
c	77	17	11.6	511	29	AQ140657	HS_3112_A2_B12_T7 CIT	1.06e-03
c	78	17	11.6	527	12	AA748460	oa56g09.s1 NCI_CGAP_GC	1.06e-03
	79	17	11.6	531	15	AA905049	ol13b01.s1 Soares_NFL_	1.06e-03
c	80	17	11.6	533	11	AA668875	aa80f10.s1 NCI_CGAP_GC	1.06e-03
	81	17	11.6	593	17	T46724	9987 Lambda-PRL2 Arabi	1.06e-03
	82	17	11.6	604	26	FR0022351	P.rubripes GSS sequenc	1.06e-03
c	83	17	11.6	610	12	AA745076	np73h10.s1 NCI_CGAP_Br	1.06e-03
	84	17	11.6	621	28	AQ208298	HS_3219_A1_F09_T7 CIT	1.06e-03
	85	17	11.6	648	26	AG008472	Homo sapiens genomic D	1.06e-03
c	86	16	10.9	165	28	AQ110174	CIT-HSP-2381I1.TR CIT-	3.07e-02
c	87	16	10.9	169	24	AA321120	EST23570 Adipose tissu	3.07e-02
c	88	16	10.9	170	15	AA928527	om73e08.s1 NCI_CGAP_GC	3.07e-02
	89	16	10.9	233	19	H75193	71 Pinus taeda cDNA cl	3.07e-02
c	90	16	10.9	236	6	AA294817	JA00A126.QA3 Adult SjC	3.07e-02
	91	16	10.9	240	25	AA377770	EST90387 Synovial sarc	3.07e-02
	92	16	10.9	247	24	AA326431	EST29631 Cerebellum II	3.07e-02
	93	16	10.9	256	15	C89689	Dictyostelium discoide	3.07e-02

	94	16	10.9	274	12	AA759142	ah78f08.s1 Soares test	3.07e-02
	95	16	10.9	277	14	H34358	EST111246 Rat PC-12 ce	3.07e-02
c	96	16	10.9	285	26	FR0015766	F.rubripes GSS sequenc	3.07e-02
c	97	16	10.9	291	11	AA707652	zj29e04.s1 Soares feta	3.07e-02
c	98	16	10.9	291	18	AI144893	UI-R-BT0-po-g-12-0-UI.	3.07e-02
	99	16	10.9	300	9	C29514	C.elegans cDNA clone y	3.07e-02
	100	16	10.9	300	6	C08346	C.elegans cDNA clone y	3.07e-02
	101	16	10.9	300	9	C54455	C.elegans cDNA clone y	3.07e-02
	102	16	10.9	300	9	C54693	C.elegans cDNA clone y	3.07e-02
c	103	16	10.9	300	9	C57076	C.elegans cDNA clone y	3.07e-02
c	104	16	10.9	300	9	C58746	C.elegans cDNA clone y	3.07e-02
	105	16	10.9	307	25	AA363285	EST73387 Ovary, subtra	3.07e-02
c	106	16	10.9	308	10	AA629037	zu77c11.s1 Soares test	3.07e-02
c	107	16	10.9	312	14	AA864495	oh38g08.s1 NCI_CGAP_Ki	3.07e-02
	108	16	10.9	316	17	AI072015	UI-R-C2-nd-d-06-0-UI.s	3.07e-02
c	109	16	10.9	317	22	AA035609	ze75b12.s1 Soares feta	3.07e-02
c	110	16	10.9	320	16	H16195	yl28b12.s1 Homo sapien	3.07e-02
c	111	16	10.9	321	17	AI126995	qb96f09.x1 Soares_feta	3.07e-02
c	112	16	10.9	322	20	HSC3ID062	H. sapiens partial cDN	3.07e-02
	113	16	10.9	323	16	H40963	yo05g01.r1 Homo sapien	3.07e-02
	114	16	10.9	324	24	AA324457	EST27282 Cerebellum II	3.07e-02
c	115	16	10.9	329	25	AA350771	EST58210 Infant brain	3.07e-02
c	116	16	10.9	330	19	N25076	yw40g05.s1 Homo sapien	3.07e-02
	117	16	10.9	333	27	AQ072212	HS_3024_B1_C04_T7 CIT	3.07e-02
c	118	16	10.9	333	15	AI023117	ow53d03.s1 Soares_para	3.07e-02
c	119	16	10.9	346	16	H25896	yl55b09.r1 Homo sapien	3.07e-02
c	120	16	10.9	348	5	T96610	ye51d05.s1 Homo sapien	3.07e-02
	121	16	10.9	350	11	AA675140	vq53h10.r1 Barstead mo	3.07e-02
	122	16	10.9	352	24	AA324265	EST27058 Cerebellum II	3.07e-02
c	123	16	10.9	360	9	C64219	C.elegans cDNA clone y	3.07e-02
	124	16	10.9	363	27	AQ030247	HS_2200_B1_H07_MF CIT	3.07e-02
	125	16	10.9	365	17	AA407332	EST00652 Mouse 7.5 dpc	3.07e-02
	126	16	10.9	365	5	T96632	ye51g07.s1 Homo sapien	3.07e-02
	127	16	10.9	367	27	B84676	RPCI11-27P4.TV RPCI11	3.07e-02
c	128	16	10.9	369	21	H49090	yq18g03.s1 Soares feta	3.07e-02
	129	16	10.9	373	24	AA301644	EST14672 Aorta endothe	3.07e-02
c	130	16	10.9	373	9	C63191	C.elegans cDNA clone y	3.07e-02
	131	16	10.9	385	9	AA575055	vm34g06.r1 Knowles Sol	3.07e-02
c	132	16	10.9	387	21	W31632	zb97d07.s1 Soares para	3.07e-02
c	133	16	10.9	387	20	W04651	zb93f06.s1 Soares para	3.07e-02
	134	16	10.9	388	28	AQ205787	HS_3245_A1_E02_MR CIT	3.07e-02
c	135	16	10.9	390	29	AQ143494	HS_3066_B2_H03_MR CIT	3.07e-02
c	136	16	10.9	393	8	AA234894	zs36b11.s1 Soares_NhHM	3.07e-02
	137	16	10.9	394	13	AA884851	am28f10.s1 Soares NFL	3.07e-02
c	138	16	10.9	394	7	AA512069	vj43e01.r1 Stratagene	3.07e-02
c	139	16	10.9	400	20	H77767	yz83d02.r1 Homo sapien	3.07e-02
c	140	16	10.9	401	8	AA243498	zr67d06.s1 Soares_NhHM	3.07e-02
c	141	16	10.9	401	29	AQ135651	HS_3053_B2_G04_MR CIT	3.07e-02
	142	16	10.9	407	19	N42279	yw85h01.r1 Homo sapien	3.07e-02
	143	16	10.9	409	16	H50034	yo26d07.r1 Homo sapien	3.07e-02
c	144	16	10.9	409	23	W86844	zh59c01.s1 Soares feta	3.07e-02

	145	16	10.9	413	26	B77228	T31N9TR TAMU Arabidops 3.07e-02
	146	16	10.9	415	27	AQ034271	CIT-HSP-2329I22.TR CIT 3.07e-02
c	147	16	10.9	418	6	AA009775	zi04d08.s1 Soares feta 3.07e-02
	148	16	10.9	419	8	AA102098	z179f12.s1 Stratagene 3.07e-02
c	149	16	10.9	422	8	AA282261	zt13g02.s1 NCI_CGAP_GC 3.07e-02
c	150	16	10.9	424	21	N92293	zb60f02.s1 Soares feta 3.07e-02
c	151	16	10.9	427	7	HUM279C07B	Human aorta cDNA 5'-en 3.07e-02
c	152	16	10.9	431	9	AA577305	nm87h02.s1 NCI_CGAP_Co 3.07e-02
	153	16	10.9	435	14	C92007	Dictyostelium discoide 3.07e-02
c	154	16	10.9	435	18	AI142837	ox64a09.s1 Soares_NhHM 3.07e-02
	155	16	10.9	437	28	AQ202465	RPCI11-58C22.TK RPCI11 3.07e-02
c	156	16	10.9	438	17	AI022406	ow72g11.x1 Soares_feta 3.07e-02
	157	16	10.9	439	22	W51760	zc36b12.s1 Soares sene 3.07e-02
	158	16	10.9	441	22	N59564	yv76g06.s1 Soares feta 3.07e-02
c	159	16	10.9	443	27	AQ038647	CIT-HSP-2337D5.TR CIT- 3.07e-02
	160	16	10.9	444	17	AI067601	EST209283 Schistosoma 3.07e-02
	161	16	10.9	446	6	AA403021	zt62a02.r1 Soares test 3.07e-02
	162	16	10.9	449	23	AA198890	mu05g10.r1 Soares mous 3.07e-02
	163	16	10.9	451	8	AA235165	zs36b11.r1 Soares NhHM 3.07e-02
	164	16	10.9	452	5	T47785	yb17a10.r1 Homo sapien 3.07e-02
	165	16	10.9	454	21	W48836	zc42h02.s1 Soares sene 3.07e-02
c	166	16	10.9	458	16	H49129	yo21f07.r1 Homo sapien 3.07e-02
	167	16	10.9	461	17	AI118678	uc12f03.r1 Soares mous 3.07e-02
	168	16	10.9	461	26	B49675	CIT-HSP-263P3.TP CIT-H 3.07e-02
	169	16	10.9	461	8	AA282260	zt13g02.r1 NCI_CGAP_GC 3.07e-02
c	170	16	10.9	468	27	AQ058534	CIT-HSP-2343E6.TF CIT- 3.07e-02
c	171	16	10.9	468	19	N54158	yv63h09.s1 Homo sapien 3.07e-02
c	172	16	10.9	469	10	AA628086	ns67d08.s1 NCI_CGAP_Pr 3.07e-02
c	173	16	10.9	470	18	AI095309	oy20f08.s1 Soares_sene 3.07e-02
c	174	16	10.9	471	6	W90632	zh78h07.s1 Soares feta 3.07e-02
	175	16	10.9	474	14	C92329	Dictyostelium discoide 3.07e-02
	176	16	10.9	479	21	W44678	zc29g09.s1 Soares sene 3.07e-02
	177	16	10.9	480	23	AA138125	mr05g10.r1 Soares mous 3.07e-02
	178	16	10.9	480	23	AA244694	mx07h05.r1 Soares mous 3.07e-02
	179	16	10.9	492	10	AA651753	ns67d08.r1 NCI_CGAP_Pr 3.07e-02
c	180	16	10.9	495	14	AA864988	oh01c02.s1 NCI_CGAP_Ki 3.07e-02
	181	16	10.9	498	19	N51454	yr04e01.s1 Homo sapien 3.07e-02
	182	16	10.9	501	23	AA198995	mu05c10.r1 Soares mous 3.07e-02
	183	16	10.9	502	29	AQ141835	HS_3154_B1_H10_T7 CIT 3.07e-02
c	184	16	10.9	502	17	AI038062	ox21b06.x1 Soares_feta 3.07e-02
c	185	16	10.9	505	16	H50480	yo25c11.r1 Homo sapien 3.07e-02
c	186	16	10.9	505	12	AA729672	nx37c03.s1 NCI_CGAP_GC 3.07e-02
c	187	16	10.9	530	8	AA527678	nh86c11.s1 NCI_CGAP_Br 3.07e-02
	188	16	10.9	535	23	AA106432	ml94c03.r1 Stratagene 3.07e-02
	189	16	10.9	537	14	C90157	Dictyostelium discoide 3.07e-02
	190	16	10.9	537	11	C83907	Dictyostelium discoide 3.07e-02
	191	16	10.9	538	18	AI171355	EST217310 Normalized r 3.07e-02
c	192	16	10.9	543	17	AI074463	oy96f10.x1 Soares_feta 3.07e-02
c	193	16	10.9	544	5	R18177	yf98b09.r1 Homo sapien 3.07e-02
c	194	16	10.9	545	27	AQ002534	CIT-HSP-2283I16.TF CIT 3.07e-02
c	195	16	10.9	546	28	AQ182572	HS_3084_A1_C10_MF CIT 3.07e-02

c	196	16	10.9	557	8	AA292838	zt64d01.r1 Soares test 3.07e-02
c	197	16	10.9	559	19	N46851	yy73d04.s1 Homo sapien 3.07e-02
	198	16	10.9	560	23	AA204477	mu32e11.r1 Soares 2NbM 3.07e-02
	199	16	10.9	569	15	C93960	Dictyostelium discoide 3.07e-02
	200	16	10.9	571	14	C89716	Dictyostelium discoide 3.07e-02
	201	16	10.9	583	26	AQ023696	CpGR0110A Cryptosporid 3.07e-02
c	202	16	10.9	586	17	AI131650	ui41b01.x1 Sugano mous 3.07e-02
	203	16	10.9	592	14	C92605	Dictyostelium discoide 3.07e-02
c	204	16	10.9	592	6	AA135749	zl50e06.s1 Soares preg 3.07e-02
c	205	16	10.9	598	11	AA651323	31437 Lambda-PRL2 Arab 3.07e-02
	206	16	10.9	603	12	AA794906	vr48e06.s1 Knowles Sol 3.07e-02
	207	16	10.9	609	14	C93421	Dictyostelium discoide 3.07e-02
c	208	16	10.9	636	28	AQ109709	CIT-HSP-237209.TF CIT- 3.07e-02
c	209	16	10.9	656	21	W84446	zd89f03.s1 Soares feta 3.07e-02
	210	16	10.9	658	14	C92254	Dictyostelium discoide 3.07e-02
	211	16	10.9	674	27	AQ083616	CpG0501B CpIOWAgDNA1 C 3.07e-02
	212	16	10.9	692	14	C92013	Dictyostelium discoide 3.07e-02
	213	16	10.9	713	14	C90956	Dictyostelium discoide 3.07e-02
	214	16	10.9	880	27	AQ026352	l(3)neo63 Drosophila m 3.07e-02
c	215	16	10.9	941	26	B09148	T17M6-T7 TAMU Arabidop 3.07e-02
c	216	16	10.9	1187	26	B11102	F19C22-T7 IGF Arabidop 3.07e-02
c	217	16	10.9	1248	26	B11336	F19M10-Sp6 IGF Arabido 3.07e-02
c	218	15	10.2	91	10	AA638880	vn75c08.r1 Knowles Sol 7.49e-01
	219	15	10.2	105	19	N32628	yw95h02.s1 Homo sapien 7.49e-01
c	220	15	10.2	111	23	AA221242	mw11c04.r1 Soares mous 7.49e-01
c	221	15	10.2	112	23	AA174653	mt11h12.r1 Soares mous 7.49e-01
	222	15	10.2	116	14	AA892243	EST196046 Normalized r 7.49e-01
c	223	15	10.2	116	11	AA675122	vq52g01.r1 Barstead mo 7.49e-01
c	224	15	10.2	118	10	AA612536	v170g06.r1 Knowles Sol 7.49e-01
c	225	15	10.2	120	23	AA176038	mt04f08.r1 Soares mous 7.49e-01
c	226	15	10.2	121	15	AI021130	ub01f09.r1 Soares mous 7.49e-01
c	227	15	10.2	121	9	AA600625	vm76g01.r1 Knowles Sol 7.49e-01
c	228	15	10.2	127	8	AA230243	nc13d04.s1 NCI_CGAP_Pr 7.49e-01
c	229	15	10.2	133	23	AA231561	mw31h01.r1 Soares mous 7.49e-01
c	230	15	10.2	139	19	H55103	CHR220042 Homo sapiens 7.49e-01
	231	15	10.2	147	15	AI008266	EST202717 Normalized r 7.49e-01
	232	15	10.2	153	14	AA848777	EST191538 Normalized r 7.49e-01
c	233	15	10.2	153	9	AA596312	vo26c11.r1 Barstead mo 7.49e-01
c	234	15	10.2	153	23	AA245407	my52g07.r1 Barstead mo 7.49e-01
c	235	15	10.2	162	14	AA930167	va71a04.r1 Stratagene 7.49e-01
c	236	15	10.2	163	22	AA027498	mi04d09.r1 Soares mous 7.49e-01
c	237	15	10.2	164	27	AQ009429	CIT-HSP-2283N4.TRB CIT 7.49e-01
c	238	15	10.2	166	12	AA795489	vq95e11.r1 Knowles Sol 7.49e-01
c	239	15	10.2	172	8	AA262607	zs23f09.s1 NCI_CGAP_GC 7.49e-01
	240	15	10.2	177	14	AA894117	EST197920 Normalized r 7.49e-01
	241	15	10.2	180	9	C51606	C.elegans cDNA clone y 7.49e-01
	242	15	10.2	188	9	C75454	Homo sapiens mRNA expr 7.49e-01
c	243	15	10.2	192	20	W72882	zd62f09.s1 Soares feta 7.49e-01
	244	15	10.2	193	25	AA361474	EST70758 T-cell lympho 7.49e-01
	245	15	10.2	194	14	AA957694	UI-R-E1-fw-f-12-0-UI.s 7.49e-01
	246	15	10.2	197	19	N64123	yv60h04.r1 Homo sapien 7.49e-01

c	247	15	10.2	199	5	R38393	yc96e10.s1 Homo sapien 7.49e-01
	248	15	10.2	201	10	AA582467	nn54f02.s1 NCI_CGAP_Ki 7.49e-01
	249	15	10.2	201	26	PR0019435	F.rubripes GSS sequenc 7.49e-01
	250	15	10.2	202	19	N56236	J9441F Homo sapiens cD 7.49e-01
	251	15	10.2	203	17	AA908549	og83h02.s1 NCI_CGAP_Ov 7.49e-01
c	252	15	10.2	204	10	AA580076	zu62d01.s1 Soares test 7.49e-01
c	253	15	10.2	205	21	W57855	zd21b07.s1 Soares feta 7.49e-01
	254	15	10.2	206	17	AA908546	og83g02.s1 NCI_CGAP_Ov 7.49e-01
c	255	15	10.2	208	5	R40459	yf78e04.s1 Homo sapien 7.49e-01
	256	15	10.2	209	22	N52237	yv46f12.s1 Soares feta 7.49e-01
	257	15	10.2	211	18	AI174295	am67c07.s1 Barstead sp 7.49e-01
	258	15	10.2	211	22	AA185003	mt88d11.r1 Soares mous 7.49e-01
c	259	15	10.2	214	15	AA926693	om24g02.s1 Soares_NFL_ 7.49e-01
	260	15	10.2	215	9	AA573881	nk08g12.s1 NCI_CGAP_Co 7.49e-01
c	261	15	10.2	217	13	AA804440	ns38b11.s1 NCI_CGAP_GC 7.49e-01
	262	15	10.2	221	13	AA859721	UI-R-E0-bx-h-08-0-UI.s 7.49e-01
c	263	15	10.2	221	23	AA244831	mx25f11.r1 Soares mous 7.49e-01
	264	15	10.2	225	12	AA742502	nx20h10.s1 NCI_CGAP_GC 7.49e-01
c	265	15	10.2	225	15	AA908921	ol96b03.s1 NCI_CGAP_PN 7.49e-01
c	266	15	10.2	226	10	AA624820	vn79f11.r1 Knowles Sol 7.49e-01
	267	15	10.2	227	6	C05790	similar to profilin I, 7.49e-01
	268	15	10.2	228	17	AI095528	qb29c08.x1 Soares_preg 7.49e-01
	269	15	10.2	231	16	H47401	yp74c03.r1 Homo sapien 7.49e-01
	270	15	10.2	233	12	AA776076	ae79c06.s1 Stratagene 7.49e-01
c	271	15	10.2	233	16	R46838	yg54h01.s1 Homo sapien 7.49e-01
	272	15	10.2	234	6	AA389856	JA00A128.QA3 Adult 5jC 7.49e-01
c	273	15	10.2	234	9	AA590859	vn60f03.r1 Barstead mo 7.49e-01
c	274	15	10.2	235	23	AA254723	mz77h10.r1 Soares mous 7.49e-01
c	275	15	10.2	236	28	AQ104630	HS_3047_B2_F01_MR CIT 7.49e-01
	276	15	10.2	237	9	AA563950	nk23d08.s1 NCI_CGAP_Co 7.49e-01
	277	15	10.2	238	25	AA376272	EST88669 HSC172 cells 7.49e-01
	278	15	10.2	239	19	N74268	za77e01.s1 Homo sapien 7.49e-01
c	279	15	10.2	240	23	AA222610	mv70f11.r1 Soares mous 7.49e-01
	280	15	10.2	241	10	AA586712	nn67b01.s1 NCI_CGAP_La 7.49e-01
	281	15	10.2	244	15	AI015285	ov55e01.s1 Soares_test 7.49e-01
	282	15	10.2	247	15	AA995291	os27d05.s1 NCI_CGAP_Ki 7.49e-01
c	283	15	10.2	248	10	AA667218	vq89g08.s1 Knowles Sol 7.49e-01
c	284	15	10.2	249	26	PR0020841	F.rubripes GSS sequenc 7.49e-01
c	285	15	10.2	253	12	AA796472	vp26c06.r1 Barstead mo 7.49e-01
	286	15	10.2	257	7	AA443916	zv46h01.s1 Soares ovar 7.49e-01
	287	15	10.2	257	8	AA230300	nc13d04.r1 NCI_CGAP_Pr 7.49e-01
	288	15	10.2	259	22	AA071506	ze88c03.s1 Soares feta 7.49e-01
	289	15	10.2	263	11	AA650535	nt02e03.s1 NCI_CGAP_Ly 7.49e-01
	290	15	10.2	263	25	AA340561	EST45783 Fetal kidney 7.49e-01
	291	15	10.2	265	6	AA436351	zv32d01.s1 Soares ovar 7.49e-01
c	292	15	10.2	265	19	N59224	yz81b09.s1 Homo sapien 7.49e-01
c	293	15	10.2	266	23	AA216826	mu91c01.r1 Soares mous 7.49e-01
	294	15	10.2	268	6	AA401200	zu52f01.s1 Soares ovar 7.49e-01
c	295	15	10.2	269	13	AA841222	MB3D6AA5B02T3 Brugia m 7.49e-01
c	296	15	10.2	269	19	N63502	yy62c01.s1 Homo sapien 7.49e-01
c	297	15	10.2	270	15	AA972889	op24h02.s1 Soares_NFL_ 7.49e-01

c	298	15	10.2	270	14	AA956441	UI-R-E1-fj-h-06-0-UI.s 7.49e-01
c	299	15	10.2	271	10	AA649036	na38d04.s1 NCI_CGAP_GC 7.49e-01
	300	15	10.2	273	11	AA650545	nt02f03.s1 NCI_CGAP_Ly 7.49e-01
	301	15	10.2	274	23	AA199975	mu01f11.r1 Soares mous 7.49e-01
	302	15	10.2	276	25	AA170877	EST82588 Prostate glan 7.49e-01
c	303	15	10.2	276	21	AA048339	mj27e11.r1 Soares mous 7.49e-01
	304	15	10.2	278	26	PR0009803	F.rubripes GSS sequenc 7.49e-01
c	305	15	10.2	279	14	AA934945	ny33c03.s1 NCI_CGAP_Pr 7.49e-01
c	306	15	10.2	281	12	AA738280	nx13e02.s1 NCI_CGAP_GC 7.49e-01
c	307	15	10.2	282	25	AA356329	EST65118 Jurkat T-cell 7.49e-01
c	308	15	10.2	283	8	AA226685	nc21e08.s1 NCI_CGAP_Pr 7.49e-01
c	309	15	10.2	283	22	AA212273	mu78f04.r1 Stratagene 7.49e-01
c	310	15	10.2	283	22	AA158686	zo78c12.s1 Stratagene 7.49e-01
c	311	15	10.2	284	24	AA289417	vb36e05.r1 Soares mous 7.49e-01
	312	15	10.2	286	8	AA486477	zx91g12.r1 Soares ovar 7.49e-01
c	313	15	10.2	286	7	RICS1619A	Rice cDNA, partial seq 7.49e-01
c	314	15	10.2	289	20	HSCODD051	H. sapiens partial cDN 7.49e-01
	315	15	10.2	290	25	AA336292	EST40862 Epithilioid s 7.49e-01
c	316	15	10.2	291	10	AA435782	zt77d09.s1 Soares test 7.49e-01
c	317	15	10.2	295	19	N25061	yw40e03.s1 Homo sapien 7.49e-01
c	318	15	10.2	296	22	AA191884	rs09e05.r1 Sommer Pris 7.49e-01
	319	15	10.2	297	13	AA852472	NHTBCae13g06r1 Normal 7.49e-01
	320	15	10.2	297	14	AA900159	UI-R-E0-dm-d-08-0-UI.s 7.49e-01
c	321	15	10.2	299	20	T15389	IB1052 Infant brain, B 7.49e-01
	322	15	10.2	300	9	C54938	C.elegans cDNA clone y 7.49e-01
c	323	15	10.2	300	5	R42931	yg14f11.s1 Homo sapien 7.49e-01
c	324	15	10.2	300	8	AA546423	vk64e08.r1 Stratagene 7.49e-01
c	325	15	10.2	301	7	AA470200	vf69d09.r1 Barstead mo 7.49e-01
c	326	15	10.2	302	12	AA664737	nu65a02.s1 NCI_CGAP_Al 7.49e-01
c	327	15	10.2	304	15	AA953909	oo83g05.s1 NCI_CGAP_Ki 7.49e-01
c	328	15	10.2	304	7	HUMGS01154	Human HL60 3'directed 7.49e-01
	329	15	10.2	305	13	AA853839	NHTBCae09c10f1 Normal 7.49e-01
	330	15	10.2	305	27	AQ002154	CIT-HSP-2282E6.TF CIT- 7.49e-01
	331	15	10.2	306	27	AQ029257	RPCI11-370I8.TP RPCI11 7.49e-01
	332	15	10.2	307	5	T94478	ye35d02.r1 Homo sapien 7.49e-01
c	333	15	10.2	307	27	AQ028432	CIT-HSP-231313.TP CIT- 7.49e-01
c	334	15	10.2	307	27	AQ067615	HS_2238_B2_E12_MR CIT 7.49e-01
c	335	15	10.2	307	7	AA529075	vi32c03.r1 Barstead mo 7.49e-01
	336	15	10.2	309	14	AA955429	UI-R-A1-ew-a-12-0-UI.s 7.49e-01
	337	15	10.2	311	5	R47836	yj56f01.r1 Homo sapien 7.49e-01
c	338	15	10.2	311	16	T32421	EST48299 Homo sapiens 7.49e-01
c	339	15	10.2	312	29	AQ236481	HS_2033_B1_E03_MR CIT 7.49e-01
c	340	15	10.2	313	11	AA707501	ah40g01.s1 Soares test 7.49e-01
	341	15	10.2	314	13	T25475	PTHI260 HTCDL1 Homo sa 7.49e-01
c	342	15	10.2	315	21	W84052	mf26a11.r1 Soares mous 7.49e-01
	343	15	10.2	316	25	AA349358	EST56169 Infant brain 7.49e-01
c	344	15	10.2	317	8	AA523775	ni68e04.s1 NCI_CGAP_Pr 7.49e-01
c	345	15	10.2	318	18	AI152181	ud18g12.r1 Soares 2NbM 7.49e-01
c	346	15	10.2	318	9	AA536113	nj81e05.s1 NCI_CGAP_Pr 7.49e-01
c	347	15	10.2	318	15	AI049714	an37b04.x1 Gessler Wil 7.49e-01
	348	15	10.2	320	23	AA152712	ms12g02.r1 Stratagene 7.49e-01

	349	15	10.2	321	24	AA321081	EST23517 Adipose tissu 7.49e-01
c	350	15	10.2	322	22	AA014089	mh28d02.r1 Soares mous 7.49e-01
c	351	15	10.2	326	20	W96347	ze42d11.s1 Soares reti 7.49e-01
c	352	15	10.2	328	23	AA198129	mv13d10.r1 GuayWoodfor 7.49e-01
c	353	15	10.2	328	12	AA764513	vp05b01.r1 Soares mous 7.49e-01
c	354	15	10.2	328	23	AA164097	mr23f07.r1 Soares mous 7.49e-01
c	355	15	10.2	329	13	AA833574	aj46c12.s1 Soares test 7.49e-01
	356	15	10.2	330	21	W36366	mb83a11.r1 Soares mous 7.49e-01
	357	15	10.2	330	18	AI170218	EST216143 Normalized r 7.49e-01
	358	15	10.2	330	9	AA559016	nl11a09.s1 NCI_CGAP_Br 7.49e-01
	359	15	10.2	331	17	AA407236	EST02231 Mouse 7.5 dpc 7.49e-01
	360	15	10.2	331	21	W30476	mc18a11.r1 Soares mous 7.49e-01
	361	15	10.2	332	16	T34080	EST62376 Homo sapiens 7.49e-01
c	362	15	10.2	332	8	AA501449	ne67f02.s1 NCI_CGAP_Al 7.49e-01
c	363	15	10.2	332	24	AA270811	va70g01.r1 Soares mous 7.49e-01
c	364	15	10.2	332	9	AA596262	vo38g02.r1 Barstead mo 7.49e-01
c	365	15	10.2	333	12	AA742963	ny15c12.s1 NCI_CGAP_GC 7.49e-01
c	366	15	10.2	334	5	T52247	yb22c12.r1 Homo sapien 7.49e-01
c	367	15	10.2	335	27	AQ065085	HS_2263_B2_MF_B05 CIT 7.49e-01
c	368	15	10.2	337	23	AA220811	my26g11.r1 Barstead mo 7.49e-01
c	369	15	10.2	337	21	AA049735	mj36g10.r1 Soares mous 7.49e-01
c	370	15	10.2	338	13	AA832458	oc92h06.s1 NCI_CGAP_GC 7.49e-01
	371	15	10.2	339	19	H62353	yu39c02.s1 Homo sapien 7.49e-01
c	372	15	10.2	339	24	AA319694	EST21948 Adrenal gland 7.49e-01
	373	15	10.2	341	14	AA903525	ok60f07.s1 NCI_CGAP_GC 7.49e-01
	374	15	10.2	344	28	AQ131388	HS_3038_B2_D08_MR CIT 7.49e-01
c	375	15	10.2	344	5	R28346	yh66h10.r1 Homo sapien 7.49e-01
c	376	15	10.2	344	19	N72599	za46f07.s1 Homo sapien 7.49e-01
	377	15	10.2	347	19	N20266	yx42c05.s1 Homo sapien 7.49e-01
c	378	15	10.2	347	15	AA988503	or89e04.s1 NCI_CGAP_Lu 7.49e-01
	379	15	10.2	348	16	H27512	y157d06.s1 Homo sapien 7.49e-01
c	380	15	10.2	348	15	AI047961	vn20a09.r1 Knowles Sol 7.49e-01
	381	15	10.2	349	8	AA283745	zt18c01.s1 Soares ovar 7.49e-01
	382	15	10.2	350	10	AA427358	zw20h05.r1 Soares ovar 7.49e-01
c	383	15	10.2	352	23	AA200623	mu01e08.r1 Soares mous 7.49e-01
c	384	15	10.2	354	8	AA504546	aa60c08.r1 NCI_CGAP_GC 7.49e-01
	385	15	10.2	355	16	H51972	yq32c04.r1 Homo sapien 7.49e-01
	386	15	10.2	355	25	AA338096	EST42917 Endometrial t 7.49e-01
c	387	15	10.2	355	8	AA490831	aa49g10.s1 NCI_CGAP_GC 7.49e-01
c	388	15	10.2	357	27	AQ069079	HS_2253_B1_B02_MR CIT 7.49e-01
c	389	15	10.2	358	7	AA518012	vi18e08.r1 Barstead mo 7.49e-01
c	390	15	10.2	358	22	N62139	yz62b04.s1 Soares mult 7.49e-01
c	391	15	10.2	358	7	AA473362	vd64a06.r1 Knowles Sol 7.49e-01
	392	15	10.2	360	19	H82052	yv78b01.s1 Homo sapien 7.49e-01
	393	15	10.2	360	9	C39209	C.elegans cDNA clone y 7.49e-01
c	394	15	10.2	360	9	C42394	C.elegans cDNA clone y 7.49e-01
c	395	15	10.2	360	9	C47359	C.elegans cDNA clone y 7.49e-01
c	396	15	10.2	360	9	C65444	C.elegans cDNA clone y 7.49e-01
c	397	15	10.2	360	6	C13088	C.elegans cDNA clone y 7.49e-01
c	398	15	10.2	360	9	C64767	C.elegans cDNA clone y 7.49e-01
c	399	15	10.2	360	9	C42428	C.elegans cDNA clone y 7.49e-01

c	400	15	10.2	360	9	C63355	C.elegans cDNA clone y 7.49e-01
c	401	15	10.2	360	23	AA240176	my23g04.r1 Barstead mo 7.49e-01
	402	15	10.2	361	16	R72077	yj87c07.s1 Homo sapien 7.49e-01
	403	15	10.2	361	10	AA400568	zu62h12.r1 Soares test 7.49e-01
c	404	15	10.2	362	7	AA458477	zx75e09.s1 Soares ovar 7.49e-01
	405	15	10.2	363	25	AA346091	EST521195 Greater omment 7.49e-01
	406	15	10.2	363	14	AA892215	EST196018 Normalized r 7.49e-01
	407	15	10.2	364	14	AA899839	UI-R-E0-cq-f-11-0-UI.s 7.49e-01
	408	15	10.2	366	16	R74534	yl01e12.s1 Homo sapien 7.49e-01
	409	15	10.2	366	15	AA947212	on51g02.s1 NCI_CGAP_Co 7.49e-01
	410	15	10.2	366	16	R72414	yj90b05.s1 Homo sapien 7.49e-01
	411	15	10.2	368	25	AA358044	EST66941 Fetal lung II 7.49e-01
	412	15	10.2	369	7	AA508315	MAAD0428.M3R Schistoso 7.49e-01
c	413	15	10.2	369	7	AA529818	vj12f06.r1 Barstead mo 7.49e-01
	414	15	10.2	371	5	T98460	ye60h03.r1 Homo sapien 7.49e-01
c	415	15	10.2	371	8	AA293603	zt26a11.r1 Soares ovar 7.49e-01
c	416	15	10.2	372	5	T72641	yd19d01.s1 Homo sapien 7.49e-01
	417	15	10.2	374	15	C77148	Mus musculus 3.5-dpc b 7.49e-01
	418	15	10.2	374	10	AA645982	vs29a04.r1 Stratagene 7.49e-01
	419	15	10.2	374	22	AA119628	mo27a01.r1 Life Tech m 7.49e-01
	420	15	10.2	374	20	W74925	me43g10.r1 Soares mous 7.49e-01
	421	15	10.2	374	8	AA562080	vl33f09.r1 Stratagene 7.49e-01
c	422	15	10.2	375	15	AA996746	UI-R-C0-hg-c-07-0-UI.s 7.49e-01
	423	15	10.2	376	22	AA212171	mu77h09.r1 Stratagene 7.49e-01
	424	15	10.2	376	8	AA225248	nc23d05.r1 NCI_CGAP_Pr 7.49e-01
c	425	15	10.2	376	20	AA003778	mg62d09.r1 Soares mous 7.49e-01
	426	15	10.2	377	9	C66669	C.elegans cDNA clone y 7.49e-01
	427	15	10.2	377	5	R40362	yf81f03.s1 Homo sapien 7.49e-01
	428	15	10.2	379	13	AA765727	oa05b04.s1 NCI_CGAP_GC 7.49e-01
	429	15	10.2	379	23	AA068201	mm48b04.r1 Stratagene 7.49e-01
c	430	15	10.2	379	28	AQ131564	HS_3045_A2_E07_MR CIT 7.49e-01
c	431	15	10.2	379	24	AA312317	EST183060 Jurkat T-cel 7.49e-01
c	432	15	10.2	380	16	T33764	EST59062 Homo sapiens 7.49e-01
c	433	15	10.2	380	28	AQ197851	CIT-HSP-2381D11.TF CIT 7.49e-01
	434	15	10.2	381	6	AA411421	zv27c02.s1 Soares ovar 7.49e-01
	435	15	10.2	381	21	W46638	zc33h05.s1 Soares sene 7.49e-01
	436	15	10.2	381	29	AQ135852	HS_3064_A2_H09_MR CIT 7.49e-01
c	437	15	10.2	381	21	W40682	mc66a04.r1 Soares mous 7.49e-01
c	438	15	10.2	381	21	W41532	mc45e03.r1 Soares mous 7.49e-01
c	439	15	10.2	382	27	AQ090321	HS_3000_B1_A04_MR CIT 7.49e-01
	440	15	10.2	383	21	AA038550	mi85e02.r1 Soares mous 7.49e-01
c	441	15	10.2	383	13	AA633758	ac27b12.s1 Stratagene 7.49e-01
	442	15	10.2	384	27	AQ085966	HS_2163_B1_C09_MR CIT 7.49e-01
	443	15	10.2	384	26	B34418	HS-1024-B2-A03-MR.abi 7.49e-01
	444	15	10.2	384	17	AI033266	ov37c04.x1 Soares_test 7.49e-01
c	445	15	10.2	384	20	W62384	md98f05.r1 Soares mous 7.49e-01
c	446	15	10.2	384	15	AI016151	ot73e01.s1 Soares_tota 7.49e-01
c	447	15	10.2	385	21	W35717	mc13f09.r1 Soares mous 7.49e-01
	448	15	10.2	389	8	AA490444	aa51f01.r1 NCI_CGAP_GC 7.49e-01
	449	15	10.2	389	23	AA098391	mn86f09.r1 Stratagene 7.49e-01
c	450	15	10.2	389	27	AQ066844	HS_2227_B2_A05_MR CIT 7.49e-01

	451	15	10.2	390	7	HUM425G03B	Human fetal brain cDNA 7.49e-01
	452	15	10.2	390	27	AQ027076	CIT-HSP-2313K14.TF CIT 7.49e-01
c	453	15	10.2	390	23	AA122972	mq23f11.r1 Barstead MP 7.49e-01
c	454	15	10.2	394	17	AA991748	ot88a03.s1 Soares_tota 7.49e-01
	455	15	10.2	395	15	C82947	Oryctolagus cuniculus 7.49e-01
c	456	15	10.2	395	14	AA897117	am08b08.s1 Soares_NFL_ 7.49e-01
c	457	15	10.2	395	22	N97957	1718C3 czapPPDd2.1, De 7.49e-01
c	458	15	10.2	395	15	C83803	Oryctolagus cuniculus 7.49e-01
c	459	15	10.2	396	19	H86403	ys93c06.r1 Homo sapien 7.49e-01
c	460	15	10.2	396	20	W78358	me88d07.r1 Soares mous 7.49e-01
	461	15	10.2	397	10	AA110231	mp42b07.r1 Barstead MP 7.49e-01
c	462	15	10.2	397	5	T99212	ye63a03.s1 Homo sapien 7.49e-01
c	463	15	10.2	397	21	W35701	mc13b09.r1 Soares mous 7.49e-01
c	464	15	10.2	397	13	AA805175	oa89f03.s1 NCI_CGAP_GC 7.49e-01
c	465	15	10.2	397	8	AA292039	zt46b01.s1 Soares ovar 7.49e-01
	466	15	10.2	398	19	H57214	yr08d08.r1 Homo sapien 7.49e-01
	467	15	10.2	398	28	AQ119266	HS_2165_B1_C08_MF CIT 7.49e-01
c	468	15	10.2	398	27	AQ060861	CIT-HSP-2349I15.TF CIT 7.49e-01
c	469	15	10.2	399	15	AA926928	om68e06.s1 NCI_CGAP_GC 7.49e-01
c	470	15	10.2	399	12	AA764694	ETN157 Trypanosoma bru 7.49e-01
	471	15	10.2	400	13	AA628805	af58b02.s1 Soares tota 7.49e-01
	472	15	10.2	400	27	AQ064517	HS_2211_A2_D08_MR CIT 7.49e-01
c	473	15	10.2	400	8	AA544401	vj65b04.r1 Knowles Sol 7.49e-01
c	474	15	10.2	400	19	H58741	yr25g04.s1 Homo sapien 7.49e-01
c	475	15	10.2	400	18	AI154304	ud31b12.r1 Soares mous 7.49e-01
c	476	15	10.2	400	11	AA711187	yt56g02.r2 Barstead mo 7.49e-01
c	477	15	10.2	401	20	W98550	mg14b04.r1 Soares mous 7.49e-01
c	478	15	10.2	401	27	AQ030704	HS_2184_B1_B02_MF CIT 7.49e-01
c	479	15	10.2	401	13	AA875608	UI-R-E0-cv-c-09-0-UI.s 7.49e-01
	480	15	10.2	403	13	AA158962	zo60b12.r1 Stratagene 7.49e-01
	481	15	10.2	404	7	AA455196	zx79f02.r1 Soares ovar 7.49e-01
	482	15	10.2	405	15	AI003371	am78d11.s1 Stratagene 7.49e-01
	483	15	10.2	405	14	AA968058	uh08h04.r1 Soares mous 7.49e-01
	484	15	10.2	406	17	AI128589	qa61e03.s1 Soares_feta 7.49e-01
	485	15	10.2	406	28	AQ172320	HS_3183_A1_B07_MR CIT 7.49e-01
	486	15	10.2	407	28	AQ131142	HS_3036_A1_E01_MR CIT 7.49e-01
c	487	15	10.2	408	8	AA543411	vk37a02.r1 Soares mous 7.49e-01
	488	15	10.2	409	8	AA235517	zt30b04.s1 Soares ovar 7.49e-01
c	489	15	10.2	409	29	AQ139237	HS_3090_A1_F10_MR CIT 7.49e-01
	490	15	10.2	410	28	AQ167119	HS_3150_A1_B01_MR CIT 7.49e-01
	491	15	10.2	410	6	AA290944	zt18h11.s1 Soares ovar 7.49e-01
c	492	15	10.2	410	12	AA755810	vu04a11.r1 Soares mous 7.49e-01
c	493	15	10.2	410	9	AA564385	ng47h03.s1 NCI_CGAP_Co 7.49e-01
c	494	15	10.2	410	12	AA729506	nx54a01.s1 NCI_CGAP_Al 7.49e-01
	495	15	10.2	411	18	AI040253	ox92b09.x1 Soares_sene 7.49e-01
c	496	15	10.2	411	5	R07410	ye96g02.r1 Homo sapien 7.49e-01
c	497	15	10.2	411	22	AA100562	zn51e01.r1 Stratagene 7.49e-01
c	498	15	10.2	412	9	AA579013	nf30h10.s1 NCI_CGAP_Pr 7.49e-01
c	499	15	10.2	413	14	C89496	Mus musculus early bla 7.49e-01
	500	15	10.2	414	28	AQ111472	CIT-HSP-2372O22.TF CIT 7.49e-01
	501	15	10.2	414	7	AA464706	zx82b09.s1 Soares ovar 7.49e-01

	502	15	10.2	414	27	AQ006086	CIT-HSP-2297B12.TF CIT 7.49e-01
	503	15	10.2	414	5	R15715	ya51b08.r2 Homo sapien 7.49e-01
c	504	15	10.2	414	16	R81954	yi98e11.s1 Homo sapien 7.49e-01
c	505	15	10.2	414	8	AA414142	vc59a03.s1 Knowles Sol 7.49e-01
c	506	15	10.2	415	6	AA412401	zv12d01.r1 Soares ovar 7.49e-01
c	507	15	10.2	415	17	AI089952	qa20f08.x1 NCI_CGAP_Br 7.49e-01
c	508	15	10.2	415	11	AA723457	zg69c11.s1 Soares feta 7.49e-01
	509	15	10.2	416	19	H72828	ya06a03.s1 Homo sapien 7.49e-01
c	510	15	10.2	416	24	AA277697	va81b01.r1 Soares mous 7.49e-01
c	511	15	10.2	416	8	AA234832	zs38b05.s1 Soares NhHM 7.49e-01
	512	15	10.2	417	28	AQ195796	RPCI11-60I18.TK RPCI11 7.49e-01
	513	15	10.2	417	21	AA032361	mi44c07.r1 Soares mous 7.49e-01
	514	15	10.2	418	28	AQ182645	HS_3090_B1_B03_MR CIT 7.49e-01
c	515	15	10.2	420	17	AI026737	ov91e07.x1 Soares_test 7.49e-01
c	516	15	10.2	421	28	AQ203015	RPCI11-46J10.TJ RPCI11 7.49e-01
c	517	15	10.2	423	18	AI105337	EST214626 Normalized r 7.49e-01
	518	15	10.2	424	24	AA305589	EST176663 Colon carcin 7.49e-01
	519	15	10.2	424	20	W03411	za07c10.r1 Soares mela 7.49e-01
c	520	15	10.2	424	15	AI011987	EST206438 Normalized r 7.49e-01
c	521	15	10.2	424	28	AQ218453	HS_2001_B1_G10_T7 CIT 7.49e-01
c	522	15	10.2	424	26	B37025	HS-1042-B1-A04-MF.abi 7.49e-01
	523	15	10.2	425	8	R78586	yi73d09.s1 Soares plac 7.49e-01
	524	15	10.2	428	11	AA669369	ad47d06.s1 Stratagene 7.49e-01
c	525	15	10.2	428	9	AA533189	nj11b10.s1 NCI_CGAP_Pr 7.49e-01
c	526	15	10.2	428	10	AA616585	vn65h10.r1 Barstead mo 7.49e-01
	527	15	10.2	431	28	AQ182226	HS_3209_B2_H02_MR CIT 7.49e-01
	528	15	10.2	431	20	W98593	mg22g04.r1 Soares mous 7.49e-01
c	529	15	10.2	431	8	AA236576	zr66b12.s1 Soares NhHM 7.49e-01
	530	15	10.2	432	15	AA907513	ok96g05.s1 NCI_CGAP_Lu 7.49e-01
	531	15	10.2	432	14	AA970220	op09d02.s1 NCI_CGAP_Ki 7.49e-01
	532	15	10.2	433	8	AA513378	ne59c10.s1 NCI_CGAP_Co 7.49e-01
c	533	15	10.2	433	12	AA140787	CK00471.3prime CK Dros 7.49e-01
c	534	15	10.2	433	27	AQ046035	RPCI11-36O19.TJ RPCI11 7.49e-01
	535	15	10.2	434	6	AA052983	zl70f01.s1 Stratagene 7.49e-01
	536	15	10.2	434	6	AA029230	zk09h09.s1 Soares preg 7.49e-01
	537	15	10.2	435	21	W74307	zd75a04.s1 Soares feta 7.49e-01
c	538	15	10.2	435	29	AQ250798	T19H7-T7 TAMU Arabidop 7.49e-01
c	539	15	10.2	435	6	AA383308	EST96667 Testis I Homo 7.49e-01
	540	15	10.2	436	13	AA857752	oe42a04.s1 NCI_CGAP_Pr 7.49e-01
c	541	15	10.2	436	23	AA153686	mq59a02.r1 Soares 2NbM 7.49e-01
	542	15	10.2	437	28	AQ223912	HS_2259_B2_B08_T7 CIT 7.49e-01
	543	15	10.2	437	23	AA152898	mq56g02.r1 Soares 2NbM 7.49e-01
	544	15	10.2	437	15	AA932919	om83h06.s1 NCI_CGAP_Ki 7.49e-01
c	545	15	10.2	437	8	AA543684	vj83c11.r1 Soares mous 7.49e-01
c	546	15	10.2	437	10	AI155508	ue05a03.r1 Soares mous 7.49e-01
	547	15	10.2	438	5	R32831	yh71f01.s1 Homo sapien 7.49e-01
c	548	15	10.2	438	6	W88912	zh69b02.r1 Soares feta 7.49e-01
	549	15	10.2	439	15	AA944619	EST200118 Normalized r 7.49e-01
	550	15	10.2	439	28	AQ180532	HS_3219_A2_D04_MR CIT 7.49e-01
c	551	15	10.2	439	16	R61038	yh05d07.r1 Homo sapien 7.49e-01
c	552	15	10.2	439	8	AA293075	zt48d12.r1 Soares ovar 7.49e-01

	553	15	10.2	440	29	AQ241881	CITBI-E1-2501021.TF.1	7.49e-01
	554	15	10.2	440	20	N80754	za98e06.s1 Homo sapien	7.49e-01
c	555	15	10.2	440	7	AA511204	vh56a06.r1 Soares mous	7.49e-01
	556	15	10.2	441	26	B45640	HS-1062-A1-G01-MF.abi	7.49e-01
	557	15	10.2	442	8	AA135204	zo27d02.s1 Stratagene	7.49e-01
c	558	15	10.2	442	28	AQ183632	HS_3197_B2_G07_MR CIT	7.49e-01
c	559	15	10.2	442	13	AA804239	nz29h04.s1 NCI_CGAP_GC	7.49e-01
c	560	15	10.2	443	7	AA459289	zx89b02.s1 Soares ovar	7.49e-01
	561	15	10.2	444	18	AI042580	oy14f02.x1 Soares_sene	7.49e-01
	562	15	10.2	446	27	AQ033831	HS_2194_B2_D07_MR CIT	7.49e-01
	563	15	10.2	446	6	AA411152	zt37h04.s1 Soares ovar	7.49e-01
	564	15	10.2	447	8	AA293487	zt48d12.s1 Soares ovar	7.49e-01
	565	15	10.2	447	11	AA708014	zg05c05.s1 Soares pine	7.49e-01
c	566	15	10.2	447	26	HSU77522	Human genomic clone B2	7.49e-01
c	567	15	10.2	448	12	AA789948	vx70g10.r1 Stratagene	7.49e-01
c	568	15	10.2	448	19	H66320	yu14a06.s1 Homo sapien	7.49e-01
	569	15	10.2	449	21	W49572	zc40h09.s1 Soares sene	7.49e-01
c	570	15	10.2	449	21	W58027	zd21d07.s1 Soares feta	7.49e-01
c	571	15	10.2	449	20	W65657	me10b11.r1 Soares mous	7.49e-01
	572	15	10.2	450	15	AI002204	oq85f06.s1 NCI_CGAP_Ki	7.49e-01
c	573	15	10.2	451	21	W81308	zd85b12.r1 Soares feta	7.49e-01
c	574	15	10.2	451	22	N58411	yv70g03.s1 Soares feta	7.49e-01
c	575	15	10.2	452	28	AQ113200	CIT-HSP-2374E19.TR CIT	7.49e-01
	576	15	10.2	454	14	C91416	Dictyostelium discoide	7.49e-01
c	577	15	10.2	454	14	AA924672	UI-R-A1-ea-a-04-0-UI.s	7.49e-01
	578	15	10.2	455	27	AQ009469	CIT-HSP-2284E9.TR3 CIT	7.49e-01
	579	15	10.2	457	19	H93662	yw55c09.r1 Homo sapien	7.49e-01
c	580	15	10.2	457	28	AQ170461	HS_2248_A1_F01_MR CIT	7.49e-01
	581	15	10.2	458	21	H49633	yq20e02.s1 Soares feta	7.49e-01
	582	15	10.2	459	11	AA069012	zm10b06.s1 Stratagene	7.49e-01
c	583	15	10.2	459	16	H20714	ym47a12.r1 Homo sapien	7.49e-01
c	584	15	10.2	459	11	AA690449	vt31g03.r1 Barstead mo	7.49e-01
	585	15	10.2	460	28	AQ193692	CIT-HSP-2385K10.TF CIT	7.49e-01
	586	15	10.2	460	22	AA159966	zo75b12.s1 Stratagene	7.49e-01
	587	15	10.2	462	22	AA186868	zp74a02.s1 Stratagene	7.49e-01
	588	15	10.2	463	14	AA977803	oq68g03.s1 NCI_CGAP_Ki	7.49e-01
	589	15	10.2	463	20	W65942	zm14g08.r1 Soares mous	7.49e-01
	590	15	10.2	463	20	W46984	zc43e01.s1 Soares sene	7.49e-01
c	591	15	10.2	464	5	T50218	yb78b10.s1 Homo sapien	7.49e-01
	592	15	10.2	465	18	AI042546	oy06f07.x1 Soares_sene	7.49e-01
	593	15	10.2	465	22	AA186929	zp74h01.s1 Stratagene	7.49e-01
	594	15	10.2	466	16	R95120	yq44f09.r1 Homo sapien	7.49e-01
	595	15	10.2	466	23	AA162875	mr30e10.r1 Soares mous	7.49e-01
c	596	15	10.2	466	29	AQ144744	HS_3092_A1_D07_MF CIT	7.49e-01
c	597	15	10.2	466	10	AA430337	zw20h05.s1 Soares ovar	7.49e-01
	598	15	10.2	468	18	AI079597	ox91b08.s1 Soares_sene	7.49e-01
c	599	15	10.2	473	19	N48803	yy76d08.s1 Homo sapien	7.49e-01
c	600	15	10.2	473	22	AA020091	mh49a01.r1 Soares mous	7.49e-01
c	601	15	10.2	473	6	AA041519	zk44c07.s1 Soares preg	7.49e-01
	602	15	10.2	474	22	AA026359	ze92a07.s1 Soares feta	7.49e-01
c	603	15	10.2	474	10	AA664357	af25h07.s1 Soares tota	7.49e-01

c	604	15	10.2	474	5	R25277	yg53c08.r1 Homo sapien	7.49e-01
	605	15	10.2	475	26	B75268	RPCI11-1406.TV RPCI11	7.49e-01
c	606	15	10.2	475	26	AQ010714	F27N1TFB IGF Arabidops	7.49e-01
c	607	15	10.2	475	22	AA198176	mv49e10.r1 Soares mous	7.49e-01
	608	15	10.2	476	7	AA441783	zw62b03.s1 Soares tota	7.49e-01
	609	15	10.2	477	13	AA859846	UI-R-E0-cc-h-07-0-UI.s	7.49e-01
	610	15	10.2	478	5	R27250	yh53c03.r1 Homo sapien	7.49e-01
	611	15	10.2	479	5	T86102	yd76g12.r1 Homo sapien	7.49e-01
c	612	15	10.2	480	27	AQ040850	CIT-HSP-2336B9.TF CIT-	7.49e-01
	613	15	10.2	481	5	T52171	yb22c12.s1 Homo sapien	7.49e-01
	614	15	10.2	482	18	AI186334	qd30a10.x1 Soares_plac	7.49e-01
	615	15	10.2	483	29	AQ228036	HS_2024_A1_F01_MR CIT	7.49e-01
c	616	15	10.2	483	27	AQ032310	HS_2196_A1_F04_MF CIT	7.49e-01
c	617	15	10.2	484	15	AI006277	ua68c03.r1 Soares 2NbM	7.49e-01
	618	15	10.2	485	28	AQ209361	HS_3240_A2_C06_T7 CIT	7.49e-01
	619	15	10.2	485	14	AA903097	ok46f07.s1 NCI_CGAP_Le	7.49e-01
	620	15	10.2	486	14	AA865042	oh32b09.s1 NCI_CGAP_Ki	7.49e-01
	621	15	10.2	486	10	AA641460	nr93e09.s1 NCI_CGAP_Pr	7.49e-01
	622	15	10.2	487	18	AI144467	qc07d05.x1 Soares_feta	7.49e-01
c	623	15	10.2	487	28	AQ186484	HS_3094_B2_H09_MF CIT	7.49e-01
	624	15	10.2	488	13	AA833802	od61h10.s1 NCI_CGAP_GC	7.49e-01
	625	15	10.2	488	28	AQ155354	HS_3058_A1_B06_MF CIT	7.49e-01
c	626	15	10.2	488	11	AA687162	nv60a05.s1 NCI_CGAP_GC	7.49e-01
c	627	15	10.2	488	6	AA405206	zu52h08.r1 Soares ovar	7.49e-01
	628	15	10.2	490	14	AA827432	of32d03.s1 NCI_CGAP_Ki	7.49e-01
c	629	15	10.2	490	9	AA568165	n186c07.s1 NCI_CGAP_Co	7.49e-01
	630	15	10.2	491	27	AQ087613	HS_2236_A2_D05_MR CIT	7.49e-01
	631	15	10.2	492	10	AA403006	zu54e12.s1 Soares ovar	7.49e-01
c	632	15	10.2	493	29	AQ146940	HS_2248_A2_C05_MR CIT	7.49e-01
c	633	15	10.2	494	5	T79800	yd74f06.s1 Homo sapien	7.49e-01
c	634	15	10.2	495	18	AI038929	ox96h04.x1 Soares_sene	7.49e-01
c	635	15	10.2	496	26	B23995	F18B22TF IGF Arabidops	7.49e-01
	636	15	10.2	498	23	AA110678	mp48d07.r1 Barstead MP	7.49e-01
c	637	15	10.2	498	14	AA920365	vx94h10.r1 Stratagene	7.49e-01
	638	15	10.2	499	28	AQ191526	HS_3230_B1_G01_T7 CIT	7.49e-01
c	639	15	10.2	499	23	AA241090	mv23h10.r1 GuayWoodfor	7.49e-01
	640	15	10.2	501	7	AA449173	zx02h04.r1 Soares tota	7.49e-01
	641	15	10.2	501	12	AA780607	ad40d11.s1 Stratagene	7.49e-01
	642	15	10.2	501	16	R56130	yg94f02.s1 Homo sapien	7.49e-01
c	643	15	10.2	501	10	AA648540	ns23g10.s1 NCI_CGAP_GC	7.49e-01
	644	15	10.2	502	14	AA931452	om89c04.s1 NCI_CGAP_Ki	7.49e-01
c	645	15	10.2	502	10	AA394404	25987 Lambda-PRL2 Arab	7.49e-01
c	646	15	10.2	503	9	AA516511	nh52e03.r1 NCI_CGAP_Pr	7.49e-01
c	647	15	10.2	503	8	AA214533	zr92a11.s1 NCI_CGAP_GC	7.49e-01
	648	15	10.2	504	6	C05801	similar to profilin I,	7.49e-01
	649	15	10.2	504	28	AQ185466	HS_3124_B2_F06_T7 CIT	7.49e-01
	650	15	10.2	504	14	AA931246	SWBmL38A336T3 Brugia m	7.49e-01
	651	15	10.2	504	28	AQ126775	HS_3049_A1_E10_MF CIT	7.49e-01
c	652	15	10.2	504	7	AA521899	vi16a08.r1 Barstead mo	7.49e-01
c	653	15	10.2	504	19	N74242	za59g07.s1 Homo sapien	7.49e-01
	654	15	10.2	505	17	AA922086	oi29g11.s1 NCI_CGAP_HN	7.49e-01

	655	15	10.2	505	28	AQ093382	HS_3003_A2_D03_T7 CIT	7.49e-01
	656	15	10.2	505	23	AA073291	mm95c03.r1 Stratagene	7.49e-01
c	657	15	10.2	505	9	AA590044	vn57b02.r1 Barstead mo	7.49e-01
	658	15	10.2	506	18	AI149462	qc78b11.x1 Soares_plac	7.49e-01
c	659	15	10.2	506	5	T83048	yd43d07.r1 Homo sapien	7.49e-01
	660	15	10.2	507	22	N53683	yv68b10.s1 Soares feta	7.49e-01
c	661	15	10.2	507	15	AA944353	EST199852 Normalized r	7.49e-01
c	662	15	10.2	507	23	AA222500	mw20b06.r1 Soares mous	7.49e-01
	663	15	10.2	508	14	C92435	Dictyostelium discoide	7.49e-01
c	664	15	10.2	508	28	AQ215827	HS_3238_B2_A12_T7 CIT	7.49e-01
c	665	15	10.2	508	18	AI138919	qc56g03.x1 Soares_plac	7.49e-01
c	666	15	10.2	509	7	AA522075	vi16e07.r1 Barstead mo	7.49e-01
	667	15	10.2	510	8	AA394143	zt49f07.s1 Soares ovar	7.49e-01
c	668	15	10.2	510	14	AA851224	EST193992 Normalized r	7.49e-01
	669	15	10.2	511	13	AA609290	af13e11.s1 Soares test	7.49e-01
	670	15	10.2	512	5	T61661	yb87d09.s1 Homo sapien	7.49e-01
	671	15	10.2	513	29	AQ140898	HS_3126_B2_C02_T7 CIT	7.49e-01
	672	15	10.2	513	27	B90384	CIT-HSP-2165P15.TF CIT	7.49e-01
	673	15	10.2	514	14	AA953106	oq70f10.s1 NCI_CGAP_Ki	7.49e-01
	674	15	10.2	514	14	AA984997	am68c05.s1 Barstead sp	7.49e-01
	675	15	10.2	515	15	AA826385	of27d04.s1 NCI_CGAP_Ki	7.49e-01
	676	15	10.2	515	15	AA912221	ol92e10.s1 NCI_CGAP_PN	7.49e-01
	677	15	10.2	516	29	AQ140018	HS_3106_A1_F09_T7 CIT	7.49e-01
	678	15	10.2	517	22	AA168913	ms39d04.r1 Life Tech m	7.49e-01
c	679	15	10.2	518	10	AA655515	vs25d11.r1 Barstead mo	7.49e-01
	680	15	10.2	520	13	AA874855	UI-R-E0-cq-h-10-0-UI.s	7.49e-01
	681	15	10.2	521	13	AA599836	ag33e02.s1 Jia bone ma	7.49e-01
	682	15	10.2	522	27	B85301	RPCI11-1406.TV8 RPCI11	7.49e-01
	683	15	10.2	522	14	C90495	Dictyostelium discoide	7.49e-01
	684	15	10.2	524	28	AQ175541	HS_3215_A2_C10_T7 CIT	7.49e-01
c	685	15	10.2	524	7	AA444013	zv44a12.r1 Soares ovar	7.49e-01
c	686	15	10.2	525	17	AI130885	qb81e04.x1 Soares_feta	7.49e-01
c	687	15	10.2	525	6	AA115355	zl10b09.s1 Soares preg	7.49e-01
c	688	15	10.2	527	11	AA071349	zm73e01.s1 Stratagene	7.49e-01
c	689	15	10.2	530	27	B88150	RPCI11-21C23.TV8 RPCI1	7.49e-01
	690	15	10.2	532	19	H43974	zs74h08.s1 Homo sapien	7.49e-01
c	691	15	10.2	532	20	W65780	me14b02.r1 Soares mous	7.49e-01
	692	15	10.2	533	23	AA255009	mz76b06.r1 Soares mous	7.49e-01
	693	15	10.2	536	15	C77085	Mus musculus 3.5-dpc b	7.49e-01
c	694	15	10.2	536	28	AQ223786	HS_2218_A1_B01_MR CIT	7.49e-01
	695	15	10.2	538	28	AQ174640	HS_3205_B2_B04_T7 CIT	7.49e-01
	696	15	10.2	539	27	B60693	CIT-HSP-2015N10.TR8 CI	7.49e-01
	697	15	10.2	540	14	AA830574	oc55b01.s1 NCI_CGAP_GC	7.49e-01
	698	15	10.2	540	15	AA932117	oo62h10.s1 NCI_CGAP_Lu	7.49e-01
	699	15	10.2	541	8	AA102360	zl91c05.s1 Stratagene	7.49e-01
	700	15	10.2	545	29	AQ237437	RPCI11-66N17.TR RPCI11	7.49e-01
	701	15	10.2	545	15	AA931254	om88e03.s1 NCI_CGAP_Ki	7.49e-01
c	702	15	10.2	546	23	AA162589	mr26e04.r1 Soares mous	7.49e-01
	703	15	10.2	547	6	AA007643	zh99e05.r1 Soares feta	7.49e-01
	704	15	10.2	547	28	AQ179347	HS_3165_A2_D03_T7 CIT	7.49e-01
	705	15	10.2	548	27	AQ037222	CIT-HSP-2335F16.TF CIT	7.49e-01

c	706	15	10.2	548	24	AA277269	vb68d06.r1 Barstead mo	7.49e-01
	707	15	10.2	551	12	AA781382	aj25f04.s1 Soares test	7.49e-01
	708	15	10.2	551	10	AA614570	np53d02.s1 NCI_CGAP_Br	7.49e-01
c	709	15	10.2	551	15	AI017079	ou31f05.x1 Soares_NFL_	7.49e-01
	710	15	10.2	553	23	AA254711	ms80e01.r1 Soares mous	7.49e-01
	711	15	10.2	554	10	AA603126	np51e01.s1 NCI_CGAP_Br	7.49e-01
	712	15	10.2	554	13	AA857007	oh94d03.s1 NCI_CGAP_HN	7.49e-01
c	713	15	10.2	554	10	AA655494	vs25b11.r1 Barstead mo	7.49e-01
	714	15	10.2	555	26	B23332	F28F8TR IGF Arabidopsi	7.49e-01
	715	15	10.2	559	13	AA826780	nr89e12.s1 NCI_CGAP_Pr	7.49e-01
	716	15	10.2	560	14	C90396	Dictyostelium discoide	7.49e-01
c	717	15	10.2	560	11	AA712753	32312 Lambda-PRL2 Arab	7.49e-01
	718	15	10.2	561	27	FR0032793	Fugu rubripes GSS sequ	7.49e-01
c	719	15	10.2	562	18	AI102593	EST211882 Normalized r	7.49e-01
	720	15	10.2	563	6	AA045232	zk65e03.s1 Soares preg	7.49e-01
	721	15	10.2	563	14	C90518	Dictyostelium discoide	7.49e-01
c	722	15	10.2	563	26	AQ024010	CpGR0349A Cryptosporid	7.49e-01
	723	15	10.2	564	11	AA670520	SWOv3MCAM02G08SK Oncho	7.49e-01
	724	15	10.2	567	21	W68286	zd35e11.r1 Soares feta	7.49e-01
	725	15	10.2	568	8	AA549867	0940m3 gmbPfHB3.1, G.	7.49e-01
	726	15	10.2	569	13	AA826748	nr89b02.s1 NCI_CGAP_Pr	7.49e-01
	727	15	10.2	569	22	AA148276	zo44g12.r1 Stratagene	7.49e-01
	728	15	10.2	571	7	AA453073	zx41e06.r1 Soares tota	7.49e-01
	729	15	10.2	573	26	B88475	CpG0218B CpIOWAgDNA1 C	7.49e-01
	730	15	10.2	574	17	AI117688	ui48f04.x1 Sugano mous	7.49e-01
	731	15	10.2	576	22	W91555	zd95e04.s1 Soares feta	7.49e-01
c	732	15	10.2	577	18	AI177505	EST221137 Normalized r	7.49e-01
c	733	15	10.2	581	28	AQ174093	HS_3200_B1_E11_T7 CIT	7.49e-01
	734	15	10.2	582	10	AA641699	nr79g02.s1 NCI_CGAP_Pr	7.49e-01
	735	15	10.2	582	13	AA161304	zq38a09.r1 Stratagene	7.49e-01
	736	15	10.2	585	9	AA521431	aa69b11.s1 NCI_CGAP_GC	7.49e-01
c	737	15	10.2	585	8	AA404312	zt43g12.r1 Soares ovar	7.49e-01
c	738	15	10.2	586	26	FR0021323	F.rubripes GSS sequenc	7.49e-01
c	739	15	10.2	587	13	C87911	Mus musculus fertilize	7.49e-01
	740	15	10.2	589	10	AA654356	nt12c11.s1 NCI_CGAP_Pr	7.49e-01
	741	15	10.2	590	21	AA081801	zn32a10.s1 Stratagene	7.49e-01
	742	15	10.2	590	22	AA028203	ze75b12.r1 Soares feta	7.49e-01
	743	15	10.2	590	27	B68808	CIT-HSP-2023H12.TF CIT	7.49e-01
	744	15	10.2	591	7	AA537267	vk46a06.r1 Soares mous	7.49e-01
c	745	15	10.2	594	28	AQ159190	nbxb0013I17f CUGI Rice	7.49e-01
c	746	15	10.2	595	15	AI013792	EST208467 Normalized r	7.49e-01
c	747	15	10.2	597	7	AA516806	vh97b12.r1 Barstead mo	7.49e-01
c	748	15	10.2	597	15	AI012585	EST207036 Normalized r	7.49e-01
c	749	15	10.2	599	27	AQ056489	CIT-HSP-2341P3.TF CIT-	7.49e-01
	750	15	10.2	600	29	CPU90782	Culex pipiens pipiens	7.49e-01
c	751	15	10.2	600	7	AA516801	vh97b06.r1 Barstead mo	7.49e-01
	752	15	10.2	601	26	FR0025981	F.rubripes GSS sequenc	7.49e-01
	753	15	10.2	610	10	AA600850	np45g07.s1 NCI_CGAP_Br	7.49e-01
c	754	15	10.2	617	8	AA543187	vj85f03.s1 Knowles Sol	7.49e-01
c	755	15	10.2	624	27	B93836	CIT-HSP-2168H6.TR CIT-	7.49e-01
c	756	15	10.2	626	28	AQ158628	nbxb0011J22r CUGI Rice	7.49e-01

	757	15	10.2	627	17	MMAJ6264	Mus musculus mRNA; EST 7.49e-01
	758	15	10.2	638	10	AA660642	00530 MtRRE Medicago t 7.49e-01
	759	15	10.2	645	22	AA151681	zo29g01.s1 Stratagene 7.49e-01
	760	15	10.2	645	8	AA496671	zv38b09.s1 Soares ovar 7.49e-01
	761	15	10.2	665	6	AA149656	z139g08.s1 Soares preg 7.49e-01
	762	15	10.2	670	22	AA165699	zo75g06.s1 Stratagene 7.49e-01
c	763	15	10.2	670	20	W01369	yy66h09.r1 Soares mult 7.49e-01
	764	15	10.2	678	27	AG012272	Homo sapiens genomic D 7.49e-01
	765	15	10.2	678	26	AG005208	Homo sapiens genomic D 7.49e-01
	766	15	10.2	681	26	B67769	T24F22TF TAMU Arabidop 7.49e-01
	767	15	10.2	686	18	AI084757	oz79g12.x1 Soares_sene 7.49e-01
	768	15	10.2	693	8	AA522611	ni38h01.s1 NCI_CGAP_Lu 7.49e-01
c	769	15	10.2	697	26	B88564	CpG0126B CpIOWAgDNA1 C 7.49e-01
	770	15	10.2	702	7	AA454672	zx76a04.s1 Soares ovar 7.49e-01
c	771	15	10.2	702	26	AG008022	Homo sapiens genomic D 7.49e-01
	772	15	10.2	707	9	AA551088	nk75b05.s1 NCI_CGAP_Sc 7.49e-01
	773	15	10.2	708	27	AQ007717	CIT-HSP-2291P4.TR CIT- 7.49e-01
	774	15	10.2	712	26	B78507	T30N14TPC TAMU Arabido 7.49e-01
c	775	15	10.2	712	26	AG008423	Homo sapiens genomic D 7.49e-01
c	776	15	10.2	719	28	AQ193453	CIT-HSP-2381J15.TR CIT 7.49e-01
c	777	15	10.2	726	29	AQ239904	CIT-HSP-2386P17.TF.1 C 7.49e-01
c	778	15	10.2	732	11	AA696037	GM04945.3prime GM Dros 7.49e-01
c	779	15	10.2	735	27	AG012730	Homo sapiens genomic D 7.49e-01
	780	15	10.2	758	20	W25914	14h3 Human retina cDNA 7.49e-01
c	781	15	10.2	760	26	B06506	cSRL-81d7-u cSRL flow 7.49e-01
c	782	15	10.2	761	13	AA643228	nr61e04.s1 NCI_CGAP_Ly 7.49e-01
c	783	15	10.2	761	27	AQ049984	nbxb0003aF03f CUGI Ric 7.49e-01
	784	15	10.2	766	28	AQ157144	nbxb0009C03r CUGI Rice 7.49e-01
	785	15	10.2	773	8	AA550581	1746m3 gmbPfHB3.1, G. 7.49e-01
	786	15	10.2	777	26	B19527	P23F10-T7 IGF Arabidop 7.49e-01
c	787	15	10.2	779	18	AI148549	qc68f10.x1 Soares_plac 7.49e-01
	788	15	10.2	807	26	B20747	T19E19-T7 TAMU Arabido 7.49e-01
	789	15	10.2	817	26	B20751	T19J11-T7 TAMU Arabido 7.49e-01
	790	15	10.2	864	26	B10741	F16A24-T7 IGF Arabidop 7.49e-01
c	791	15	10.2	880	17	AI069145	mgae0005cH02f Magnapor 7.49e-01
	792	15	10.2	922	22	AA203697	zx52h05.r1 Soares feta 7.49e-01
	793	15	10.2	928	26	B09439	F10O3-T7 IGF Arabidops 7.49e-01
	794	15	10.2	945	26	B09191	T30D19-Sp6 TAMU Arabid 7.49e-01
	795	15	10.2	1082	17	AI068747	mgae0004aD01f Magnapor 7.49e-01
c	796	15	10.2	1502	6	T46905	EST014 BL29 Burkitt's 7.49e-01
	797	14	9.5	105	22	N97394	0561M7 gmbPfHB3.1, G. 1.49e+01
	798	14	9.5	118	26	B49840	CIT-HSP-461H3.TV CIT-H 1.49e+01
c	799	14	9.5	156	24	AA324090	EST26957 Cerebellum II 1.49e+01
	800	14	9.5	170	10	AA627745	nq52f01.s1 NCI_CGAP_Co 1.49e+01
c	801	14	9.5	173	25	AA348612	EST55152 Hippocampus I 1.49e+01
	802	14	9.5	196	27	B89132	CIT-HSP-2174M5.TR CIT- 1.49e+01
c	803	14	9.5	209	22	AA009987	ze16g10.s1 Soares feta 1.49e+01
c	804	14	9.5	213	24	AA323708	EST26534 Cerebellum II 1.49e+01
c	805	14	9.5	214	19	N64067	yx41b09.s1 Homo sapien 1.49e+01
c	806	14	9.5	216	20	HSC03H112	H. sapiens partial cDN 1.49e+01
	807	14	9.5	221	29	AQ249744	F14P9-Sp6.1 IGF Arabid 1.49e+01

c	808	14	9.5	225	14	AA905442	oj96g11.s1 Soares_NFL_ 1.49e+01
c	809	14	9.5	225	25	AA342311	EST47875 Fetal spleen 1.49e+01
c	810	14	9.5	228	22	N64229	yv63d05.r1 Soares feta 1.49e+01
	811	14	9.5	230	23	AA168588	ms32f01.r1 Stratagene 1.49e+01
	812	14	9.5	231	23	AA247786	hfe0360.seq.F Human fe 1.49e+01
c	813	14	9.5	231	6	AA035152	zk28e07.s1 Soares preg 1.49e+01
	814	14	9.5	239	9	AA523458	ni63f12.s1 NCI_CGAP_Pr 1.49e+01
c	815	14	9.5	246	25	AA348474	EST54932 Hippocampus I 1.49e+01
c	816	14	9.5	251	15	AI053867	qi70d05.x1 NCI_CGAP_Ov 1.49e+01
c	817	14	9.5	252	11	AA681237	vr74a10.s1 Knowles Sol 1.49e+01
c	818	14	9.5	252	15	AI053607	qi72f08.x1 NCI_CGAP_Ov 1.49e+01
c	819	14	9.5	261	15	AA897755	oj77b10.s1 Soares_NFL_ 1.49e+01
c	820	14	9.5	263	13	AA181833	zp64b12.s1 Stratagene 1.49e+01
	821	14	9.5	264	14	AA849364	EST192131 Normalized r 1.49e+01
c	822	14	9.5	272	25	AA348621	EST55467 Infant adrena 1.49e+01
c	823	14	9.5	272	10	AA583280	nn41h05.s1 NCI_CGAP_GC 1.49e+01
c	824	14	9.5	273	24	AA323389	EST26344 Cerebellum II 1.49e+01
c	825	14	9.5	280	19	HSC1DG011	H. sapiens partial cDN 1.49e+01
c	826	14	9.5	287	21	R93346	yt70h06.s1 Soares feta 1.49e+01
c	827	14	9.5	288	10	AA600758	np44c03.s1 NCI_CGAP_Br 1.49e+01
c	828	14	9.5	293	16	H19873	yn57c07.s1 Homo sapien 1.49e+01
	829	14	9.5	298	21	AA032726	mi31e06.r1 Soares mous 1.49e+01
	830	14	9.5	303	20	HSC3FG031	H. sapiens partial cDN 1.49e+01
c	831	14	9.5	304	24	AA326687	EST30013 Cerebellum II 1.49e+01
c	832	14	9.5	304	8	AA542964	ni92h03.s1 NCI_CGAP_Pr 1.49e+01
	833	14	9.5	308	5	R48730	yj68a12.r1 Homo sapien 1.49e+01
c	834	14	9.5	309	22	AA192440	zp91f01.s1 Stratagene 1.49e+01
c	835	14	9.5	309	5	T55523	yb38f03.r1 Homo sapien 1.49e+01
c	836	14	9.5	311	27	B80100	CIT-HSP-2042I15.TR CIT 1.49e+01
	837	14	9.5	316	6	AA386031	EST99769 Pancreas tumo 1.49e+01
	838	14	9.5	319	21	W56023	zd14a09.r1 Soares feta 1.49e+01
	839	14	9.5	321	8	AA507095	nh41c05.s1 NCI_CGAP_Pr 1.49e+01
	840	14	9.5	325	25	AA373600	EST85638 HSC172 cells 1.49e+01
	841	14	9.5	325	17	AI071697	UI-R-C2-nc-b-10-0-UI.s 1.49e+01
	842	14	9.5	327	14	AA946516	EST202015 Normalized r 1.49e+01
c	843	14	9.5	328	16	H50579	yp83f09.s1 Homo sapien 1.49e+01
c	844	14	9.5	331	24	AA320670	EST23057 Adipose tissu 1.49e+01
c	845	14	9.5	333	17	AA407203	EST02188 Mouse 7.5 dpc 1.49e+01
	846	14	9.5	337	8	AA253289	zr53f12.r1 Soares NhHM 1.49e+01
c	847	14	9.5	337	8	AA467992	nc72a10.s1 NCI_CGAP_Pr 1.49e+01
	848	14	9.5	338	6	W87890	zh66d06.r1 Soares feta 1.49e+01
	849	14	9.5	338	26	AQ022983	HS_2185_A2_A12_MR CIT 1.49e+01
	850	14	9.5	339	25	AA373887	EST85980 HSC172 cells 1.49e+01
c	851	14	9.5	340	8	AA280567	zt10c11.r1 NCI_CGAP_GC 1.49e+01
c	852	14	9.5	341	5	T93262	ye25c04.s1 Homo sapien 1.49e+01
c	853	14	9.5	344	24	AA325322	EST28312 Cerebellum II 1.49e+01
c	854	14	9.5	345	9	AA228686	nc46f09.s1 NCI_CGAP_Pr 1.49e+01
c	855	14	9.5	345	5	R26362	yh49h05.r1 Homo sapien 1.49e+01
c	856	14	9.5	345	16	H47851	yp76g11.s1 Homo sapien 1.49e+01
	857	14	9.5	349	20	ATTS5044	A. thaliana transcribe 1.49e+01
	858	14	9.5	349	17	AA408325	EST02804 Mouse 7.5 dpc 1.49e+01

c	859	14	9.5	350	5	R26331	yh49d05.r1 Homo sapien 1.49e+01
	860	14	9.5	356	17	AI072588	UI-R-C2-nn-b-11-0-UI.s 1.49e+01
	861	14	9.5	358	15	AI012622	EST207073 Normalized r 1.49e+01
	862	14	9.5	358	10	AA477915	zu39b03.r1 Soares ovar 1.49e+01
	863	14	9.5	360	6	HSPD04993	H.sapiens EST sequence 1.49e+01
	864	14	9.5	360	20	ATTS2456	A. thaliana transcribe 1.49e+01
c	865	14	9.5	363	28	AQ107709	HS_3109_A1_P03_T7 CIT 1.49e+01
c	866	14	9.5	367	29	AQ142381	HS_3073_A2_H10_MR CIT 1.49e+01
	867	14	9.5	369	9	AA548402	nj14f01.s1 NCI_CGAP_Pr 1.49e+01
c	868	14	9.5	369	17	AU019313	Mus musculus 8-cell em 1.49e+01
	869	14	9.5	371	7	HUM300H11B	Human aorta cDNA 5'-en 1.49e+01
c	870	14	9.5	373	25	AA364061	EST74616 Pineal gland 1.49e+01
c	871	14	9.5	373	9	C47848	C.elegans cDNA clone y 1.49e+01
	872	14	9.5	374	11	AA734112	vv22f03.r1 Stratagene 1.49e+01
c	873	14	9.5	378	13	AA199610	zq74d03.s1 Stratagene 1.49e+01
c	874	14	9.5	380	24	AA326303	EST29447 Cerebellum II 1.49e+01
c	875	14	9.5	382	24	AA326254	EST29447 Cerebellum II 1.49e+01
c	876	14	9.5	383	26	B71934	RPCI11-13K7.TV RPCI11 1.49e+01
c	877	14	9.5	385	5	R24535	yh36g02.s1 Homo sapien 1.49e+01
	878	14	9.5	388	6	C01552	HUMGS0008548, Human Ge 1.49e+01
	879	14	9.5	389	14	AA965081	UI-R-C0-gz-b-03-0-UI.s 1.49e+01
	880	14	9.5	389	28	AQ117034	HS_3027_A2_D12_MR CIT 1.49e+01
c	881	14	9.5	390	15	AI048787	ub32e07.r1 Soares 2NbM 1.49e+01
c	882	14	9.5	391	9	AA011919	TgESTzz10e04.r1 TgME49 1.49e+01
	883	14	9.5	400	27	AQ006533	CIT-HSP-2193A15.TF CIT 1.49e+01
c	884	14	9.5	400	5	T97622	ye55g02.s1 Homo sapien 1.49e+01
c	885	14	9.5	404	8	AA562852	vl54e10.r1 Knowles Sol 1.49e+01
	886	14	9.5	405	5	T66821	ya50b07.s2 Homo sapien 1.49e+01
	887	14	9.5	409	15	AA988595	or79g10.s1 NCI_CGAP_Lu 1.49e+01
c	888	14	9.5	411	6	AA004811	zh94a05.s1 Soares feta 1.49e+01
	889	14	9.5	412	8	AA546371	vk06g11.r1 Knowles Sol 1.49e+01
	890	14	9.5	413	5	T59641	yc13d03.s1 Homo sapien 1.49e+01
c	891	14	9.5	414	27	AQ031787	HS_2228_B1_B03_MF CIT 1.49e+01
	892	14	9.5	415	19	H72925	ys05a04.s1 Homo sapien 1.49e+01
c	893	14	9.5	416	24	AA325239	EST28192 Cerebellum II 1.49e+01
c	894	14	9.5	416	18	AI146274	qb87b02.x1 Soares_feta 1.49e+01
	895	14	9.5	417	29	AQ246782	HS_2197_A2_F04_MR CIT 1.49e+01
c	896	14	9.5	418	27	AQ041942	CIT-HSP-2136D2.TF CIT- 1.49e+01
c	897	14	9.5	418	5	R32003	yh63a07.s1 Homo sapien 1.49e+01
c	898	14	9.5	421	28	AQ191647	HS_3225_B1_G04_MR CIT 1.49e+01
	899	14	9.5	423	16	H47403	yp74c05.r1 Homo sapien 1.49e+01
	900	14	9.5	426	15	AI007710	EST202161 Normalized r 1.49e+01
	901	14	9.5	430	18	AI176070	EST219648 Normalized r 1.49e+01
	902	14	9.5	431	5	R18156	yf97d12.r1 Homo sapien 1.49e+01
c	903	14	9.5	431	18	AI152127	ud20b12.r1 Soares 2NbM 1.49e+01
	904	14	9.5	438	27	AQ088871	HS_3005_B1_E12_MF CIT 1.49e+01
c	905	14	9.5	439	21	AA044016	zk50e01.s1 Soares preg 1.49e+01
c	906	14	9.5	441	17	AI015113	ot90c11.s1 Soares_tota 1.49e+01
	907	14	9.5	442	12	AA732934	zg78c04.s1 Soares feta 1.49e+01
	908	14	9.5	442	11	AA727368	yu95g04.r1 Stratagene 1.49e+01
	909	14	9.5	442	27	AQ030329	HS_2177_B1_A11_MR CIT 1.49e+01

c	910	14	9.5	442	9	AA605829	fa19h08.s1 Ekkerearly 1.49e+01
c	911	14	9.5	442	16	R40701	yf73e11.s1 Homo sapien 1.49e+01
c	912	14	9.5	451	5	T62937	yb99f09.s1 Homo sapien 1.49e+01
c	913	14	9.5	451	15	AI017302	ov31e07.x1 Soares_test 1.49e+01
c	914	14	9.5	453	17	AI127118	qb98b03.x1 Soares_feta 1.49e+01
c	915	14	9.5	453	6	AA099281	zk84c11.s1 Soares preg 1.49e+01
c	916	14	9.5	454	9	N60451	TgESTzy08c09.r1 TgRH T 1.49e+01
	917	14	9.5	455	10	AA429275	zv50c02.s1 Soares ovar 1.49e+01
	918	14	9.5	456	12	AA769035	oa80g05.s1 NCI_CGAP_GC 1.49e+01
	919	14	9.5	458	27	AQ057996	CIT-HSP-2343O11.TF CIT 1.49e+01
c	920	14	9.5	459	13	AA872571	oa16g10.s1 NCI_CGAP_GC 1.49e+01
	921	14	9.5	463	19	N33608	yv22d09.s1 Homo sapien 1.49e+01
c	922	14	9.5	464	11	AA703219	zi68g05.s1 Soares feta 1.49e+01
c	923	14	9.5	465	13	AA856249	vw98e05.r1 Soares 2NbM 1.49e+01
	924	14	9.5	467	27	B88143	RPCI11-18O15.TPC RPCI1 1.49e+01
	925	14	9.5	468	23	AA073953	mm99h05.r1 Stratagene 1.49e+01
c	926	14	9.5	468	21	N23938	zb74b07.s1 Soares feta 1.49e+01
	927	14	9.5	469	19	N24895	yx16g09.r1 Homo sapien 1.49e+01
	928	14	9.5	469	27	B83482	RPCI11-15E11.TVB RPCI1 1.49e+01
	929	14	9.5	470	11	H76551	18256 Lambda-PRL2 Arab 1.49e+01
	930	14	9.5	471	18	AI138815	qd98g12.x1 Soares_test 1.49e+01
c	931	14	9.5	471	19	N39247	yy50f01.s1 Homo sapien 1.49e+01
	932	14	9.5	472	22	AA210379	mu72e06.r1 Soares mous 1.49e+01
c	933	14	9.5	472	14	AA948513	on53d09.s1 NCI_CGAP_Co 1.49e+01
c	934	14	9.5	473	13	AA839397	vw49e07.r1 Soares mous 1.49e+01
c	935	14	9.5	475	6	AA001748	zh86h07.s1 Soares feta 1.49e+01
c	936	14	9.5	478	13	AA630299	ac08e02.s1 Stratagene 1.49e+01
	937	14	9.5	480	8	AA291031	zs45f03.r1 NCI_CGAP_GC 1.49e+01
	938	14	9.5	480	26	B38750	HS-1048-A2-C10-MF.abi 1.49e+01
c	939	14	9.5	482	16	R55562	yg88a07.r1 Homo sapien 1.49e+01
	940	14	9.5	483	17	AI032168	ou94c03.s1 NCI_CGAP_Ki 1.49e+01
	941	14	9.5	487	26	B37125	HS-1042-B1-H04-MF.abi 1.49e+01
c	942	14	9.5	488	16	R70419	yj82b07.r1 Homo sapien 1.49e+01
	943	14	9.5	494	26	FR0020879	F.rubripes GSS sequenc 1.49e+01
	944	14	9.5	496	22	AA203481	zx55f02.r1 Soares feta 1.49e+01
c	945	14	9.5	497	7	AA511245	vh78a04.r1 Knowles Sol 1.49e+01
	946	14	9.5	499	20	AA000592	mg25e11.r1 Soares mous 1.49e+01
c	947	14	9.5	502	7	AA442693	zv72a12.r1 Soares tota 1.49e+01
c	948	14	9.5	504	14	AA969336	on58a07.s1 Soares_NFL_ 1.49e+01
	949	14	9.5	510	20	W17276	zb17h04.r1 Soares feta 1.49e+01
c	950	14	9.5	511	19	N23026	yx65g10.s1 Homo sapien 1.49e+01
	951	14	9.5	512	26	B76551	RPCI11-15L23.TP RPCI11 1.49e+01
	952	14	9.5	519	16	H08016	y191c06.r1 Homo sapien 1.49e+01
	953	14	9.5	521	17	AI120248	ub68a10.r1 Soares mous 1.49e+01
c	954	14	9.5	522	28	AQ120475	HS_3030_A1_F12_MF CIT 1.49e+01
c	955	14	9.5	527	14	AA925674	UI-R-A1-er-c-06-0-UI.s 1.49e+01
c	956	14	9.5	529	23	AA117688	mo64d05.r1 Stratagene 1.49e+01
	957	14	9.5	530	26	AQ023920	CpGR0253A Cryptosporid 1.49e+01
	958	14	9.5	537	8	AA397449	nc65e09.r1 NCI_CGAP_Pr 1.49e+01
	959	14	9.5	538	15	AI008547	EST202998 Normalized r 1.49e+01
	960	14	9.5	540	27	B94536	CIT-HSP-2164D6.TR CIT- 1.49e+01

```
c  961     14     9.5     540 18  AI169276     EST215111 Normalized r  1.49e+01
   962     14     9.5     542 23  AA260602     mx76h10.r1 Soares mous  1.49e+01
c  963     14     9.5     545 20  W37888       zc13d06.s1 Soares para  1.49e+01
c  964     14     9.5     550 15  AI029933     UI-R-C0-ja-c-11-0-UI.s  1.49e+01
c  965     14     9.5     551 22  AA181361     zp42c03.s1 Stratagene   1.49e+01
   966     14     9.5     555 12  AA797369     vw27b04.r1 Soares mous  1.49e+01
   967     14     9.5     556 24  AA288179     vb13e11.r1 Soares mous  1.49e+01
   968     14     9.5     559  8  AA546162     vk62d08.r1 Stratagene   1.49e+01
c  969     14     9.5     567 13  AA858511     UI-R-E0-bo-d-08-0-UI.s  1.49e+01
c  970     14     9.5     572 10  AA595259     no33c10.s1 NCI_CGAP_Pr  1.49e+01
c  971     14     9.5     576 14  C90440       Dictyostelium discoide  1.49e+01
   972     14     9.5     586 26  B72178       RPCI11-7K14.TP RPCI11   1.49e+01
   973     14     9.5     587 12  AA718944     ah45e06.s1 Soares test  1.49e+01
c  974     14     9.5     588 27  AQ025269     ep(3)3088 Drosophila m  1.49e+01
c  975     14     9.5     589 26  FR0004170    F.rubripes GSS sequenc  1.49e+01
c  976     14     9.5     596 12  AA264571     LD08203.3prime LD Dros  1.49e+01
c  977     14     9.5     603 26  B55721       CIT-HSP-2003G21.TR CIT  1.49e+01
   978     14     9.5     605 26  FR0027990    F.rubripes GSS sequenc  1.49e+01
   979     14     9.5     606 26  FR0020622    F.rubripes GSS sequenc  1.49e+01
c  980     14     9.5     606 13  AA490011     ab02a02.s1 Stratagene   1.49e+01
   981     14     9.5     610 27  B95697       CIT-HSP-2169I22.TR CIT  1.49e+01
c  982     14     9.5     619 12  AA759577     vw55c03.r1 Soares mous  1.49e+01
c  983     14     9.5     628 26  B27428       T7K22TR TAMU Arabidops  1.49e+01
c  984     14     9.5     641 26  B54599       CIT-HSP-34704.TVC CIT-  1.49e+01
   985     14     9.5     650 27  AQ030074     RPCI11-39I8.TV RPCI11   1.49e+01
   986     14     9.5     651 28  AQ199101     RPCI11-45C13.TJ RPCI11  1.49e+01
c  987     14     9.5     654 15  AI055708     coau0004N20 Cotton Bol  1.49e+01
   988     14     9.5     668  8  AA550635     1824m3 gmbPfHB3.1, G.   1.49e+01
c  989     14     9.5     669 26  B73239       RPCI11-13G13.TV RPCI11  1.49e+01
   990     14     9.5     677 21  W73855       zd52f10.r1 Soares feta  1.49e+01
c  991     14     9.5     678 26  AG005729     Homo sapiens genomic D  1.49e+01
c  992     14     9.5     708 18  AI076597     oz31a05.x1 Soares_tota  1.49e+01
c  993     14     9.5     729  6  AA149524     zl28f08.r1 Soares preg  1.49e+01
   994     14     9.5     740 29  AQ244894     HS_2061_A2_G03_T7 CIT   1.49e+01
c  995     14     9.5     749 20  W22496       68F4 Human retina cDNA  1.49e+01
   996     14     9.5     767 26  B19064       T5Q16-Sp6 TAMU Arabido  1.49e+01
c  997     14     9.5     779 15  AI049017     uc85d08.y1 Sugano mous  1.49e+01
   998     14     9.5     814 27  AG010456     Homo sapiens genomic D  1.49e+01
   999     14     9.5     814 26  AG004330     Homo sapiens genomic D  1.49e+01
c 1000     14     9.5     860 26  B18910       T21G19-Sp6 TAMU Arabid  1.49e+01

                              ALIGNMENTS

RESULT    1
LOCUS       HUMGS00783     147 bp    mRNA            EST       17-JUN-1996
DEFINITION  Human HL60 3'directed MboI cDNA, HUMGS00783, clone mm0468, mRNA
            sequence.
ACCESSION   D19827
NID         g500143
KEYWORDS    EST; EST(expressed sequence tag); HL60; gene signature(GS);
```



```
            granulocyte; macrophage; monocyte; promyelocyte.
SOURCE      Homo sapiens adult female promyelocyte cell_line:HL60 cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (sites)
  AUTHORS   Murakawa,K., Matsubara,K., Fukushima,A., Yoshii,J. and Okubo,K.
  TITLE     Chromosomal assignments of 3'-directed partial cDNA sequences
            representing novel genes expressed in granulocytoid cells
  JOURNAL   Genomics 23 (2), 379-389 (1994)
  MEDLINE   95137584
REFERENCE   2  (sites)
  AUTHORS   Okubo,K., Itoh,K., Fukushima,A., Yoshii,J. and Matsubara,K.
  TITLE     Monitoring cell physiology by expression profiles and discovering
            cell type-specific genes by compiled expression profiles
  JOURNAL   Genomics 30 (2), 178-186 (1995)
  MEDLINE   96163870
REFERENCE   3  (bases 1 to 147)
  AUTHORS   Okubo,K., Fukushima,A., Yoshii,J., Niiyama,T., Kojima,Y.,
            Yoshinari,H., Arimoto,J. and Matsubara,K.
  TITLE     Gene expression of human promyelocytic cell line HL60before and
            after induction of differentiation. A new application of
            3'directedcDNA sequencing
  JOURNAL   Unpublished (1993)
COMMENT     Submitted (10-Sep-1993) to DDBJ by:
            Kousaku Okubo
            Institute for Molecular and Cellular Biology
            Osaka University
            3-1, Yamadaoka
            Suita, Osaka, 565
            Japan
            Phone: 06-877-5111
            Fax  : 06-875-1922
            DDBJ CHROMOSOMAL_ASSIGNMENT
                 :method='PCR WITH HYBRID CELL PANEL'
                 ::primer_f='CCCAGGGTTTCTTTGTCTTA'
                 ::primer_r='TCTGTAACTTCGGTTCCAAG'
                 ::background='mouse:1,2,5,6,7,8,10,11,12,14,15,16,17,18,19,
                              20,21,X'
                 ::background='chinese hamster:3,4,9,13,22,Y'
                 :chromosome='6'
                 :chromosome='12'.
FEATURES            Location/Qualifiers
     source         1. .147
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /cell_line="HL60"
                    /cell_type="promyelocyte"
                    /dev_stage="adult"
                    /sex="female"
```



```
BASE COUNT       39 a     39 c     30 g     39 t
ORIGIN

  Query Match          100.0%;  Score 147;  DB 7;  Length 147;
  Best Local Similarity 100.0%;  Pred. No. 0.00e+00;
  Matches  147;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 GATCCCCAGGGTTTCTTTGTCTTATTTATGGAGAAAAACCGGTCACTTTGTCCAGCGCAC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      1 GATCCCCAGGGTTTCTTTGTCTTATTTATGGAGAAAAACCGGTCACTTTGTCCAGCGCAC 60

Db     61 TGTGAGGCCCCCACTCAGGCCAGCCCTGGCCCCCCCCTTGGTACTTGGAACCGAAGTTACA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     61 TGTGAGGCCCCCACTCAGGCCAGCCCTGGCCCCCCCCTTGGTACTTGGAACCGAAGTTACA 120

Db    121 GATTATATTAAAATAATAATGTACAAA 147
          |||||||||||||||||||||||||||
Qy    121 GATTATATTAAAATAATAATGTACAAA 147


RESULT    2
LOCUS       AI018794       349 bp    mRNA            EST       16-JUN-1998
DEFINITION  ov32g09.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1639072
            3', mRNA sequence.
ACCESSION   AI018794
NID         g3232592
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 349)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40m13 fwd. ET from Amersham.
FEATURES            Location/Qualifiers
     source         1. .349
```



```
                    /organism="Homo sapiens"
                    /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                    polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                    was prepared from mRNA obtained from Clontech
                    Laboratories, Inc., and primed with  a Not I - oligo(dT)
                    primer [5'
                    TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3']."
                    Double-stranded cDNA was ligated to Eco RI adaptors
                    (Pharmacia), digested with Not I and cloned into the Not I
                    and Eco RI sites of the modified pT7T3 vector. Library
                    went through one round of normalization to Cot5, and was
                    constructed  by Bento Soares and M. Fatima Bonaldo."
                    /db_xref="taxon:9606"
                    /clone="IMAGE:1639072"
                    /clone_lib="Soares_testis_NHT"
                    /sex="male"
                    /lab_host="DH10B"
BASE COUNT      114 a     70 c     74 g     91 t
ORIGIN

  Query Match          100.0%;  Score 147;  DB 15;  Length 349;
  Best Local Similarity 100.0%;  Pred. No. 0.00e+00;
  Matches  147;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      4 TTTGTACATTATTATTTTAATATAATCTGTAACTTCGGTTCCAAGTACCAAGGGGGGGCC 63
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp    147 TTTGTACATTATTATTTTAATATAATCTGTAACTTCGGTTCCAAGTACCAAGGGGGGGCC 88

Db     64 AGGGCTGGCCTGAGTGGGGGCCTCACAGTGCGCTGGACAAAGTGACCGGTTTTTCTCCAT 123
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     87 AGGGCTGGCCTGAGTGGGGGCCTCACAGTGCGCTGGACAAAGTGACCGGTTTTTCTCCAT 28

Db    124 AAATAAGACAAAGAAACCCTGGGGATC 150
          |||||||||||||||||||||||||||
Cp     27 AAATAAGACAAAGAAACCCTGGGGATC 1


RESULT    3
LOCUS       AI017469       478 bp    mRNA            EST       27-AUG-1998
DEFINITION  ou23c08.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
            IMAGE:1627118 3', mRNA sequence.
ACCESSION   AI017469
NID         g3231805
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 478)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
```

```
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 566   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 386.
FEATURES              Location/Qualifiers
     source           1. .478
                      /organism="Homo sapiens"
                      /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                      a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                      Equal amounts of plasmid DNA from three normalized
                      libraries (fetal lung NbHL19W, testis NHT, and B-cell
                      NCI_CGAP_GCB1) were mixed, and ss circles were made in
                      vitro. Following HAP purification, this DNA was used as
                      tracer in a subtractive hybridization reaction. The driver
                      was PCR-amplified cDNAs from pools of 5,000 clones made
                      from the same 3 libraries. The pools consisted of
                      I.M.A.G.E. clones 297480-302087, 682632-687239,
                      726408-728711, and 729096-731399. Subtraction by Bento
                      Soares and M. Fatima Bonaldo."
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1627118"
                      /clone_lib="Soares_NFL_T_GBC_S1"
                      /lab_host="DH10B"
BASE COUNT      147 a    109 c    117 g    105 t
ORIGIN

  Query Match          100.0%;  Score 147;  DB 17;  Length 478;
  Best Local Similarity 100.0%;  Pred. No. 0.00e+00;
  Matches  147;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      2 TTTGTACATTATTATTTTAATATAATCTGTAACTTCGGTTCCAAGTACCAAGGGGGGGCC 61
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp    147 TTTGTACATTATTATTTTAATATAATCTGTAACTTCGGTTCCAAGTACCAAGGGGGGGCC 88

Db     62 AGGGCTGGCCTGAGTGGGGGCCTCACAGTGCGCTGGACAAAGTGACCGGTTTTTCTCCAT 121
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     87 AGGGCTGGCCTGAGTGGGGGCCTCACAGTGCGCTGGACAAAGTGACCGGTTTTTCTCCAT 28

Db    122 AAATAAGACAAAGAAACCCTGGGGATC 148
          |||||||||||||||||||||||||||
Cp     27 AAATAAGACAAAGAAACCCTGGGGATC 1


RESULT    4
```



```
LOCUS       AA303166      385 bp    mRNA            EST       18-APR-1997
DEFINITION  EST13228 Thymus tumor I Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION   AA303166
NID         g1955497
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 385)
  AUTHORS   Adams,M.D., Soares,M.B., Kerlavage,A.R., Fields,C. and Venter,J.C.
  TITLE     Rapid cDNA sequencing (expressed sequence tags) from a
            directionally cloned human infant brain cDNA library
  JOURNAL   Nature Genet. 4, 373-380 (1993)
  MEDLINE   94004965
COMMENT     Other_ESTs: THC177288
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES              Location/Qualifiers
     source           1. .385
                      /organism="Homo sapiens"
                      /note="Organ: thymus; Vector: pBluescript SK-; Site_1:
                      EcoRI; Site_2: XhoI"
                      /db_xref="ATCC (inhost):192577"
                      /db_xref="taxon:9606"
                      /clone_lib="Thymus tumor I"
                      /dev_stage="adult"
     mRNA             <1. .>385
BASE COUNT      110 a     78 c     67 g    125 t      5 others
ORIGIN

  Query Match          99.3%;  Score 146;  DB 24;  Length 385;
  Best Local Similarity 99.3%;  Pred. No. 0.00e+00;
  Matches  146;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db     12 GATCCCCAGGGTTTCTTTGTNTTATTTATGGAGAAAAACCGGTCACTTTGTCCAGCGCAC 71
          ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Qy      1 GATCCCCAGGGTTTCTTTGTCTTATTTATGGAGAAAAACCGGTCACTTTGTCCAGCGCAC 60

Db     72 TGTGAGGCCCCCACTCAGGCCAGCCCTGGCCCCCCCTTGGTACTTGGAACCGAAGTTACA 131
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```



```
Qy     61 TGTGAGGCCCCCACTCAGGCCAGCCCTGGCCCCCCCTTGGTACTTGGAACCGAAGTTACA 120

Db    132 GATTATATTAAAATAATAATGTACAAA 158
          |||||||||||||||||||||||||||
Qy    121 GATTATATTAAAATAATAATGTACAAA 147


RESULT    5
LOCUS       AA909023      399 bp    mRNA            EST       13-APR-1998
DEFINITION  o111f10.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
            IMAGE:1523179 3', mRNA sequence.
ACCESSION   AA909023
NID         g3048428
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 399)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 387.
FEATURES              Location/Qualifiers
     source           1. .399
                      /organism="Homo sapiens"
                      /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                      a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                      Equal amounts of plasmid DNA from three normalized
                      libraries (fetal lung NbHL19W, testis NHT, and B-cell
                      NCI_CGAP_GCB1) were mixed, and ss circles were made in
                      vitro. Following HAP purification, this DNA was used as
                      tracer in a subtractive hybridization reaction. The driver
                      was PCR-amplified cDNAs from pools of 5,000 clones made
                      from the same 3 libraries. The pools consisted of
                      I.M.A.G.E. clones 297480-302087, 682632-687239,
                      726408-728711, and 729096-731399. Subtraction by Bento
                      Soares and M. Fatima Bonaldo."
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1523179"
                      /clone_lib="Soares_NFL_T_GBC_S1"
                      /lab_host="DH10B"
BASE COUNT      132 a     92 c     82 g     93 t
```



```
ORIGIN

  Query Match          57.8%;  Score 85;  DB 14;  Length 399;
  Best Local Similarity 98.6%;  Pred. No. 1.69e-154;
  Matches  145;  Conservative    0;  Mismatches   2;  Indels   0;  Gaps   0;

Db      1 TTTGTACATTATTATTTTAATATAATCTGTAACTTCGGTTCCAAGTACCAAGGGGGGGCC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp    147 TTTGTACATTATTATTTTAATATAATCTGTAACTTCGGTTCCAAGTACCAAGGGGGGGCC 88

Db     61 AGGGCTGGCTCGAGTGGGGGCCTCACAGTGCGCTGGACAAAGTGACCGGTTTTTCTCCAT 120
          |||||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     87 AGGGCTGGCCTGAGTGGGGGCCTCACAGTGCGCTGGACAAAGTGACCGGTTTTTCTCCAT 28

Db    121 AAATAAGACAAAGAAACCCTGGGGATC 147
          |||||||||||||||||||||||||||
Cp     27 AAATAAGACAAAGAAACCCTGGGGATC 1


RESULT    6
LOCUS       H41523        390 bp    mRNA            EST       31-JUL-1995
DEFINITION  yn90g09.r1 Homo sapiens cDNA clone 175744 5'.
ACCESSION   H41523
NID         g917575
KEYWORDS    EST.
SOURCE      Homo sapiens clone=175744 library=Soares adult brain N2b5HB55Y
            vector=pT7T3D (Pharmacia) with a modified polylinker host=DH10B
            (ampicillin resistant) primer=M13RP1 Rsitel=Not I Rsite2=Eco RI
            55-year old male. 1st strand cDNA was primed with a Not I -
            oligo(dT) primer [5'
            TGTTACCAATCTGAAGTGGGAGCGGCCGCGCTTTTTTTTTTTTTTTTTT 3'],
            double-stranded cDNA was size selected, ligated to Eco RI adapters
            (Pharmacia), digested with Not I and cloned into the Not I and Eco
            RI sites of a modified pT7T3 vector (Pharmacia). Library went
            through one round of normalization to a Cot = 53. Library
            constructed by Bento Soares and M.Fatima Bonaldo. The adult brain
            RNA was provided by Dr. Donald H. Gilden. Tissue was acquired 17-18
            hours after death which occurred in consequence of a ruptured
            aortic aneurysm. RNA was prepared from a pool of tissues
            representing the following areas of the brain: frontal, parietal,
            temporal and occipital cortex from the left and right hemispheres,
            subcortical white matter, basal ganglia, thalamus, cerebellum,
            midbrain, pons and medulla.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 390)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
```

```
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 329
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES             Location/Qualifiers
     source          1. .390
                     /organism="Homo sapiens"
                     /clone="175744"
BASE COUNT       93 a      79 c      78 g     134 t      6 others
ORIGIN

  Query Match           56.5%;  Score 83;  DB 16;  Length 390;
  Best Local Similarity 100.0%;  Pred. No. 1.41e-149;
  Matches    83;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    221 TTCTTTGTCTTATTTATGGAGAAAAACCGGTCACTTTGTCCAGCGCACTGTGAGGCCCCC 280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     13 TTCTTTGTCTTATTTATGGAGAAAAACCGGTCACTTTGTCCAGCGCACTGTGAGGCCCCC 72

Db    281 ACTCAGGCCAGCCCCTGGCCCCCC 303
          |||||||||||||||||||||||||
Qy     73 ACTCAGGCCAGCCCCTGGCCCCCC 95


RESULT    7
LOCUS       AA972448      313 bp    mRNA            EST       07-JUL-1998
DEFINITION  op41b02.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
            IMAGE:1579371 3', mRNA sequence.
ACCESSION   AA972448
NID         g3145212
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 313)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
```



```
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 419   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .313
                     /organism="Homo sapiens"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from three normalized
                     libraries (fetal lung NbHL19W, testis NHT, and B-cell
                     NCI_CGAP_GCB1) were mixed, and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR-amplified cDNAs from pools of 5,000 clones made
                     from the same 3 libraries. The pools consisted of
                     I.M.A.G.E. clones 297480-302087, 682632-687239,
                     726408-728711, and 729096-731399. Subtraction by Bento
                     Soares and M. Fatima Bonaldo."
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1579371"
                     /clone_lib="Soares_NFL_T_GBC_S1"
                     /lab_host="DH10B"
BASE COUNT      104 a      59 c      70 g      80 t
ORIGIN

  Query Match           55.8%;  Score 82;  DB 15;  Length 313;
  Best Local Similarity 98.6%;  Pred. No. 4.02e-147;
  Matches   142;  Conservative    0;  Mismatches   2;  Indels   0;  Gaps   0;

Db      1 GTACATTATTATTTTAATATAATCTGTAACTTCGGTTCCAAGTACCAAGGGGGGGCCAGG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp    144 GTACATTATTATTTTAATATAATCTGTAACTTCGGTTCCAAGTACCAAGGGGGGGCCAGG 85

Db     61 GCTGGCTCGAGTGGGGGCCTCACAGTGCGCTGGACAAAGTGACCGGTTTTTCTCCATAAA 120
          ||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     84 GCTGGCCTGAGTGGGGGCCTCACAGTGCGCTGGACAAAGTGACCGGTTTTTCTCCATAAA 25

Db    121 TAAGACAAAGAAACCCTGGGGATC 144
          |||||||||||||||||||||||||
Cp     24 TAAGACAAAGAAACCCTGGGGATC 1


RESULT    8
LOCUS       AA557377      273 bp    mRNA            EST       09-SEP-1997
```



```
DEFINITION  nl81f07.s1 NCI_CGAP_Br2 Homo sapiens cDNA clone IMAGE:1057093, mRNA
            sequence.
ACCESSION   AA557377
NID         g2327854
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 273)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 2819   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 240.
FEATURES             Location/Qualifiers
     source          1. .273
                     /organism="Homo sapiens"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; 1st strand cDNA was prepared from pooled bulk
                     breast tumor tissue, and was then primed with a Not I -
                     oligo(dT) primer. Double-stranded cDNA was ligated to Eco
                     RI adaptors (Pharmacia), digested with Not I and cloned
                     into the Not I and Eco RI sites of the modified pT7T3
                     vector.  This library is the normalized version of
                     NCI_CGAP_Br1.1.  Library was constructed by Bento Soares
                     and M. Fatima Bonaldo."
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1057093"
                     /clone_lib="NCI_CGAP_Br2"
                     /sex="female, pooled"
                     /tissue_type="breast"
                     /lab_host="DH10B"
     mRNA            <1. .>273
BASE COUNT       97 a      53 c      55 g      68 t
ORIGIN
```



```
  Query Match           49.7%;  Score 73;  DB 9;  Length 273;
  Best Local Similarity 100.0%;  Pred. No. 3.64e-125;
  Matches    73;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     70 TGGGGGCCTCACAGTGCGCTGGACAAAGTGACCGGTTTTTCTCCATAAATAAGACAAAGA 129
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     73 TGGGGGCCTCACAGTGCGCTGGACAAAGTGACCGGTTTTTCTCCATAAATAAGACAAAGA 14

Db    130 AACCCTGGGGATC 142
          |||||||||||||
Cp     13 AACCCTGGGGATC 1


RESULT    9
LOCUS       AA923630      318 bp    mRNA            EST       29-APR-1998
DEFINITION  nq98c09.s1 NCI_CGAP_Co10 Homo sapiens cDNA clone IMAGE:1160368 3',
            mRNA sequence.
ACCESSION   AA923630
NID         g3070939
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 318)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck,
            M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 422   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 269.
FEATURES             Location/Qualifiers
     source          1. .318
                     /organism="Homo sapiens"
                     /note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; 1st strand cDNA was prepared from
```

```
                    RER+ colon tumor, and was then primed with a Not I -
                    oligo(dT) primer. Double-stranded cDNA was ligated to Eco
                    RI adaptors (Pharmacia), digested with Not I and cloned
                    into the Not I and Eco RI sites of the modified pT7T3
                    vector.  Library is normalized. Library was constructed by
                    Bento Soares and M. Fatima Bonaldo (M-Soares4)."
                    /db_xref="taxon:9606"
                    /clone="IMAGE:1160368"
                    /clone_lib="NCI_CGAP_Col0"
                    /tissue_type="colon tumor RER+"
                    /lab_host="DH10B"
BASE COUNT      105 a      61 c      70 g      82 t
ORIGIN

  Query Match           49.7%;  Score 73;  DB 14;  Length 318;
  Best Local Similarity 100.0%;  Pred. No. 3.64e-125;
  Matches    73;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      72 TGGGGGCCTCACAGTGCGCTGGACAAAGTGACCGGTTTTTCTCCATAAATAAGACAAAGA 131
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp      73 TGGGGGCCTCACAGTGCGCTGGACAAAGTGACCGGTTTTTCTCCATAAATAAGACAAAGA 14

Db     132 AACCCTGGGGATC 144
           |||||||||||||
Cp      13 AACCCTGGGGATC 1


RESULT   10
LOCUS       AA862961      232 bp    mRNA            EST       11-MAR-1998
DEFINITION  og98f02.s1 NCI_CGAP_Kid3 Homo sapiens cDNA clone IMAGE:1456347 3',
            mRNA sequence.
ACCESSION   AA862961
NID         g2955440
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 232)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
```



```
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40m13 fwd. ET from Amersham.
FEATURES              Location/Qualifiers
     source           1. .232
                      /organism="Homo sapiens"
                      /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                      a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                      strand cDNA was primed with a Not I - oligo(dT) primer,
                      double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia),  digested with Not I and cloned into the Not
                      I and Eco RI sites  of the modified pT7T3 vector.  mRNA
                      source: 2 pooled kidneys. Library went through one round
                      of normalization. Library constructed by Bento Soares and
                      M. Fatima Bonaldo."
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1456347"
                      /clone_lib="NCI_CGAP_Kid3"
                      /lab_host="DH10B"
BASE COUNT       68 a      45 c      57 g      62 t
ORIGIN

  Query Match           49.0%;  Score 72;  DB 13;  Length 232;
  Best Local Similarity 100.0%;  Pred. No. 9.60e-123;
  Matches    72;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      77 GGGGGCCTCACAGTGCGCTGGACAAAGTGACCGGTTTTTCTCCATAAATAAGACAAAGAA 136
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp      72 GGGGGCCTCACAGTGCGCTGGACAAAGTGACCGGTTTTTCTCCATAAATAAGACAAAGAA 13

Db     137 ACCCTGGGGATC 148
           ||||||||||||
Cp      12 ACCCTGGGGATC 1


RESULT   11
LOCUS       R59962        287 bp    mRNA            EST       24-MAY-1995
DEFINITION  yh11g12.s1 Homo sapiens cDNA clone 43010 3'.
ACCESSION   R59962
NID         g830657
KEYWORDS    EST.
SOURCE      human clone=43010 library=Soares infant brain 1NIB vector=Lafmid BA
            host=DH10B (ampicillin resistant) primer=SP6 Rsite1=Not I
            Rsite2=Hind III Whole brain from a 73 days post natal female. 1st
            strand cDNA was primed with a Not I - oligo(dT) primer [5'
            AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTT 3']; double-stranded
            cDNA was ligated to Hind III adaptors (Pharmacia), digested with
            Not I and directionally cloned into the Not I and Hind III sites of
```



```
            the Lafmid BA vector. Library went through one round of
            normalization. Library constructed by Bento Soares and M.Fatima
            Bonaldo.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 287)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            GDB: G00-415-551
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 142
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES              Location/Qualifiers
     source           1. .287
                      /organism="Homo sapiens"
                      /clone="43010"
BASE COUNT       93 a      48 c      55 g      88 t       3 others
ORIGIN

  Query Match           21.8%;  Score 32;  DB 16;  Length 287;
  Best Local Similarity  94.1%;  Pred. No. 2.92e-31;
  Matches    32;  Conservative    0;  Mismatches   2;  Indels   0;  Gaps   0;

Db     161 TTTGTACATNATTATTTTANTATAATCTGTAACT 194
           ||||||||| ||||||||| ||||||||||||||
Cp     147 TTTGTACATTATTATTTTAATATAATCTGTAACT 114


RESULT   12
LOCUS       HUMJ93B06B    212 bp    mRNA            EST       14-DEC-1995
DEFINITION  Human aorta cDNA 5'-end GEN-293B06, mRNA sequence.
ACCESSION   D79625
NID         g1179976
KEYWORDS    EST; EST(expressed sequence tag); Human aorta.
```



```
SOURCE      Homo sapiens aorta cDNA to mRNA, clone_lib:human aorta polyA+.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 212)
  AUTHORS   Fujiwara,T., Hirano,H., Katagiri,T., Kawai,A., Kuga,Y., Nagata,M.,
            Okuno,S., Ozaki,K., Shimizu,F., Shimada,Y., Shinomiya,H.,
            Takaichi,A., Takeda,S., Watanabe,T., Takahashi,E., Hirai,Y.,
            Maekawa,H., Shin,S. and Nakamura,Y.
  TITLE     Large-scale sequencing project at Otsuka GEN Research Institute
  JOURNAL   Unpublished (1995)
REFERENCE   2  (bases 1 to 212)
  AUTHORS   Fujiwara,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-NOV-1995) to the DDBJ/EMBL/GenBank databases. Tsutomu
            Fujiwara, Otsuka GEN Research Institute,Otsuka Pharmaceutical
            Co.,Ltd, 463-10 Kagasuno Kawauchi-cho, Tokushima, Tokushima 771-01,
            Japan
COMMENT     Submitted (7-Nov-1995) to DDBJ by:
            Tsutomu Fujiwara
            Otsuka GEN Research Institute
            Otsuka Pharmaceutical CO.,Ltd
            463-10 Kagasuno Kawauchi-cho
            Tokushima, Tokushima
            771-01
            Japan
            Phone:0886-65-2888
            Fax  :0886-37-1035.
FEATURES              Location/Qualifiers
     source           1. .212
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone_lib="human aorta polyA+"
                      /tissue_type="aorta"
BASE COUNT       79 a      33 c      31 g      68 t       1 others
ORIGIN

  Query Match           13.6%;  Score 20;  DB 7;  Length 212;
  Best Local Similarity 100.0%;  Pred. No. 1.75e-08;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     191 TATATTAAAATAATAATGTA 210
           ||||||||||||||||||||
Qy     124 TATATTAAAATAATAATGTA 143


RESULT   13
LOCUS       AA122818      310 bp    mRNA            EST       16-FEB-1997
DEFINITION  mr03b11.r1 Soares mouse 3NbMS Mus musculus cDNA clone 596349 5',
            mRNA sequence.
```

```
ACCESSION   AA122818
NID         g1681775
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Mus.
REFERENCE   1  (bases 1 to 310)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:361781
            Putative full length read
            vector to vector length is 311
            Seq primer: -28M13 rev2 from Amersham.
FEATURES                Location/Qualifiers
     source             1. .310
                        /organism="Mus musculus"
                        /strain="C57BL/6J"
                        /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                        polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                        was primed with a Not I - oligo(dT) primer [5'
                        TGTTACCAATCTGAAGTGGGAGCGGCCGCGCTGTTTTTTTTTTTTTTTTTTTTTTT
                        3']; double-stranded cDNA was ligated to Eco RI adaptors
                        (Pharmacia), digested with Not I and cloned into the Not I
                        and Eco RI sites of the modified pT7T3 vector. RNA
                        provided by Dr. Bertrand Jordan. Library went through
                        three rounds of normalization, and was constructed by
                        Bento Soares and M.Fatima Bonaldo."
                        /db_xref="taxon:10090"
                        /clone="596349"
                        /clone_lib="Soares mouse 3NbMS"
                        /sex="male"
                        /tissue_type="Spleen"
                        /dev_stage="4 weeks"
                        /lab_host="DH10B"
```



```
     mRNA            <1. .>310
BASE COUNT       86 a     55 c     57 g    112 t
ORIGIN

  Query Match          13.6%;  Score 20;  DB 23;  Length 310;
  Best Local Similarity 100.0%;  Pred. No. 1.75e-08;
  Matches   20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     67 TGTCCAGCGCACTGTGAGGC 86
          ||||||||||||||||||||
Qy     49 TGTCCAGCGCACTGTGAGGC 68


RESULT   14
LOCUS       AA924990      360 bp    mRNA            EST       21-APR-1998
DEFINITION  UI-R-A1-eh-b-12-0-UI.s1 UI-R-A1 Rattus norvegicus cDNA clone
            UI-R-A1-eh-b-12-0-UI 3' similar to gi|2075271|gb|AA415181|AA415181
            ve98e08.s1 Knowles Solter mouse 2 cell Mus musculus cDNA clone
            791078 5', mRNA sequence.
ACCESSION   AA924990
NID         g3072126
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 360)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT
            Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence tag present in the cDNA between the NotI site and the
            oligo-dT track (not shown) served to identify it as a clone from
            the normalized adult Spleen library. cDNA Library Preparation: M.
            Fatima Bonaldo, Ph.D. Clone distribution: clones will be available
            through Research Genetics
            Seq primer: M13 Forward.
FEATURES                Location/Qualifiers
     source             1. .360
                        /organism="Rattus norvegicus"
                        /strain="Sprague-Dawley"
                        /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
```



```
                        polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-A1
                        library is a subtracted library derived from the UI-R-A0
                        library. The UI-R-A0 library  consisted of a mixture of
                        individually tagged normalized libraries constructed from
                        rat placenta, adult lung, brain, liver, kidney, heart,
                        spleen, ovary, and muscle. The tag is a string of 3-5
                        nucleotides present between the Not I site and the
                        oligo-dT track which allows identification of the library
                        of origin of a clone within the mixture. The subtracted
                        library (UI-R-A1) was constructed as follows: PCR
                        amplified cDNA inserts from a pool of approximately 3,840
                        UI-R-A0 clones from which 3' ESTs had been derived was
                        used as a driver in a hybridization with the UI-R-A0
                        library in the form of single-stranded circles. The
                        remaining single-stranded circles (subtracted library) was
                        purified by hydroxyapatite column chromatography,
                        converted to double-stranded circles and electroporated
                        into DH10B bacteria (Life Technologies) to generate the
                        UI-R-A1 library. This procedure has been previously
                        described (Bonaldo, Lennon and Soares, Genome Research 6:
                        791-806, 1996)"
                        /db_xref="taxon:10116"
                        /clone="UI-R-A1-eh-b-12-0-UI"
                        /clone_lib="UI-R-A1"
                        /dev_stage="adult"
                        /lab_host="DH10B (Life Technologies)"
BASE COUNT      122 a     65 c     59 g    114 t
ORIGIN

  Query Match          13.6%;  Score 20;  DB 14;  Length 360;
  Best Local Similarity 100.0%;  Pred. No. 1.75e-08;
  Matches   20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    177 TCGGTTCCAAGTACCAAGGG 196
          ||||||||||||||||||||
Cp    113 TCGGTTCCAAGTACCAAGGG 94


RESULT   15
LOCUS       AA399462      369 bp    mRNA            EST       16-MAY-1997
DEFINITION  zt53c08.s1 Soares ovary tumor NbHOT Homo sapiens cDNA clone 726062
            3', mRNA sequence.
ACCESSION   AA399462
NID         g2053208
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 369)
```



```
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J.,
            Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-Merck EST Project 1997
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -41m13 fwd. ET from Amersham
            High quality sequence stop: 360.
FEATURES                Location/Qualifiers
     source             1. .369
                        /organism="Homo sapiens"
                        /note="Organ: ovary; Vector: pT7T3D (Pharmacia) with a
                        modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                        strand cDNA was primed with a Not I - oligo(dT) primer [5'
                        TGTTACCAATCTGAAGTGGGAGCGGCCGCGGTTTTTTTTTTTTTTTTTT 3'],
                        double-stranded cDNA was size selected, ligated to Eco RI
                        adapters (Pharmacia), digested with Not I and cloned into
                        the Not I and Eco RI sites of a modified pT7T3 vector
                        (Pharmacia). Library constructed by Bento Soares and
                        M.Fatima Bonaldo."
                        /db_xref="taxon:9606"
                        /clone="726062"
                        /clone_lib="Soares ovary tumor NbHOT"
                        /sex="Female"
                        /tissue_type="ovarian tumor"
                        /lab_host="DH10B (ampicillin resistant)"
     mRNA               complement(<1. .>369)
                        /db_xref="GDB:5938087"
BASE COUNT       97 a     55 c     63 g    154 t
ORIGIN

  Query Match          13.6%;  Score 20;  DB 6;  Length 369;
  Best Local Similarity 100.0%;  Pred. No. 1.75e-08;
  Matches   20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      3 TACATTATTATTTTAATATA 22
          ||||||||||||||||||||
Cp    143 TACATTATTATTTTAATATA 124


RESULT   16
```

```
LOCUS       AA721239      380 bp    mRNA            EST       23-JAN-1998
DEFINITION  nz72c05.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1301000,
            mRNA sequence.
ACCESSION   AA721239
NID         g2737374
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 380)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
            Ph.D., Gerald Marti, M.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 834   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 320.
FEATURES             Location/Qualifiers
     source          1. .380
                     /organism="Homo sapiens"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from human tonsillar cells enriched for
                     germinal center B cells by flow sorting (CD20+, IgD-),
                     provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                     (NCI) and Dr. Gerald Marti (CBER).  cDNA synthesis was
                     primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTTT-
                     3']. Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo."
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1301000"
                     /clone_lib="NCI_CGAP_GCB1"
```



```
                     /tissue_type="germinal center B cell"
                     /lab_host="DH10B"
BASE COUNT       98 a      55 c      63 g     164 t
ORIGIN

  Query Match           13.6%;  Score 20;  DB 12;  Length 380;
  Best Local Similarity 100.0%;  Pred. No. 1.75e-08;
  Matches   20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      15 TACATTATTATTTTAATATA 34
           ||||||||||||||||||||
Cp     143 TACATTATTATTTTAATATA 124


RESULT   17
LOCUS       AA415181      382 bp    mRNA            EST       16-OCT-1997
DEFINITION  vc98e08.s1 Knowles Solter mouse 2 cell Mus musculus cDNA clone
            791078 5', mRNA sequence.
ACCESSION   AA415181
NID         g2075271
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 382)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:475486
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 297.
FEATURES             Location/Qualifiers
     source          1. .382
                     /organism="Mus musculus"
                     /strain="B6D2 F1/J"
                     /note="Organ: embryo; Vector: pBluescribe (modified);
```



```
                     Site_1: MluI; Site_2: SalI; Cloned unidirectionally from
                     mRNA prepared from 13,500 2-cell stage  embryos.  Primer:
                     SalI(dT): 5'-CGGTCGACCGTCGACCGTTTTTTTTTTTTTT-3'.  cDNAs
                     were cloned into the MluI/SalI sites of a modified
                     pBluescribe  vector using commercial linkers (NEB).
                     Average insert size:  1.2 kb."
                     /db_xref="taxon:10090"
                     /clone="791078"
                     /clone_lib="Knowles Solter mouse 2 cell"
                     /tissue_type="embryo"
                     /dev_stage="2-cell"
                     /lab_host="DH10B"
BASE COUNT      111 a      62 c      67 g     142 t
ORIGIN

  Query Match           13.6%;  Score 20;  DB 10;  Length 382;
  Best Local Similarity 100.0%;  Pred. No. 1.75e-08;
  Matches   20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     137 TGTCCAGCGCACTGTGAGGC 156
           ||||||||||||||||||||
Qy      49 TGTCCAGCGCACTGTGAGGC 68


RESULT   18
LOCUS       AA119227      432 bp    mRNA            EST       17-FEB-1997
DEFINITION  mp77b12.r1 Soares 2NbMT Mus musculus cDNA clone 575231 5', mRNA
            sequence.
ACCESSION   AA119227
NID         g1676717
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Mus.
REFERENCE   1  (bases 1 to 432)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
```



```
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:349879
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 401.
FEATURES             Location/Qualifiers
     source          1. .432
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. RNA
                     provided by Dr. Bertrand Jordan. Library went through two
                     rounds of normalization, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
                     /db_xref="taxon:10090"
                     /clone="575231"
                     /clone_lib="Soares 2NbMT"
                     /sex="male"
                     /tissue_type="Thymus"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
     mRNA            <1. .>432
BASE COUNT      112 a      79 c      81 g     159 t       1 others
ORIGIN

  Query Match           13.6%;  Score 20;  DB 23;  Length 432;
  Best Local Similarity 100.0%;  Pred. No. 1.75e-08;
  Matches   20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     277 TGTCCAGCGCACTGTGAGGC 296
           ||||||||||||||||||||
Qy      49 TGTCCAGCGCACTGTGAGGC 68


RESULT   19
LOCUS       W52272        447 bp    mRNA            EST       11-OCT-1996
DEFINITION  zc46d11.r1 Soares senescent fibroblasts NbHSF Homo sapiens cDNA
            clone 325365 5', mRNA sequence.
ACCESSION   W52272
NID         g1349402
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE   1  (bases 1 to 447)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1548   Std Error: 0.00
            Seq primer: mob.REGA+ET
            High quality sequence stop: 425.
FEATURES            Location/Qualifiers
     source         1. .447
                    /organism="Homo sapiens"
                    /note="Vector: pT7T3D (Pharmacia) with a modified
                    polylinker V_TYPE: phagemid; Site_1: Not I; Site_2: Eco
                    RI; TGTTACCAATCTGAAGTGGGAGCGGCCGGCATTTTTTTTTTTTTTTTTTT
                    3'], double-stranded cDNA was size selected, ligated to
                    Eco RI adapters (Pharmacia), digested with Not I and
                    cloned into the Not I and Eco RI sites of a modified pT7T3
                    vector (Pharmacia). Library went through one round of
                    normalization to a Cot = 5. Library constructed by Bento
                    Soares and M.Fatima Bonaldo."
                    /db_xref="taxon:9606"
                    /clone="325365"
                    /clone_lib="Soares senescent fibroblasts NbHSF"
                    /tissue_type="senescent fibroblast"
                    /lab_host="DH10B (ampicillin resistant)"
     mRNA           <1. .>447
BASE COUNT      156 a     78 c     79 g    132 t      2 others
ORIGIN

  Query Match           13.6%;  Score 20;  DB 21;  Length 447;
  Best Local Similarity 100.0%;  Pred. No. 1.75e-08;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    242 TATATTAAAATAATAATGTA 261
          ||||||||||||||||||||
Qy    124 TATATTAAAATAATAATGTA 143
```



```
RESULT   20
LOCUS       AA048254      457 bp    mRNA            EST       09-SEP-1996
DEFINITION  mj27e04.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
            clone 477342 5', mRNA sequence.
ACCESSION   AA048254
NID         g1527942
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Mus.
REFERENCE   1  (bases 1 to 457)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:288086
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 419.
FEATURES            Location/Qualifiers
     source         1. .457
                    /organism="Mus musculus"
                    /strain="C57BL/6J"
                    /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                    polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                    was primed with a Not I - oligo(dT) primer [5'
                    TGTTACCAATCTGAAGTGGGAGCGGCCGCGGAAATTTTTTTTTTTTTTTTTTTTT
                    T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
                    14.5dpc embryos [total RNA provided by  Minoru Ko, Wayne
                    State Univ., from 2 ]; double-stranded cDNA was ligated to
                    Eco RI adaptors (Pharmacia), digested with Not I and
                    cloned into the Not I and Eco RI sites of the modified
                    pT7T3 vector. Library went through one round of
                    normalization, and was constructed by Bento Soares and
                    M.Fatima Bonaldo."
                    /db_xref="taxon:10090"
                    /clone="477342"
```



```
                    /clone_lib="Soares mouse embryo NbME13.5 14.5"
                    /sex="unknown"
                    /tissue_type="embryo"
                    /dev_stage="13.5-14.5dpc total fetus"
                    /lab_host="DH10B"
     mRNA           <1. .>457
BASE COUNT      130 a     79 c     81 g    167 t
ORIGIN

  Query Match           13.6%;  Score 20;  DB 21;  Length 457;
  Best Local Similarity 100.0%;  Pred. No. 1.75e-08;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    141 TGTCCAGCGCACTGTGAGGC 160
          ||||||||||||||||||||
Qy     49 TGTCCAGCGCACTGTGAGGC 68


RESULT   21
LOCUS       AI050617      476 bp    mRNA            EST       09-JUL-1998
DEFINITION  ub38b11.r1 Soares 2NbMT Mus musculus cDNA clone 1379997 5', mRNA
            sequence.
ACCESSION   AI050617
NID         g3299734
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 476)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:902465
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 475.
FEATURES            Location/Qualifiers
```



```
     source         1. .476
                    /organism="Mus musculus"
                    /strain="C57BL/6J"
                    /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                    polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                    was primed with a Not I - oligo(dT) primer [5'
                    TGTTACCAATCTGAAGTGGGAGCGGCCGCGTTTTTTTTTTTTTTTTTTTTTTT
                    3']; double-stranded cDNA was ligated to Eco RI adaptors
                    (Pharmacia), digested with Not I and cloned into the Not I
                    and Eco RI sites of the modified pT7T3 vector. RNA
                    provided by Dr. Bertrand Jordan. Library went through two
                    rounds of normalization, and was constructed by Bento
                    Soares and M.Fatima Bonaldo."
                    /db_xref="taxon:10090"
                    /clone="1379997"
                    /clone_lib="Soares 2NbMT"
                    /sex="male"
                    /tissue_type="Thymus"
                    /dev_stage="4 weeks"
                    /lab_host="DH10B"
BASE COUNT      113 a     96 c     98 g    168 t      1 others
ORIGIN

  Query Match           13.6%;  Score 20;  DB 15;  Length 476;
  Best Local Similarity 100.0%;  Pred. No. 1.75e-08;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    312 TGTCCAGCGCACTGTGAGGC 331
          ||||||||||||||||||||
Qy     49 TGTCCAGCGCACTGTGAGGC 68


RESULT   22
LOCUS       AA048095      491 bp    mRNA            EST       09-SEP-1996
DEFINITION  mj23d09.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
            clone 476945 5', mRNA sequence.
ACCESSION   AA048095
NID         g1527766
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Mus.
REFERENCE   1  (bases 1 to 491)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
```

```
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:287689
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 489.
FEATURES              Location/Qualifiers
     source           1. .491
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                      was primed with a Not I - oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCGGAAATTTTTTTTTTTTTTTTTTTTTTT
                      T 3']; on equal amounts of mRNA from 2 13.5dpc and 2
                      14.5dpc embryos [total RNA provided by  Minoru Ko, Wayne
                      State Univ., from 2 ]; double-stranded cDNA was ligated to
                      Eco RI adaptors (Pharmacia), digested with Not I and
                      cloned into the Not I and Eco RI sites of the modified
                      pT7T3 vector. Library went through one round of
                      normalization, and was constructed by Bento Soares and
                      M.Fatima Bonaldo."
                      /db_xref="taxon:10090"
                      /clone="476945"
                      /clone_lib="Soares mouse embryo NbME13.5 14.5"
                      /sex="unknown"
                      /tissue_type="embryo"
                      /dev_stage="13.5-14.5dpc total fetus"
                      /lab_host="DH10B"
     mRNA             <1. .>491
BASE COUNT      144 a     90 c     83 g    174 t
ORIGIN

  Query Match           13.6%;  Score 20;  DB 21;  Length 491;
  Best Local Similarity 100.0%;  Pred. No. 1.75e-08;
  Matches   20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    141 TGTCCAGCGCACTGTGAGGC 160
          ||||||||||||||||||||
Qy     49 TGTCCAGCGCACTGTGAGGC 68


RESULT   23
```



```
LOCUS       AA185438      493 bp    mRNA            EST       19-FEB-1997
DEFINITION  mt61c06.r1 Soares 2NbMT Mus musculus cDNA clone 634378 5', mRNA
            sequence.
ACCESSION   AA185438
NID         g1759084
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Mus.
REFERENCE   1  (bases 1 to 493)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:386370
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 492.
FEATURES              Location/Qualifiers
     source           1. .493
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                      was primed with a Not I - oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCGTTTTTTTTTTTTTTTTTTTTTTTT
                      3']; double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Not I and cloned into the Not I
                      and Eco RI sites of the modified pT7T3 vector. RNA
                      provided by Dr. Bertrand Jordan. Library went through two
                      rounds of normalisation, and was constructed by Bento
                      Soares and M.Fatima Bonaldo."
                      /db_xref="taxon:10090"
                      /clone="634378"
                      /clone_lib="Soares 2NbMT"
                      /sex="male"
                      /tissue_type="Thymus"
```



```
                      /dev_stage="4 weeks"
                      /lab_host="DH10B"
     mRNA             <1. .>493
BASE COUNT      135 a    104 c    107 g    147 t
ORIGIN

  Query Match           13.6%;  Score 20;  DB 23;  Length 493;
  Best Local Similarity 100.0%;  Pred. No. 1.75e-08;
  Matches   20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     19 CCCTTGGTACTTGGAACCGA 38
          ||||||||||||||||||||
Qy     94 CCCTTGGTACTTGGAACCGA 113


RESULT   24
LOCUS       AI083687      335 bp    mRNA            EST       17-AUG-1998
DEFINITION  oy71g10.x1 NCI_CGAP_CLL1 Homo sapiens cDNA clone IMAGE:1671330 3',
            mRNA sequence.
ACCESSION   AI083687
NID         g3422110
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 335)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever,
            M.D., Louis M. Staudt, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 322.
FEATURES              Location/Qualifiers
     source           1. .335
                      /organism="Homo sapiens"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
```



```
                      was primed with a Not I - oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCATTGCTTTTTTTTTTTTTTTTTTTTTTT
                      T 3']; double-stranded cDNA was ligated to Eco RI
                      adaptors (Pharmacia), digested with Not I and cloned into
                      the Not I and Eco RI sites of the modified pT7T3 vector.
                      Library is normalized, and was constructed by Bento
                      Soares and M.Fatima Bonaldo."
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1671330"
                      /clone_lib="NCI_CGAP_CLL1"
                      /tissue_type="B-cell, chronic lymphotic leukemia"
                      /lab_host="DH10B"
BASE COUNT      121 a     41 c     53 g    120 t
ORIGIN

  Query Match           12.9%;  Score 19;  DB 17;  Length 335;
  Best Local Similarity 100.0%;  Pred. No. 7.84e-07;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    256 CATAAATAAGACAAAGAAA 274
          |||||||||||||||||||
Cp     30 CATAAATAAGACAAAGAAA 12


RESULT   25
LOCUS       B95458        716 bp    DNA             GSS       25-JUN-1998
DEFINITION  CIT-HSP-2166B16.TF CIT-HSP Homo sapiens genomic clone 2166B16,
            genomic survey sequence.
ACCESSION   B95458
NID         g2977795
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 716)
  AUTHORS   Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K.,
            Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
            Simon,M. and Venter,J.C.
  TITLE     Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: CIT-HSP-2166B16.TRB
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
```

```
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: M13-21;
            Class: BAC ends.
FEATURES                Location/Qualifiers
     source             1. .716
                        /organism="Homo sapiens"
                        /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                        HindIII"
                        /db_xref="taxon:9606"
                        /clone="2166B16"
                        /clone_lib="CIT-HSP"
                        /sex="Male"
                        /cell_type="Sperm"
BASE COUNT      318 a    101 c    107 g    189 t      1 others
ORIGIN

  Query Match           12.9%;  Score 19;  DB 27;  Length 716;
  Best Local Similarity 100.0%;  Pred. No. 7.84e-07;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    602 ATATTAAAATAATAATGTA 620
          ||||||||||||||||||
Qy    125 ATATTAAAATAATAATGTA 143


RESULT   26
LOCUS       AA996256      337 bp    mRNA            EST       27-JUL-1998
DEFINITION  os18e06.s1 NCI_CGAP_Kid5 Homo sapiens cDNA clone IMAGE:1605730 3'
            similar to gb:Z26876 60S RIBOSOMAL PROTEIN L38 (HUMAN);, mRNA
            sequence.
ACCESSION   AA996256
NID         g2182745
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 337)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
```



```
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Trace considered overall poor quality
            Insert Length: 395   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 1.
FEATURES                Location/Qualifiers
     source             1. .337
                        /organism="Homo sapiens"
                        /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                        a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                        strand cDNA was primed with a Not I - oligo(dT) primer [5'
                        AACTGGAAGAATTCGCGGCCGCAATATTTTTTTTTTTTTTTTT 3'],
                        double-stranded cDNA was ligated to Eco RI adaptors
                        (Pharmacia), digested with Not I and cloned into the Not I
                        and Eco RI sites of the modified pT7T3 vector. Library
                        went through one round of normalization. Library
                        constructed by Bento Soares and M. Fatima Bonaldo."
                        /db_xref="taxon:9606"
                        /clone="IMAGE:1605730"
                        /clone_lib="NCI_CGAP_Kid5"
                        /tissue_type="2 pooled tumors (clear cell type)"
                        /lab_host="DH10B"
BASE COUNT       69 a     81 c     60 g    127 t
ORIGIN

  Query Match           12.2%;  Score 18;  DB 15;  Length 337;
  Best Local Similarity 100.0%;  Pred. No. 3.09e-05;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      6 TTATTATTTTAATATAAT 23
          ||||||||||||||||||
Cp    139 TTATTATTTTAATATAAT 122


RESULT   27
LOCUS       AQ027251      363 bp    DNA             GSS       29-JUN-1998
DEFINITION  CIT-HSP-2314L21.TF CIT-HSP Homo sapiens genomic clone 2314L21,
            genomic survey sequence.
ACCESSION   AQ027251
NID         g3267473
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 363)
  AUTHORS   Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K.,
```



```
            Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
            Simon,M. and Venter,J.C.
  TITLE     Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13-21
            Class: BAC ends.
FEATURES                Location/Qualifiers
     source             1. .363
                        /organism="Homo sapiens"
                        /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                        HindIII"
                        /db_xref="taxon:9606"
                        /clone="2314L21"
                        /clone_lib="CIT-HSP"
                        /sex="Male"
                        /cell_type="Sperm"
BASE COUNT      111 a     79 c     91 g     82 t
ORIGIN

  Query Match           12.2%;  Score 18;  DB 27;  Length 363;
  Best Local Similarity 100.0%;  Pred. No. 3.09e-05;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     12 ATTATATTAAAATAATAA 29
          ||||||||||||||||||
Qy    122 ATTATATTAAAATAATAA 139


RESULT   28
LOCUS       CELK028B9R    400 bp    mRNA            EST       06-JAN-1997
DEFINITION  C.elegans cDNA clone yk28b9 : 3' end, single read, mRNA sequence.
ACCESSION   D33248
NID         g524175
KEYWORDS    EST; EST(expressed sequence tag).
SOURCE      Caenorhabditis elegans (strain:CB1489 him-8(e1489)) varied
            Hermaphrodite, male whole animal cDNA to mRNA, clone_lib:Yuji
            Kohara unpublished cDNA.
  ORGANISM  Caenorhabditis elegans
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Nematoda;
```



```
            Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea;
            Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 400)
  AUTHORS   Kohara,Y., Mitsuki,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and
            Tabara,H.
  TITLE     Toward an expression map of the C.elegans genome
  JOURNAL   Unpublished (1994)
REFERENCE   2  (bases 1 to 400)
  AUTHORS   Kohara,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUL-1994) to the DDBJ/EMBL/GenBank databases. Yuji
            Kohara, National Institute of Genetics, Gene Library Lab; Yata
            1111, Mishima, Shizuoka 411, Japan (E-mail:ykohara@lab.nig.ac.jp,
            Tel:0559-81-6854, Fax:0559-81-6855)
FEATURES                Location/Qualifiers
     source             1. .400
                        /organism="Caenorhabditis elegans"
                        /strain="CB1489 him-8(e1489)"
                        /db_xref="taxon:6239"
                        /clone_lib="Yuji Kohara unpublished cDNA"
                        /dev_stage="varied"
                        /sex="Hermaphrodite, male"
                        /tissue_type="whole animal"
BASE COUNT      134 a     48 c     78 g    134 t      6 others
ORIGIN

  Query Match           12.2%;  Score 18;  DB 6;  Length 400;
  Best Local Similarity 100.0%;  Pred. No. 3.09e-05;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     45 TTATATTAAAATAATAAT 62
          ||||||||||||||||||
Qy    123 TTATATTAAAATAATAAT 140


RESULT   29
LOCUS       AA595195      470 bp    mRNA            EST       08-OCT-1997
DEFINITION  no32f09.s1 NCI_CGAP_Pr23 Homo sapiens cDNA clone IMAGE:1102409,
            mRNA sequence.
ACCESSION   AA595195
NID         g2410545
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 470)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
```

```
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: Stratagene, Inc.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 1294   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 470.
FEATURES              Location/Qualifiers
     source           1. .470
                      /organism="Homo sapiens"
                      /note="Organ: prostate; Vector: Bluescript SK-; Site_1:
                      EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                      Oligo dT. Pooled prostate tumors. 5' adaptor sequence: 5'
                      GAATTCGGCACGAG 3' 3' adaptor sequence: 5'
                      CTCGAGTTTTTTTTTTTTTTTTT 3' Average insert size:  1.2 kb."
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1102409"
                      /clone_lib="NCI_CGAP_Pr23"
                      /sex="male, pooled"
                      /tissue_type="prostate tumor"
                      /lab_host="SOLR (kanamycin resistant)"
BASE COUNT      183 a     71 c     64 g    152 t
ORIGIN

  Query Match           12.2%;  Score 18;  DB 10;  Length 470;
  Best Local Similarity 100.0%;  Pred. No. 3.09e-05;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    256 AAAATAATAATGTACAAA 273
          ||||||||||||||||||
Qy    130 AAAATAATAATGTACAAA 147


RESULT   30
LOCUS       B00235        622 bp    DNA             GSS       26-JUN-1996
DEFINITION  cSRL-104g4-u cSRL flow sorted Chromosome 11 specific cosmid Homo
            sapiens genomic clone cSRL-104g4, genomic survey sequence.
ACCESSION   B00235
NID         g1409513
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
```



```
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 622)
  AUTHORS   Evans,G.A., Burbee,D., Davies,C., Hahner,L., Oliver,T., Gilbert,M.,
            Jones,D., Ward,T., Gillilan,E., Schagemann,J., Probst,S. .,
            Harris,J., DeFord,J., McFarland,J., Bursinski,K., Khan,M.,
            Kupfer,K. and Garner,H.R.
  TITLE     Genomic Sequence Sampled Map of Chromosome 11
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Evans GA, Shane Probst
            McDermott Center for Human Growth and Development
            University of Texas Southwestern Medical Center At Dallas
            5323 Harry Hines Blvd, Dallas TX 75235-8591
            Tel: 214-648-1600
            Fax: 214-648-1666
            Email: gevans@utsw.swmed.edu, shane@mcdermott.swmed.edu
            PCR PRimers
            FORWARD: GTCTCACCTTTTAAATTCCC
            BACKWARD: ATTGAGTTCTGACAATGTGG
            Seq primer: T7
            Class: cosmid ends
            High quality sequence stop: 622.
FEATURES              Location/Qualifiers
     source           1. .622
                      /organism="Homo sapiens"
                      /note="Vector: sCos-1; Human Chromosome 11 specific cosmid
                      library prepared from flow sorted human  Chromosome 11
                      derived from Chinese Hampster Ovary (CHO) monochromosomal
                      somatic cell hybrid, J1"
                      /db_xref="taxon:9606"
                      /clone="cSRL-104g4"
                      /clone_lib="cSRL flow sorted Chromosome 11 specific
                      cosmid"
                      /sex="female"
                      /cell_type="chimeric hamster somatic cell hybrid"
BASE COUNT      178 a     97 c     59 g    256 t     32 others
ORIGIN

  Query Match           12.2%;  Score 18;  DB 26;  Length 622;
  Best Local Similarity 100.0%;  Pred. No. 3.09e-05;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    393 TACATTATTATTTTAATA 410
          ||||||||||||||||||
Cp    143 TACATTATTATTTTAATA 126


RESULT   31
LOCUS       AA999800       68 bp    mRNA            EST       05-JUN-1998
DEFINITION  os41c07.s1 NCI_CGAP_Br2 Homo sapiens cDNA clone IMAGE:1607916 3',
```



```
            mRNA sequence.
ACCESSION   AA999800
NID         g3190355
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 68)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 68.
FEATURES              Location/Qualifiers
     source           1. .68
                      /organism="Homo sapiens"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; 1st strand cDNA was prepared from pooled bulk
                      breast tumor tissue, and was then primed with a Not I -
                      oligo(dT) primer. Double-stranded cDNA was ligated to Eco
                      RI adaptors (Pharmacia), digested with Not I and cloned
                      into the Not I and Eco RI sites of the modified pT7T3
                      vector.  This library is the normalized version of
                      NCI_CGAP_Br1.1. Library was constructed by Bento Soares
                      and M. Fatima Bonaldo."
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1607916"
                      /clone_lib="NCI_CGAP_Br2"
                      /sex="female, pooled"
                      /tissue_type="breast"
                      /lab_host="DH10B"
BASE COUNT       20 a     11 c      4 g     33 t
ORIGIN

  Query Match           11.6%;  Score 17;  DB 15;  Length 68;
  Best Local Similarity 100.0%;  Pred. No. 1.06e-03;
```



```
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     20 TTATTTTAATATAATCT 36
          |||||||||||||||||
Cp    136 TTATTTTAATATAATCT 120


RESULT   32
LOCUS       AA224351      150 bp    mRNA            EST       11-MAR-1998
DEFINITION  zr12f12.s1 Stratagene hNT neuron (#937233) Homo sapiens cDNA clone
            648623 3', mRNA sequence.
ACCESSION   AA224351
NID         g1844893
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 150)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
            Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
            Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1183   Std Error: 0.00
            Seq primer: -41m13 fwd. ET from Amersham
            High quality sequence stop: 140.
FEATURES              Location/Qualifiers
     source           1. .150
                      /organism="Homo sapiens"
                      /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                      XhoI; Cloned unidirectionally.  Primer: Oligo dT.
                      Differentiated, post mitotic hNT neurons. Average insert
                      size: 1.5 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5'
                      GAATTCGGCACGAG 3' -3' adaptor sequence: 5'
                      CTCGAGTTTTTTTTTTTTTTTTT 3'"
                      /db_xref="GDB:5424652"
                      /db_xref="taxon:9606"
                      /clone="648623"
                      /clone_lib="Stratagene hNT neuron (#937233)"
                      /dev_stage="hNT neurons"
```

```
                    /lab_host="SOLR (kanamycin resistant)"
BASE COUNT       43 a      25 c      18 g      64 t
ORIGIN

  Query Match           11.6%;  Score 17;  DB 13;  Length 150;
  Best Local Similarity 100.0%;  Pred. No. 1.06e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      23 ATTTTAATATAATCTGT 39
           |||||||||||||||||
Cp     134 ATTTTAATATAATCTGT 118


RESULT   33
LOCUS       M79056        191 bp    mRNA            EST       26-MAY-1992
DEFINITION  EST01204 Homo sapiens cDNA clone HHCPJ21.
ACCESSION   M79056
NID         g273369
KEYWORDS    EST.
SOURCE      human clone=HHCPJ21 library=Subtracted Hippocampus, Stratagene
            (cat. #936205) vector=lambdaZAP-II primer=M13 Forward The
            hippocampus library (#4 above) was subtracted with a fibroblast
            cell line cDNA library (Stratagene cat. #936209; WI38 lung
            fibroblast cell line; oligo-dT + random primed cDNA synthesis;
            lambdaZAP-II vector, 1.0kb average insert size.) by the method of
            Sive & St. John (Nucl. Acids Res. 16:10937, 1988).
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 191)
  AUTHORS   Adams,M.D., Dubnick,M., Kerlavage,A.R., Moreno,R., Kelley,J.M.,
            Utterback,T.R., Nagle,J.W., Fields,C. and Venter,J.C.
  TITLE     Sequence identification of 2,375 human brain genes
  JOURNAL   Nature 355 (6361), 632-634 (1992)
  MEDLINE   92168112
COMMENT
            Contact: Kerlavage AR
            The Institute for Genomic Research
            932 Clopper Road,Gaithersburg,MD 20878
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
FEATURES             Location/Qualifiers
     source          1. .191
                     /organism="Homo sapiens"
                     /clone="HHCPJ21"
     mRNA            <1. .>191
                     /gene="D0S2025E"
     gene            <1. .>191
```



```
                    /gene="D0S2025E"
BASE COUNT       35 a      40 c      87 g      27 t       2 others
ORIGIN

  Query Match           11.6%;  Score 17;  DB 16;  Length 191;
  Best Local Similarity  94.4%;  Pred. No. 1.06e-03;
  Matches   17;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db     127 ACTNAGGCCAGCCCTGGC 144
           ||| ||||||||||||||||
Qy      73 ACTCAGGCCAGCCCTGGC 90


RESULT   34
LOCUS       N54751        197 bp    mRNA            EST       16-FEB-1996
DEFINITION  yz16g11.r1 Homo sapiens cDNA clone 283268 5'.
ACCESSION   N54751
NID         g1196071
KEYWORDS    EST.
SOURCE      human clone=283268 primer=T7 library=Soares multiple sclerosis
            2NbHMSP vector=pT7T3D (Pharmacia) with a modified polylinker
            V_TYPE: phagemid host=DH10B (ampicillin resistant) Rsite1=Not I
            Rsite2=Eco RI 46 year old male. 1st strand cDNA was primed with a
            Not I - oligo(dT) primer
            [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCATTTTTTTTTTTTTTTTTT-3'],
            double-stranded cDNA was size selected, ligated to Eco RI adapters
            (Pharmacia), digested with Not I and cloned into the Not I and Eco
            RI sites of a modified pT7T vector (Pharmacia). Library went
            through one round of normalization to a Cot = 5. Library
            constructed by Bento Soares and M.Fatima Bonaldo. RNA from 4
            multiple sclerosis lesions from one patient was kindly provided by
            Dr. Kevin G. Becker (NINDS/NIH).
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 197)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT

            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
```



```
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 161
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES             Location/Qualifiers
     source          1. .197
                     /organism="Homo sapiens"
                     /clone="283268"
     mRNA            <1. .>197
BASE COUNT       67 a      32 c      30 g      66 t       2 others
ORIGIN

  Query Match           11.6%;  Score 17;  DB 19;  Length 197;
  Best Local Similarity 100.0%;  Pred. No. 1.06e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     148 AGATTATATTAAAATAA 164
           |||||||||||||||||
Qy     120 AGATTATATTAAAATAA 136


RESULT   35
LOCUS       AU006635      212 bp    mRNA            EST       25-JUL-1998
DEFINITION  Schizosaccharomyces pombe cDNA, clone spc00212, mRNA sequence.
ACCESSION   AU006635
NID         g3343094
KEYWORDS    EST; EST(expressed sequence tag).
SOURCE      Schizosaccharomyces pombe (strain:972) h minus cDNA to mRNA,
            clone_lib:M13mp19, late log phase cDNA library clone:spc00212.
  ORGANISM  Schizosaccharomyces pombe
            Eukaryota; Fungi; Ascomycota; Archiascomycetes;
            Schizosaccharomycetales; Schizosaccharomycetaceae;
            Schizosaccharomyces.
REFERENCE   1  (bases 1 to 212)
  AUTHORS   Morimyo,M. and Mita,K.
  TITLE     Identification of expressed sequence tags of Schizosaccharomyces
            pombe
  JOURNAL   Published Only in DataBase (1998) In press
REFERENCE   2  (bases 1 to 212)
  AUTHORS   Morimyo,M. and Mita,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUN-1998) to the DDBJ/EMBL/GenBank databases.
            Mitsuoki Morimyo, National Institute of Radiological Sciences,
            Genome Research Group; 9-1, Anagawa-4-chome, Inage-ku, Chiba, Chiba
            263-8555, Japan (E-mail:morimyo@nirs.go.jp, Tel:+81-43-206-3144,
            Fax:+81-43-252-8214)
COMMENT     The cDNA library of Schizosaccharomyces pombe was prepared by
            cloning cDNA into the SmaI site of M13mp19 DNA and the direction of
            DNA sequences was not always from 5' to 3'.
```



```
            The cDNA data of Schizosaccharomyces pombe are available for
            searching on the World Wide Web.  (URL, http://www.nirs.go.jp).
FEATURES             Location/Qualifiers
     source          1. .212
                     /organism="Schizosaccharomyces pombe"
                     /strain="972"
                     /db_xref="taxon:4896"
                     /clone="spc00212"
                     /clone_lib="M13mp19, late log phase cDNA library"
                     /sex="h minus"
BASE COUNT       71 a      43 c      30 g      68 t
ORIGIN

  Query Match           11.6%;  Score 17;  DB 15;  Length 212;
  Best Local Similarity 100.0%;  Pred. No. 1.06e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      57 AGGCCCCCACTCAGGCC 73
           |||||||||||||||||
Qy      65 AGGCCCCCACTCAGGCC 81


RESULT   36
LOCUS       HSC27F062     245 bp    RNA             EST       21-SEP-1995
DEFINITION  H. sapiens partial cDNA sequence; clone c-27f06, mRNA sequence.
ACCESSION   Z40550
NID         g566247
KEYWORDS    EST; partial cDNA sequence; transcribed sequence fragment.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 245)
  AUTHORS   Genexpress.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-OCT-1994) Genethon, B.P. 60, 91002 Evry Cedex France
            and Genetique Moleculaire et Biologie du developpement, CNRS UPR420
            B.P. 8, 94801 Villejuif Cedex France.E-mail: genexpress@genethon.fr
REFERENCE   2  (bases 1 to 245)
  AUTHORS   Genexpress.
  TITLE     The Genexpress cDNA program
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 245)
  AUTHORS   Auffray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C.,
            Devignes,M.D., Duprat,S., Houlgatte,R., Jumeau,M.N., Lamy,B.,
            Lorenso,F., Mitchell,H., Mariage-Samson,R., Pietu,G., Pouliot,Y.,
            Sebastiani-Kabattchis,C. and Tessier,A.
  TITLE     IMAGE: molecular integration of the analysis of the human genome
            and its expression
  JOURNAL   C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
  MEDLINE   95277534
```

```
COMMENT     Clone library from B.Soares, Psychiatry Dept. Columbia University
            USA;

            Cloning_method: total mRNA was oligo-(dT) primed and directionally
            cloned 5' -> 3' into the HindIII -> NotI sites of the lafmid BA
            vector;
            Sequencing_method: single read, full automatic;
            Primer: (-21)M13_universal;
            cDNA sequence complementary to mRNA (3'end)
            Stretch_removed: 20 T removed at sequence 5'end
            Normalisation_method: Bento Soares, P.N.A.S in press;
            Genexpress_library_idt: C;
            Genexpress_sequence_idt: alc-27f06;

            No significant homology found with :
            genbank release 81 swissprot release 28.
FEATURES              Location/Qualifiers
     source           1. .245
                      /organism="Homo sapiens"
                      /isolate="muscular atrophy patient"
                      /db_xref="taxon:9606"
                      /dev_stage="3 months old"
                      /tissue_type="total brain"
                      /clone_lib="normalized infant brain cDNA"
                      /sex="Female"
BASE COUNT       46 a      53 c     101 g      38 t       7 others
ORIGIN

  Query Match           11.6%;  Score 17;  DB 19;  Length 245;
  Best Local Similarity  94.4%;  Pred. No. 1.06e-03;
  Matches    17;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db    138 ACTNAGGCCAGCCCTGGC 155
          ||| ||||||||||||||
Qy     73 ACTCAGGCCAGCCCTGGC 90


RESULT   37
LOCUS       AA100296       259 bp    mRNA            EST       19-MAY-1997
DEFINITION  zl81g06.s1 Stratagene colon (#937204) Homo sapiens cDNA clone
            511066 3' similar to TR:G755872 G755872 POLY(A) POLYMERASE. ;, mRNA
            sequence.
ACCESSION   AA100296
NID         g1646587
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 259)
```



```
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
            Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Possible reversed clone: similarity on wrong strand
            Insert Length: 1407   Std Error: 0.00
            Seq primer: -40M13 fwd. from Amersham
            High quality sequence stop: 56.
FEATURES              Location/Qualifiers
     source           1. .259
                      /organism="Homo sapiens"
                      /note="Organ: colon; Vector: pBluescript SK-; Site_1:
                      EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                      Oligo dT. T-84 colonic epithelial cell line.  Average
                      insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor
                      sequence: 5' GAATTCGGCACGAG 3' -3' adaptor sequence: 5'
                      CTCGAGTTTTTTTTTTTTTTTT 3'"
                      /db_xref="taxon:9606"
                      /clone="511066"
                      /clone_lib="Stratagene colon (#937204)"
                      /lab_host="SOLR cells (kanamycin resistant)"
     mRNA             complement(<1. .>259)
                      /db_xref="GDB:3843561"
BASE COUNT       76 a      41 c      42 g      93 t       7 others
ORIGIN

  Query Match           11.6%;  Score 17;  DB 6;  Length 259;
  Best Local Similarity 100.0%;  Pred. No. 1.06e-03;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     50 TTATTTTAATATAATCT 66
          |||||||||||||||||
Cp    136 TTATTTTAATATAATCT 120


RESULT   38
LOCUS       N66308         279 bp    mRNA            EST       08-MAR-1996
DEFINITION  yz37h08.s1 Homo sapiens cDNA clone 285279 3' similar to
            SW:PAP_BOVIN P25500 POLY(A) POLYMERASE ;.
```



```
ACCESSION   N66308
NID         g1218433
KEYWORDS    EST.
SOURCE      human clone=285279 primer=m13 -40 forward library=Morton Fetal
            Cochlea vector=pBluescript SK- host=SOLR cells (kanamycin
            resistant) Rsite1=EcoRI Rsite2=XhoI The cDNA was oligo (dT) primed
            with an XhoI restriction enzyme recognition site and an 18 base
            poly dT sequence.  For the 5' end, the synthesized cDNA termini
            were treated with T4 DNA polymerase and EcoRI adaptors were ligated
            to the blunt ends. adaptor linker: GAATTCGGCACGAG.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 279)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
            Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES              Location/Qualifiers
     source           1. .279
                      /organism="Homo sapiens"
                      /clone="285279"
     mRNA             <1. .>279
BASE COUNT       89 a      42 c      49 g      99 t
ORIGIN

  Query Match           11.6%;  Score 17;  DB 19;  Length 279;
  Best Local Similarity 100.0%;  Pred. No. 1.06e-03;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     36 TTATTTTAATATAATCT 52
          |||||||||||||||||
Cp    136 TTATTTTAATATAATCT 120


RESULT   39
```



```
LOCUS       H25898         306 bp    mRNA            EST       10-JUL-1995
DEFINITION  y155b10.r1 Homo sapiens cDNA clone 162139 5'.
ACCESSION   H25898
NID         g895021
KEYWORDS    EST.
SOURCE      human clone=162139 library=Soares breast 3NbHBst vector=pT7T3D
            (Pharmacia) with a modified polylinker host=DH10B (ampicillin
            resistant) primer=M13RP1 Rsite1=Not I Rsite2=Eco RI Adult human.
            1st strand cDNA was primed with a Not I - oligo(dT) primer [5'
            TGTTACCAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTT 3'],
            double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia),
            digested with Not I and cloned into the Not I and Eco RI sites of a
            modified pT7T3 vector (Pharmacia). Library went through one round
            of normalization to a Cot = 20. Library constructed by Bento Soares
            and M.Fatima Bonaldo.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 306)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 177
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES              Location/Qualifiers
     source           1. .306
                      /organism="Homo sapiens"
                      /clone="162139"
BASE COUNT       95 a      35 c      48 g     122 t       6 others
ORIGIN

  Query Match           11.6%;  Score 17;  DB 16;  Length 306;
  Best Local Similarity 100.0%;  Pred. No. 1.06e-03;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```

```
Db      20 AGATTATATTAAAATAA 36
           |||||||||||||||||
Qy     120 AGATTATATTAAAATAA 136


RESULT   40
LOCUS       T33015        325 bp    mRNA            EST       06-SEP-1995
DEFINITION  EST56321 Homo sapiens cDNA 3' end similar to None.
ACCESSION   T33015
NID         g615113
KEYWORDS    EST.
SOURCE      human primer=M13-21 library=Human Brain.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 325)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
            Bult,C.J., Lee,N., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
            White,O., Sutton,G., Blake,J.A., Brandon,R.C., Chiu,M.-W.,
            Clayton,R.A., Cline,R.T., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
            FitzGerald,L.M., FitzHugh,W.M., Fritchman,J.L., Geoghagen,N.S.M.,
            Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle Jr,P.S.,
            Kelley,J.M., Klimek,K.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M.,
            Merrick,J.M., Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T.,
            Pellegrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L.,
            Saudek,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Utterback,T.R.,
            Weidman,J.F., Li,Y., Bednarik,D.P., Cao,L., Cepeda,M.A.,
            Coleman,T.A., Collins,E.-J., Dimke,D., Feng,P., Perrie,A.,
            Fischer,C., Hastings,G.A., He,W.-W., Hu,J.-S., Greene,J.M.,
            Gruber,J., Hudson,P., Kim,A., Kozak,D.L., Kunsch,C., Ji,H., Li,H.,
            Meissner,P.S., Olsen,H., Raymond,L., Wei,Y.-F., Wing,J., Xu,C.,
            Yu,G.-L., Ruben,S.M., Dillon,P.J., Fannon,M.R., Rosen,C.A.,
            Haseltine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.
  TITLE     Initial Assessment of Human Gene Diversity and Expression Patterns
            Based Upon 52 Million Basepairs of cDNA Sequence
  JOURNAL   Unpublished (1995)
COMMENT     Other_ESTs: EST56322
            Contact: Venter, JC
            The Institute for Genomic Research
            932 Clopper Rd, Gaithersburg, MD 20878
            Tel: 3018699056
            Fax: 3018699423
            Email: tdbinfo@tdb.tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please contact the TIGR Database
            (tdbinfo@tdb.tigr.org).
FEATURES              Location/Qualifiers
     source           1. .325
                      /organism="Homo sapiens"
```



```
     mRNA             <1. .>325
BASE COUNT       73 a     50 c     76 g    124 t      2 others
ORIGIN

  Query Match           11.6%;  Score 17;  DB 16;  Length 325;
  Best Local Similarity 100.0%;  Pred. No. 1.06e-03;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      22 ATTTTAATATAATCTGT 38
           |||||||||||||||||
Cp     134 ATTTTAATATAATCTGT 118


RESULT   41
LOCUS       T09911        340 bp    DNA             EST       29-NOV-1993
DEFINITION  0461m3 Plasmodium falciparum genomic clone 0461m.
ACCESSION   T09911
NID         g319743
KEYWORDS    EST.
SOURCE      Malaria, falciparum clone=0461m library=gmbPfHB3.1, G. Roman Reddy
            vector=pBlueScript SK(+) host=E. coli XL1-Blue primer=T3 Genomic
            DNA, from asynchronous blood stage parasites of the cloned Honduran
            HB3 isolate cultured in vitro, was digested with mung bean nuclease
            in the presence of 30% formamide at 50oC (Vernick, K.D., Imberski,
            R.B., and McCutchan, T.F. 1988. Nucleic Acids Research
            16:6883-6896).  The ends of the fragments were polished using T4
            DNA polymerase, and the fragments were ligated to EcoR V-cleaved
            and dephosphorylated pBlueScript SK(+).  Recombinant plasmids
            transformed E. coli XL1-Blue.
  ORGANISM  Plasmodium falciparum
            Eucaryotae; Protozoa; Apicomplexa; Sporozoa; Coccidia;
            Eucoccidiida; Haemosporina; Plasmodium.
REFERENCE   1  (bases 1 to 340)
  AUTHORS   Reddy,G.R., Chakrabarti,D., Schuster,S.M., Perl,R.J., Almira,E.C.
            and Dame,J.B.
  TITLE     Gene sequence tags from Plasmodium falciparum genomic DNA fragments
            prepared by the genease activity of mung bean nuclease
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 90, 9867-9871 (1993)
COMMENT     *Other_ESTs: 0461m7
            Contact: Dame JB
            Department of Infectious Diseases
            University of Florida
            Bldg. 471 Mowry Rd., Gainesville, FL  32611-0880
            Tel: 9043924700
            Fax: 9043929704
            Email: dame@icbr.ifas.ufl.edu.
FEATURES              Location/Qualifiers
     source           1. .340
                      /organism="Plasmodium falciparum"
                      /clone="0461m"
BASE COUNT      139 a     29 c     27 g    135 t     10 others
```



```
ORIGIN

  Query Match           11.6%;  Score 17;  DB 5;  Length 340;
  Best Local Similarity 100.0%;  Pred. No. 1.06e-03;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db       6 ATTATTATTTTAATATA 22
           |||||||||||||||||
Cp     140 ATTATTATTTTAATATA 124


RESULT   42
LOCUS       T16288        346 bp    mRNA            EST       25-JUL-1996
DEFINITION  NIB1062 Normalized infant brain, Bento Soares Homo sapiens cDNA
            3'end, mRNA sequence.
ACCESSION   T16288
NID         g518450
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 346)
  AUTHORS   Berry,R., Stevens,T.J., Walter,N.A.R., Wilcox,A.S., Rubano,T.,
            Hopkins,J.A., Weber,J., Goold,R., Soares,M.B. and Sikela,J.M.
  TITLE     Gene-based Sequence Tagged Sites (STSs) as the basis for a human
            gene map
  JOURNAL   Nature Genet. 10, 415-423 (1995)
  MEDLINE   95400322
COMMENT
            Contact: Sikela JM
            Department of Pharmacology
            University of Colorado Health Sciences Center
            Box C236, 4200 E. 9th Ave, Denver CO 80262-0236
            Tel: 3032708637
            Fax: 3032707097
            Email: nikki@tally.uchsc.edu
            Seq primer: -21M13 Universal.
FEATURES              Location/Qualifiers
     source           1. .346
                      /organism="Homo sapiens"
                      /note="Vector: BA, M13-derived; Site_1: HindIII; Site_2:
                      NotI; The normalized infant brain library, constructed by
                      Bento Soares, Columbia University, was oligo-(dT) primed
                      and directionally cloned into an M13-derived plasmid using
                      total brain mRNA from a 72-day old human female afflicted
                      with spinal muscular atrophy. The library was normalized
                      as described elsewhere."
                      /db_xref="taxon:9606"
                      /clone_lib="Normalized infant brain, Bento Soares"
                      /lab_host="E. coli DH5-alpha"
```



```
     mRNA            complement(<1. .>346)
BASE COUNT       77 a     55 c     83 g    129 t      2 others
ORIGIN

  Query Match           11.6%;  Score 17;  DB 21;  Length 346;
  Best Local Similarity 100.0%;  Pred. No. 1.06e-03;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      22 ATTTTAATATAATCTGT 38
           |||||||||||||||||
Cp     134 ATTTTAATATAATCTGT 118


RESULT   43
LOCUS       H16570        351 bp    mRNA            EST       29-JUN-1995
DEFINITION  ym26b04.s1 Homo sapiens cDNA clone 49255 3'.
ACCESSION   H16570
NID         g882795
KEYWORDS    EST.
SOURCE      human clone=49255 library=Soares infant brain 1NIB vector=Lafmid BA
            host=DH10B (ampicillin resistant) primer=Promega -21m13 Rsite1=Not
            I Rsite2=Hind III Whole brain from a 73 days post natal female. 1st
            strand cDNA was primed with a Not I - oligo(dT) primer (5'
            AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTTT 3'); double-stranded
            cDNA was ligated to Hind III adaptors (Pharmacia), digested with
            Not I and directionally cloned into the Not I and Hind III sites of
            the Lafmid BA vector. Library went through one round of
            normalization. Library constructed by Bento Soares and M.Fatima
            Bonaldo.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 351)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            GDB: G00-421-796
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
```

```
            High quality sequence stops: 323
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES              Location/Qualifiers
     source           1. .351
                      /organism="Homo sapiens"
                      /clone="49255"
BASE COUNT       59 a     90 c    135 g     58 t      9 others
ORIGIN

  Query Match           11.6%;  Score 17;  DB 16;  Length 351;
  Best Local Similarity  94.4%;  Pred. No. 1.06e-03;
  Matches   17;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db    140 ACTRAGGCCAGCCCTGGC 157
          ||| ||||||||||||||
Qy     73 ACTCAGGCCAGCCCTGGC 90


RESULT   44
LOCUS       AA928637      367 bp    mRNA            EST      07-JUL-1998
DEFINITION  om75e03.s1 NCI_CGAP_GC4 Homo sapiens cDNA clone IMAGE:1553020 3'
            similar to TR:Q61208 Q61208 POLY ;, mRNA sequence.
ACCESSION   AA928637
NID         g3076928
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 367)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Trace considered overall poor quality
            Insert Length: 1145   Std Error: 0.00
```



```
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 1.
FEATURES              Location/Qualifiers
     source           1. .367
                      /organism="Homo sapiens"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; 1st strand cDNA was prepared from 3 pooled
                      germ cell tumors, and was then primed with a Not I -
                      oligo(dT) primer. Double-stranded cDNA was ligated to Eco
                      RI adaptors (Pharmacia), digested with Not I and cloned
                      into the Not I and Eco RI sites of the modified pT7T3
                      vector.  Library is normalized. Library was constructed by
                      Bento Soares and M. Fatima Bonaldo."
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1553020"
                      /clone_lib="NCI_CGAP_GC4"
                      /tissue_type="pooled germ cell tumors"
                      /lab_host="DH10B"
BASE COUNT      117 a     73 c     64 g    113 t
ORIGIN

  Query Match           11.6%;  Score 17;  DB 15;  Length 367;
  Best Local Similarity 100.0%;  Pred. No. 1.06e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      9 TTATTTTAATATAATCT 25
          |||||||||||||||||
Cp    136 TTATTTTAATATAATCT 120


RESULT   45
LOCUS       AA219483      375 bp    mRNA            EST      07-FEB-1997
DEFINITION  zq17e10.r1 Stratagene fetal retina 937202 Homo sapiens cDNA clone
            629994 5' similar to contains Alu repetitive element;, mRNA
            sequence.
ACCESSION   AA219483
NID         g1833558
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 375)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
            Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
```



```
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 307.
FEATURES              Location/Qualifiers
     source           1. .375
                      /organism="Homo sapiens"
                      /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                      XhoI; Cloned unidirectionally.  Primer: Oligo dT. Pooled
                      retinal tissue. Average insert size: 1.0 kb; Uni-ZAP XR
                      Vector; -5' adaptor sequence: 5' GAATTCGGCACGAG 3' -3'
                      adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTTT 3'"
                      /db_xref="taxon:9606"
                      /clone="629994"
                      /clone_lib="Stratagene fetal retina 937202"
                      /sex="mixed"
                      /lab_host="SOLR (kanamycin resistant)"
     mRNA             <1. .>375
BASE COUNT       84 a     71 c     79 g    141 t
ORIGIN

  Query Match           11.6%;  Score 17;  DB 23;  Length 375;
  Best Local Similarity 100.0%;  Pred. No. 1.06e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    189 CTTATTTATGGAGAAAA 205
          |||||||||||||||||
Qy     21 CTTATTTATGGAGAAAA 37


RESULT   46
LOCUS       AA810850      377 bp    mRNA            EST      19-FEB-1998
DEFINITION  oa76f06.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1318211,
            mRNA sequence.
ACCESSION   AA810850
NID         g2880461
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 377)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
```



```
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
            Ph.D., Gerald Marti, M.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 1065   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 329.
FEATURES              Location/Qualifiers
     source           1. .377
                      /organism="Homo sapiens"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                      was prepared from human tonsillar cells enriched for
                      germinal center B cells by flow sorting (CD20+, IgD-),
                      provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                      (NCI) and Dr. Gerald Marti (CBER).  cDNA synthesis was
                      primed with a Not I - oligo(dT) primer
                      [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTT-
                      3']. Double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Not I and cloned into the Not I
                      and Eco RI sites of the modified pT7T3 vector. Library
                      went through one round of normalization, and was
                      constructed by Bento Soares and M. Fatima Bonaldo."
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1318211"
                      /clone_lib="NCI_CGAP_GCB1"
                      /tissue_type="germinal center B cell"
                      /lab_host="DH10B"
BASE COUNT      115 a     64 c     63 g    135 t
ORIGIN

  Query Match           11.6%;  Score 17;  DB 13;  Length 377;
  Best Local Similarity 100.0%;  Pred. No. 1.06e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     36 TTATTTTAATATAATCT 52
          |||||||||||||||||
Cp    136 TTATTTTAATATAATCT 120
```

```
RESULT   47
LOCUS       AA536015      407 bp    mRNA            EST       21-AUG-1997
DEFINITION  nj81a01.s1 NCI_CGAP_Pr10 Homo sapiens cDNA clone IMAGE:998856
            similar to contains Alu repetitive element;contains element MER4
            PTR5 repetitive element ;, mRNA sequence.
ACCESSION   AA536015
NID         g2280268
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 407)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: W. Marston Linehan, M.D., Rodrigo F. Chuaqui,
            M.D., Michael R. Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 542   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 342.
FEATURES              Location/Qualifiers
     source           1. .407
                      /organism="Homo sapiens"
                      /note="Organ: prostate; Vector: pAMP10; mRNA made from
                      invasive prostate tumor, cDNA made by oligo-dT priming.
                      Non-directionally cloned.  Size-selected on agarose gel,
                      average insert  size 600 bp.  Library made by D. Krizman,
                      NIH."
                      /db_xref="taxon:9606"
                      /clone="IMAGE:998856"
                      /clone_lib="NCI_CGAP_Pr10"
                      /sex="male"
                      /tissue_type="invasive prostate tumor"
                      /lab_host="DH10B"
     mRNA             <1. .>407
BASE COUNT       93 a    103 c     88 g    123 t
ORIGIN
```



```
  Query Match           11.6%;  Score 17;  DB 9;  Length 407;
  Best Local Similarity 100.0%;  Pred. No. 1.06e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    360 AAATAATAATGTACAAA 376
          |||||||||||||||||
Qy    131 AAATAATAATGTACAAA 147


RESULT   48
LOCUS       AA716223      410 bp    mRNA            EST       29-DEC-1997
DEFINITION  zg60e07.s1 Soares fetal heart NbHH19W Homo sapiens cDNA clone
            397764 3', mRNA sequence.
ACCESSION   AA716223
NID         g2728497
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 410)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
            Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
            Theising,B., White,T., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 402.
FEATURES              Location/Qualifiers
     source           1. .410
                      /organism="Homo sapiens"
                      /note="Organ: heart; Vector: pT7T3D (Pharmacia) with a
                      modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                      strand cDNA was primed with a Not I - oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCATCTTTTTTTTTTTTTTTT 3'],
                      double-stranded cDNA was size selected, ligated to Eco RI
                      adapters (Pharmacia), digested with Not I and cloned into
                      the Not I and Eco RI sites of a modified pT7T3 vector
                      (Pharmacia). Library went through one round of
                      normalization to a Cot = 5. Library constructed by
```



```
                      M.Fatima Bonaldo. This library was constructed from the
                      same fetus as the fetal lung library, Soares fetal lung
                      NbHL19W."
                      /db_xref="GDB:1305587"
                      /db_xref="taxon:9606"
                      /clone="397764"
                      /clone_lib="Soares fetal heart NbHH19W"
                      /sex="unknown"
                      /dev_stage="19 weeks"
                      /lab_host="DH10B (ampicillin resistant)"
BASE COUNT       123 a     96 c     55 g    136 t
ORIGIN

  Query Match           11.6%;  Score 17;  DB 11;  Length 410;
  Best Local Similarity 100.0%;  Pred. No. 1.06e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    334 ATTTTAATATAATCTGT 350
          |||||||||||||||||
Cp    134 ATTTTAATATAATCTGT 118


RESULT   49
LOCUS       B81278        417 bp    DNA             GSS       24-JUN-1998
DEFINITION  CIT-HSP-2023F15.TFB CIT-HSP Homo sapiens genomic clone 2023F15,
            genomic survey sequence.
ACCESSION   B81278
NID         g2868301
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 417)
  AUTHORS   Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K.,
            Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
            Simon,M. and Venter,J.C.
  TITLE     Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: CIT-HSP-2023F15.TR
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
```



```
            Seq primer: M13-21
            Class: BAC ends.
FEATURES              Location/Qualifiers
     source           1. .417
                      /organism="Homo sapiens"
                      /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                      HindIII"
                      /db_xref="taxon:9606"
                      /clone="2023F15"
                      /clone_lib="CIT-HSP"
                      /sex="Male"
                      /cell_type="Sperm"
BASE COUNT       128 a     72 c     69 g    148 t
ORIGIN

  Query Match           11.6%;  Score 17;  DB 27;  Length 417;
  Best Local Similarity 100.0%;  Pred. No. 1.06e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    182 TACATTATTATTTTAAT 198
          |||||||||||||||||
Cp    143 TACATTATTATTTTAAT 127


RESULT   50
LOCUS       AA128639      418 bp    mRNA            EST       27-NOV-1996
DEFINITION  zo04c03.s1 Stratagene endothelial cell 937223 Homo sapiens cDNA
            clone 566692 3' similar to TR:G1256822 G1256822 POLY(A) POLYMERASE
            III. ;, mRNA sequence.
ACCESSION   AA128639
NID         g1688582
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 418)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
            Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
```

```
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Possible reversed clone: similarity on wrong strand
            Seq primer: -40M13 fwd. from Amersham
            High quality sequence stop: 281.
FEATURES             Location/Qualifiers
     source          1..418
                     /organism="Homo sapiens"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; Cloned unidirectionally.  Primer: Oligo dT.
                     Umbilical vein endothelial cells, passaged once. Average
                     insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor
                     sequence: 5' GAATTCGGCACGAG 3' -3' adaptor sequence: 5'
                     CTCGAGTTTTTTTTTTTTTTTTT 3'"
                     /db_xref="taxon:9606"
                     /clone="566692"
                     /clone_lib="Stratagene endothelial cell 937223"
                     /dev_stage="umbilical vein, 1 passage"
                     /lab_host="SOLR (kanamycin resistant)"
     mRNA            complement(<1..>418)
BASE COUNT      138 a     78 c     68 g    131 t      3 others
ORIGIN

  Query Match           11.6%;  Score 17;  DB 22;  Length 418;
  Best Local Similarity 100.0%;  Pred. No. 1.06e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     19 TTATTTTAATATAATCT 35
          |||||||||||||||||
Cp    136 TTATTTTAATATAATCT 120


Search completed: Tue Dec 22 20:03:13 1998
Job time : 356 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Tue Dec 22 20:03:33 1998; MasPar time 35.80 Seconds
558.646 Million cell updates/sec

Tabular output not generated.

Title: >US-08-530-112A-709
Description: (1-147) from US08530112A.seq
Perfect Score: 147
N.A. Sequence: 1 GATCCCCAGGGTTTCTTTGT.....TTAAAATAATAATGTACAAA 147
Comp: CTAGGGGTCCCAAAGAAACA.....AATTTTATTATTACATGTTT

Scoring table: TABLE jmetric
Gap 60

Nmatch STD : Dbase 0; Query 0

Searched: 188442 seqs, 68026449 bases x 2

Post-processing: Minimum Match 0%
Listing first 1000 summaries

Database: n-geneseq32
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39 40:part40

Statistics: Mean 6.292; Variance 2.822; scale 2.230

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
	1	147	100.0	147	18	T19709	Human gene signature	1.55e-114
c	2	18	12.2	731	5	Q31693	RsaI restriction frag	2.37e-01
	3	17	11.6	191	8	Q61162	Human brain Expressed	1.07e+00
	4	17	11.6	789	2	Q12532	cDNA clone encoding h	1.07e+00
c	5	17	11.6	1654	29	T61680	Human alpha(1,3)-fuco	1.07e+00
c	6	17	11.6	1654	9	Q56905	pCDNA1-alpha-(1-3)Fuc	1.07e+00
c	7	17	11.6	2973	9	Q53501	Sequence of bean mali	1.07e+00
	8	17	11.6	4558	7	Q43661	Acetobacter cdg2 oper	1.07e+00
	9	16	10.9	608	40	V17547	Peptidyl prolyl cis-t	4.67e+00
	10	16	10.9	1517	7	Q47167	HPV1 L1 ORF.	4.67e+00
	11	16	10.9	1517	38	V12162	Human papillomavirus	4.67e+00
	12	16	10.9	1517	7	Q47166	BPV1 L1 ORF.	4.67e+00
	13	16	10.9	1517	39	V09855	HPV16 L1 capsid prote	4.67e+00
	14	16	10.9	1517	39	V09856	HPV16 recombinant L1	4.67e+00
	15	16	10.9	1517	38	V12161	Human papillomavirus	4.67e+00
	16	16	10.9	2526	15	Q98697	Yeast aspartyl protea	4.67e+00
	17	16	10.9	3817	33	T72814	Human SPT5 cDNA.	4.67e+00
c	18	16	10.9	3877	37	V02209	cDNA of leaderless pr	4.67e+00
c	19	16	10.9	3878	1	N81502	DNA encoding human ba	4.67e+00
c	20	16	10.9	3878	2	N70494	Sequence encoding hum	4.67e+00
c	21	16	10.9	5247	14	Q55147	Pseudomonas aeruginos	4.67e+00
c	22	16	10.9	6767	1	Q05883	Plasmid clone pUC-SK1	4.67e+00
	23	16	10.9	7902	24	T09847	Human papilloma virus	4.67e+00
	24	16	10.9	7904	35	T94724	Human papillomavirus	4.67e+00
	25	16	10.9	7904	35	T94723	Human papillomavirus	4.67e+00
	26	15	10.2	256	20	T22689	Human gene signature	1.94e+01
c	27	15	10.2	304	19	T20016	Human gene signature	1.94e+01
c	28	15	10.2	423	33	T80680	Type II topoisomerase	1.94e+01
c	29	15	10.2	423	33	T80710	Type II topoisomerase	1.94e+01
	30	15	10.2	744	9	Q52274	Soybean seed storage	1.94e+01
	31	15	10.2	758	30	T68963	Cowpox virus p35 chem	1.94e+01
	32	15	10.2	819	13	Q78979	Human immunoglobulin	1.94e+01
	33	15	10.2	1374	20	T28972	Mouse CRF receptor ge	1.94e+01
c	34	15	10.2	1892	16	Q74063	The mouse beta-actin	1.94e+01
	35	15	10.2	2020	23	T20190	Feline herpesvirus gl	1.94e+01
	36	15	10.2	2122	27	T49318	Wnt-10b gene.	1.94e+01
	37	15	10.2	2256	4	Q25698	Sequence encoding pre	1.94e+01
c	38	15	10.2	2316	21	T12938	Prophage P1-derived p	1.94e+01
	39	15	10.2	2333	11	Q66637	Human Activin recepto	1.94e+01
c	40	15	10.2	2662	1	N90703	Rhoptry membrane anti	1.94e+01
	41	15	10.2	2812	14	Q55148	Pseudomonas aeruginos	1.94e+01
c	42	15	10.2	3315	19	T18535	Receptor tyrosine kin	1.94e+01

c	43	15	10.2	3321	11	Q64897	HER4 N-terminal trunc	1.94e+01
c	44	15	10.2	3360	3	Q15130	DNA contg. Sau3AI met	1.94e+01
c	45	15	10.2	3836	13	Q73652	HSV-2 strain 25766 in	1.94e+01
c	46	15	10.2	4100	16	Q74062	The rat beta-actin ge	1.94e+01
	47	15	10.2	5198	15	Q99462	Maize Hm1 genomic DNA	1.94e+01
	48	15	10.2	5198	11	Q68433	Maize Hm1 gene.	1.94e+01
c	49	15	10.2	5501	11	Q64896	HER4.	1.94e+01
c	50	15	10.2	5501	19	T18532	Receptor tyrosine kin	1.94e+01
	51	15	10.2	5502	15	Q99864	Mouse LTBP-2 gene.	1.94e+01
	52	15	10.2	6154	10	Q56188	Wild-type Feline Herp	1.94e+01
	53	15	10.2	6170	28	T59687	Genes for (modulation	1.94e+01
	54	15	10.2	6208	13	Q81210	Feline herpesvirus-1	1.94e+01
	55	15	10.2	6387	28	T59686	Genes for (modulation	1.94e+01
c	56	14	9.5	80	20	T24075	Human gene signature	7.64e+01
	57	14	9.5	124	20	T21511	Human gene signature	7.64e+01
c	58	14	9.5	180	22	T24278	Human gene signature	7.64e+01
	59	14	9.5	229	20	T23460	Human gene signature	7.64e+01
	60	14	9.5	277	19	T20052	Human gene signature	7.64e+01
	61	14	9.5	318	22	T25977	Human gene signature	7.64e+01
	62	14	9.5	322	8	Q60013	Human brain Expressed	7.64e+01
c	63	14	9.5	334	13	Q74686	Early Ripening Tomato	7.64e+01
c	64	14	9.5	380	19	T19947	Human gene signature	7.64e+01
	65	14	9.5	389	22	T26309	Human gene signature	7.64e+01
	66	14	9.5	396	8	Q60253	Human brain Expressed	7.64e+01
	67	14	9.5	459	2	Q12772	Human polymorphic epi	7.64e+01
	68	14	9.5	541	39	V05732	Internal nucleotide s	7.64e+01
	69	14	9.5	558	4	Q24680	H23-ETA-S antigen N-t	7.64e+01
	70	14	9.5	558	4	Q24678	H23-ETA-T antigen N-t	7.64e+01
	71	14	9.5	575	1	Q05685	Islets of Langerhans	7.64e+01
	72	14	9.5	774	35	T79673	BRCA2 cancer suscepti	7.64e+01
c	73	14	9.5	1000	30	T72062	Sequence flanking mar	7.64e+01
c	74	14	9.5	1000	30	T43942	Sequence flanking mar	7.64e+01
	75	14	9.5	1105	2	N70352	DNA fragment contg. 1	7.64e+01
	76	14	9.5	1142	2	N60568	Sequence encoding gly	7.64e+01
c	77	14	9.5	1170	37	T99151	Anti-dorsalising morp	7.64e+01
	78	14	9.5	1244	17	Q99168	Human MTS2 DNA includ	7.64e+01
	79	14	9.5	1244	17	T00741	Multiple tumour suppr	7.64e+01
c	80	14	9.5	1271	13	Q80839	Xenorhabdus nematophi	7.64e+01
	81	14	9.5	1275	32	T72871	Gamma-smooth muscle i	7.64e+01
	82	14	9.5	1334	12	Q71369	Leu2-dJ allele (selec	7.64e+01
c	83	14	9.5	1374	15	Q99463	Maize HM1 gene cDNA.	7.64e+01
c	84	14	9.5	1374	11	Q68434	Maize Hm1 cDNA.	7.64e+01
	85	14	9.5	1452	4	Q29276	Encodes transmembrane	7.64e+01
	86	14	9.5	1506	22	T36070	Mullerian inhibiting	7.64e+01
	87	14	9.5	1514	17	T12244	Rat CRF2-beta recepto	7.64e+01
	88	14	9.5	1518	8	Q49764	MisrIA/misrIB.	7.64e+01
	89	14	9.5	1608	22	T34162	Fas promoter region.	7.64e+01
	90	14	9.5	1626	17	T12243	Rat CRF2-alpha recept	7.64e+01
c	91	14	9.5	1641	13	Q80319	Human 38kDa FK-506 bi	7.64e+01
	92	14	9.5	1647	14	Q83533	Sequence encoding ser	7.64e+01
	93	14	9.5	1709	4	Q29277	Encodes secreted form	7.64e+01

	94	14	9.5	1746	1	Q05360	Glycinin subunit prec	7.64e+01
	95	14	9.5	1753	17	T01051	Mouse B7-1 exon 1.	7.64e+01
	96	14	9.5	1754	3	Q20239	Gene encoding cell gr	7.64e+01
c	97	14	9.5	1758	7	Q42419	IDUA.	7.64e+01
	98	14	9.5	1763	2	Q12771	Human polymorphic epi	7.64e+01
	99	14	9.5	1775	3	Q15134	Clone ptACC1 encoding	7.64e+01
	100	14	9.5	1780	3	Q15138	Clone ptACC5 encoding	7.64e+01
	101	14	9.5	1781	3	Q15137	Clone ptACC4 encoding	7.64e+01
	102	14	9.5	1783	3	Q15136	Clone ptACC3 encoding	7.64e+01
	103	14	9.5	1793	3	Q15139	Clone ptACC6 encoding	7.64e+01
	104	14	9.5	1800	39	V09713	ACC synthase DNA.	7.64e+01
	105	14	9.5	1816	25	T42315	Oilpalm ACP thioester	7.64e+01
	106	14	9.5	1818	3	Q15135	Clone ptACC2 encoding	7.64e+01
	107	14	9.5	2003	29	T59619	G-protein parathyroid	7.64e+01
	108	14	9.5	2160	11	Q66641	Mouse Activin recepto	7.64e+01
	109	14	9.5	2167	37	T94055	Human multidrug resis	7.64e+01
c	110	14	9.5	2225	37	T99152	Anti-dorsalising morp	7.64e+01
c	111	14	9.5	2226	12	Q72270	Rat MR22 serotonin re	7.64e+01
	112	14	9.5	2230	40	V15704	Tomato ACC synthase c	7.64e+01
	113	14	9.5	2471	15	Q93879	Fas-delta-TM cDNA.	7.64e+01
c	114	14	9.5	2483	30	T63319	Yeast mitochondria CO	7.64e+01
c	115	14	9.5	2486	22	T09866	Human neurotransmitte	7.64e+01
	116	14	9.5	2508	31	T67197	Zebrafish retinoid X	7.64e+01
	117	14	9.5	2534	5	Q29959	Human cell surface an	7.64e+01
	118	14	9.5	2534	21	T16703	hFas coding sequence	7.64e+01
	119	14	9.5	2534	16	Q95297	Plasmid pP58 contg. h	7.64e+01
c	120	14	9.5	2552	13	Q73756	Gene encoding melon p	7.64e+01
c	121	14	9.5	2811	24	T45660	Rhodopseudomonas sp.	7.64e+01
	122	14	9.5	3305	30	T68806	Mutant levansucrase s	7.64e+01
	123	14	9.5	3854	37	T97610	Human E2A-binding pro	7.64e+01
c	124	14	9.5	4065	1	N92515	Sequence of insectici	7.64e+01
c	125	14	9.5	4184	2	N96139	Bacillus thuringiensi	7.64e+01
c	126	14	9.5	4212	39	V01098	Human propyl-4-hydrol	7.64e+01
c	127	14	9.5	4704	1	Q05264	Sequence encoding bea	7.64e+01
	128	14	9.5	5361	32	T78868	P. falciparum liver s	7.64e+01
	129	14	9.5	5674	5	Q32352	MAGE-1 nucleic acid.	7.64e+01
	130	14	9.5	5674	13	Q72477	Tumour rejection anti	7.64e+01
	131	14	9.5	5724	17	Q98902	Tumour rejection anti	7.64e+01
	132	14	9.5	6152	32	T78867	P. falciparum liver s	7.64e+01
c	133	14	9.5	6340	28	T48871	IamS expression vecto	7.64e+01
	134	14	9.5	7100	1	Q03418	Complete synthetic hu	7.64e+01
	135	14	9.5	7100	1	Q03360	Complete cDNA of huma	7.64e+01
	136	14	9.5	7102	2	N60044	Sequence of rhinoviru	7.64e+01
	137	14	9.5	7102	2	N81393	Rhinovirus 2 (HRV2) g	7.64e+01
c	138	14	9.5	7208	28	T48870	IamH expression vecto	7.64e+01
	139	14	9.5	7241	3	Q15140	Genomic clone LE-ACC2	7.64e+01
	140	14	9.5	7244	40	V15705	Tomato ACC synthase L	7.64e+01
c	141	14	9.5	7620	6	Q39286	Glucocerebrosidase ge	7.64e+01
c	142	14	9.5	7625	31	T71323	Hybrid activation vec	7.64e+01
	143	14	9.5	7859	7	Q44265	pSW6 for expression o	7.64e+01
	144	14	9.5	7859	2	Q12154	Shuttle vector pSW6.	7.64e+01

	145	14	9.5	7984	4	Q25185	pSW6 expression vecto 7.64e+01
c	146	14	9.5	8585	28	T48867	Brassica napus micros 7.64e+01
c	147	14	9.5	8585	1	Q05749	Microspore-specific c 7.64e+01
	148	14	9.5	8808	22	T32227	Plasmid pBE146. 7.64e+01
c	149	14	9.5	9096	38	T88015	Human interleukin 16 7.64e+01
	150	14	9.5	9100	27	T28521	H. influenzae detecti 7.64e+01
	151	14	9.5	9278	13	Q76124	Human MDC genomic DNA 7.64e+01
	152	14	9.5	10798	4	Q28505	Tomato vacuolar inver 7.64e+01
	153	14	9.5	10965	4	Q28507	Tomato vacuolar inver 7.64e+01
c	154	14	9.5	13633	7	Q40862	VAC-beta (genomic - p 7.64e+01
c	155	14	9.5	13747	8	Q47342	Myotonic dystrophy ge 7.64e+01
c	156	14	9.5	14176	34	T84564	Swinepox virus HindII 7.64e+01
	157	14	9.5	14704	3	Q20685	PKS 741 insert contg. 7.64e+01
c	158	14	9.5	14704	3	Q20685	PKS 741 insert contg. 7.64e+01
	159	14	9.5	15144	29	T47715	Mouse bone morphogene 7.64e+01
	160	14	9.5	15155	3	N50107	DNA encoding Factor-V 7.64e+01
c	161	14	9.5	15328	13	Q81139	HPLA2-8 gene. 7.64e+01
	162	14	9.5	15397	29	T58635	T-DNA insert of 1345- 7.64e+01
c	163	14	9.5	15397	29	T58635	T-DNA insert of 1345- 7.64e+01
c	164	14	9.5	20303	31	T71699	Human deoxycytidylate 7.64e+01
	165	14	9.5	22481	24	T11658	PEDF full length sequ 7.64e+01
c	166	14	9.5	26764	31	T71696	Human deoxycytidylate 7.64e+01
	167	13	8.8	20	14	Q84204	PKC-eta antisense oli 2.84e+02
	168	13	8.8	20	8	Q49700	PKC-eta translation i 2.84e+02
	169	13	8.8	20	14	Q97918	PNA oligomer targetti 2.84e+02
c	170	13	8.8	29	11	Q66000	COL2A1 3'-primer (exo 2.84e+02
	171	13	8.8	32	36	T79531	Interleukin-16 gene p 2.84e+02
	172	13	8.8	33	14	Q87248	Primer for leader seq 2.84e+02
	173	13	8.8	41	33	T62891	Human or monkey kappa 2.84e+02
	174	13	8.8	41	34	T92220	Monkey/human kappa li 2.84e+02
	175	13	8.8	41	35	T95144	Human or monkey Ig ka 2.84e+02
	176	13	8.8	41	38	V05666	Human/monkey kappa li 2.84e+02
	177	13	8.8	41	6	Q35919	Human/monkey kappa li 2.84e+02
c	178	13	8.8	41	28	T51515	Fragment of cDNA enco 2.84e+02
	179	13	8.8	42	33	T62196	Primer PP161 for C. b 2.84e+02
c	180	13	8.8	83	18	T19435	Human gene signature 2.84e+02
c	181	13	8.8	129	10	Q76608	Human genome fragment 2.84e+02
c	182	13	8.8	135	19	T20310	Human gene signature 2.84e+02
c	183	13	8.8	178	20	T21721	Human gene signature 2.84e+02
c	184	13	8.8	189	22	T26476	Human gene signature 2.84e+02
	185	13	8.8	194	33	T62164	Promoter from Candida 2.84e+02
	186	13	8.8	198	22	T23019	Human gene signature 2.84e+02
	187	13	8.8	210	29	T48084	EGFL4 DNA. 2.84e+02
	188	13	8.8	210	10	Q62845	EGFL4. 2.84e+02
	189	13	8.8	210	10	Q58326	EGFL4. 2.84e+02
	190	13	8.8	210	20	T30992	Epidermal growth fact 2.84e+02
	191	13	8.8	210	13	Q74917	Human epidermal like 2.84e+02
	192	13	8.8	210	19	T06736	Epidermal growth fact 2.84e+02
	193	13	8.8	218	8	Q59129	Human brain Expressed 2.84e+02
	194	13	8.8	218	6	Q39717	Expressed Sequence Ta 2.84e+02
	195	13	8.8	246	21	T24422	Human gene signature 2.84e+02

	196	13	8.8	253	21	T24716	Human gene signature 2.84e+02
	197	13	8.8	257	22	T22954	Human gene signature 2.84e+02
c	198	13	8.8	258	4	Q24952	HIV-1 vpu gene amplif 2.84e+02
	199	13	8.8	279	38	V02150	Human secreted protei 2.84e+02
	200	13	8.8	279	39	T88082	3' portion of cDNA cl 2.84e+02
c	201	13	8.8	279	18	T19584	Human gene signature 2.84e+02
c	202	13	8.8	281	22	T22935	Human gene signature 2.84e+02
c	203	13	8.8	297	18	T20994	Human gene signature 2.84e+02
	204	13	8.8	325	8	Q59615	Human brain Expressed 2.84e+02
	205	13	8.8	326	8	Q60540	Human brain Expressed 2.84e+02
	206	13	8.8	350	8	Q61438	Human brain Expressed 2.84e+02
c	207	13	8.8	351	19	T20540	Human gene signature 2.84e+02
	208	13	8.8	359	6	Q39676	Expressed Sequence Ta 2.84e+02
	209	13	8.8	359	8	Q59088	Human brain Expressed 2.84e+02
	210	13	8.8	363	8	Q59485	Human brain Expressed 2.84e+02
c	211	13	8.8	378	8	Q60725	Human brain Expressed 2.84e+02
c	212	13	8.8	417	33	T80698	Type II topoisomerase 2.84e+02
c	213	13	8.8	423	33	T85012	Type II topoisomerase 2.84e+02
c	214	13	8.8	423	33	T85010	Type II topoisomerase 2.84e+02
c	215	13	8.8	423	33	T85005	Type II topoisomerase 2.84e+02
c	216	13	8.8	423	33	T80711	Type II topoisomerase 2.84e+02
c	217	13	8.8	423	33	T80676	Type II topoisomerase 2.84e+02
c	218	13	8.8	423	33	T80685	Type II topoisomerase 2.84e+02
c	219	13	8.8	423	33	T80700	Type II topoisomerase 2.84e+02
	220	13	8.8	507	19	T14354	Nsk2 receptor cDNA al 2.84e+02
	221	13	8.8	514	13	Q78961	Human immunoglobulin 2.84e+02
c	222	13	8.8	523	17	Q98945	bg28 insert encoding 2.84e+02
c	223	13	8.8	525	17	Q98944	Partial bg28 insert e 2.84e+02
c	224	13	8.8	564	16	Q89755	Plasmid pAR6 partial 2.84e+02
c	225	13	8.8	564	15	Q87557	Human T cell lymphotr 2.84e+02
c	226	13	8.8	596	12	Q70065	Sequence encoding ant 2.84e+02
c	227	13	8.8	597	16	Q92942	Natural resistance-as 2.84e+02
	228	13	8.8	606	13	Q78960	Human immunoglobulin 2.84e+02
	229	13	8.8	642	33	T62165	Promoter from Candida 2.84e+02
c	230	13	8.8	700	14	Q89191	Birch pollen P14 alle 2.84e+02
c	231	13	8.8	700	3	Q22352	P14(T) allergen clone 2.84e+02
	232	13	8.8	800	24	T38528	Chlamydia trachomatis 2.84e+02
	233	13	8.8	801	24	T38529	Chlamydia trachomatis 2.84e+02
	234	13	8.8	805	24	T38530	Chlamydia trachomatis 2.84e+02
	235	13	8.8	805	24	T13820	Chlamydia trachomatis 2.84e+02
	236	13	8.8	808	24	T38531	Chlamydia trachomatis 2.84e+02
	237	13	8.8	819	21	T12557	DNA encoding LIP3. 2.84e+02
	238	13	8.8	839	33	T62166	Promoter from Candida 2.84e+02
c	239	13	8.8	861	34	T90376	Human cytostatin I ge 2.84e+02
c	240	13	8.8	861	35	T73590	Human cytostatin I ge 2.84e+02
	241	13	8.8	898	24	T38523	Chlamydia trachomatis 2.84e+02
	242	13	8.8	900	32	T77777	Fragment of SR Ca2+ r 2.84e+02
	243	13	8.8	900	29	T60094	Human SR calcium ion 2.84e+02
c	244	13	8.8	900	34	T72329	Staphylococcus aureus 2.84e+02
	245	13	8.8	918	26	T28561	Bacterial antibiotic 2.84e+02
c	246	13	8.8	919	35	T47510	Partial P. rubrum csc 2.84e+02

	247	13	8.8	938	12	Q67721	Nematode-inducible tr 2.84e+02
c	248	13	8.8	938	12	Q67722	Nematode-inducible tr 2.84e+02
c	249	13	8.8	944	37	T91981	Human cytostatin III 2.84e+02
c	250	13	8.8	960	5	Q30850	3' UTR of Type III pr 2.84e+02
c	251	13	8.8	989	33	T62167	3'-UTR from Candida b 2.84e+02
c	252	13	8.8	1044	35	T64686	cDNA for form 2 of pr 2.84e+02
c	253	13	8.8	1044	18	T06881	Prostate/colon tumour 2.84e+02
c	254	13	8.8	1047	35	T64685	cDNA for form 1 of pr 2.84e+02
c	255	13	8.8	1047	18	T06880	Prostate/colon tumour 2.84e+02
	256	13	8.8	1050	32	T77772	WT human SR Ca2+ rele 2.84e+02
	257	13	8.8	1050	29	T60090	Human SR calcium ion 2.84e+02
	258	13	8.8	1050	29	T60089	Human SR calcium ion 2.84e+02
	259	13	8.8	1050	32	T77773	Mutant human SR Ca2+ 2.84e+02
	260	13	8.8	1064	15	Q87093	Tpa-1 phage insert. 2.84e+02
	261	13	8.8	1064	29	T63677	Peroxidase gene, Tpa- 2.84e+02
	262	13	8.8	1100	5	Q28534	Bacillus promoter and 2.84e+02
	263	13	8.8	1110	37	V02402	Human G protein-coupl 2.84e+02
	264	13	8.8	1110	21	T18371	Human pituitary G-pro 2.84e+02
	265	13	8.8	1147	8	Q48975	Encodes RB35 protein 2.84e+02
	266	13	8.8	1152	33	T87884	Consensus human hypot 2.84e+02
c	267	13	8.8	1154	21	T16015	DNA encoding ubiquiti 2.84e+02
	268	13	8.8	1155	2	N71213	Sequence of trans-act 2.84e+02
	269	13	8.8	1173	33	T87886	Human hypothalamic re 2.84e+02
	270	13	8.8	1174	33	T87888	Human hypothalamic re 2.84e+02
	271	13	8.8	1174	33	T87889	Human hypothalamic re 2.84e+02
	272	13	8.8	1174	33	T87887	Human hypothalamic re 2.84e+02
	273	13	8.8	1195	25	T27974	Hepatitis C virus typ 2.84e+02
	274	13	8.8	1198	35	T79694	BRCA2 cancer susceptl 2.84e+02
	275	13	8.8	1212	28	T61367	Creatinine amidinohyd 2.84e+02
	276	13	8.8	1212	35	T38807	Creatine amidinohydro 2.84e+02
	277	13	8.8	1215	22	T13291	Creatine amidinohydro 2.84e+02
	278	13	8.8	1253	14	Q82973	Cluster-A cDNA encodi 2.84e+02
	279	13	8.8	1269	16	Q94256	CAIR-1 cDNA clone 21D 2.84e+02
	280	13	8.8	1270	29	T63685	Full-length Tpa-1 gen 2.84e+02
	281	13	8.8	1270	15	Q87101	The Tpa gene. 2.84e+02
	282	13	8.8	1299	17	Q99306	Genomic 5'-region of 2.84e+02
c	283	13	8.8	1305	2	Q10948	Bovine Coronavirus E3 2.84e+02
c	284	13	8.8	1305	37	T89388	Bovine coronavirus E3 2.84e+02
c	285	13	8.8	1317	34	T71706	Staphylococcus aureus 2.84e+02
	286	13	8.8	1332	30	T67087	Mouse K-ATP channel s 2.84e+02
c	287	13	8.8	1342	35	T64687	DNA for prostate/colo 2.84e+02
c	288	13	8.8	1342	35	T64688	Prostate/colon tumour 2.84e+02
c	289	13	8.8	1342	18	T06878	PTSG clone N33(7). 2.84e+02
c	290	13	8.8	1342	18	T06879	Reconstructed prostat 2.84e+02
c	291	13	8.8	1361	29	T61698	Mouse muscle creatine 2.84e+02
c	292	13	8.8	1362	34	T91873	Human YT521 gene. 2.84e+02
c	293	13	8.8	1422	16	Q95109	Human papillomavirus 2.84e+02
c	294	13	8.8	1434	24	T28132	Senescence-related ge 2.84e+02
c	295	13	8.8	1440	9	Q62299	Factor VIII. 2.84e+02
	296	13	8.8	1478	33	T62163	5'-UTR from Candida b 2.84e+02
	297	13	8.8	1494	40	V03305	Thermococcus phosphat 2.84e+02

	298	13	8.8	1526	34	T80107	Human KF-1 gene clone 2.84e+02
c	299	13	8.8	1559	37	T91855	DUB-1 enhancer/promot 2.84e+02
	300	13	8.8	1611	2	Q12528	Thymidylate phosphohy 2.84e+02
c	301	13	8.8	1611	2	Q12528	Thymidylate phosphohy 2.84e+02
	302	13	8.8	1641	34	T80102	KF-1 gene open readin 2.84e+02
c	303	13	8.8	1671	4	Q24134	50 kD subunit of SceI 2.84e+02
	304	13	8.8	1692	16	Q91099	Human herpesvirus-6 E 2.84e+02
	305	13	8.8	1692	16	Q91098	Human herpesvirus-6 M 2.84e+02
c	306	13	8.8	1699	11	Q65605	Porcine zona pellucid 2.84e+02
c	307	13	8.8	1703	16	Q93871	Bovine endothelin con 2.84e+02
	308	13	8.8	1725	40	V12205	Mouse retinoid metabo 2.84e+02
	309	13	8.8	1728	22	T34292	HEK4 binding protein 2.84e+02
	310	13	8.8	1741	33	T84300	Quail Flt4 receptor t 2.84e+02
	311	13	8.8	1770	37	T89887	Caenorhabditis frizzl 2.84e+02
	312	13	8.8	1788	13	Q78161	V-MM65 gene homologue 2.84e+02
c	313	13	8.8	1797	3	Q23001	Prostaglandin endoper 2.84e+02
c	314	13	8.8	1816	17	Q98948	bg14 insert encoding 2.84e+02
c	315	13	8.8	1819	28	T59640	Human prostaglandin H 2.84e+02
c	316	13	8.8	1822	17	Q98949	bg3 insert encoding a 2.84e+02
	317	13	8.8	1824	30	T62073	Lipase encoding cDNA 2.84e+02
	318	13	8.8	1824	30	T62074	Lipase encoding cDNA 2.84e+02
	319	13	8.8	1828	34	T80109	Mouse KF-1 gene fragm 2.84e+02
	320	13	8.8	1830	17	T03731	TNF-R p75IC-binding p 2.84e+02
	321	13	8.8	1837	35	T72174	Alzheimer's disease D 2.84e+02
	322	13	8.8	1839	24	T18897	Human AL-1 cDNA. 2.84e+02
c	323	13	8.8	1850	1	Q01838	Gene encoding a polyp 2.84e+02
c	324	13	8.8	1863	29	T47705	Human FRK1 gene. 2.84e+02
	325	13	8.8	1908	2	Q11977	Bovine pancreatic cho 2.84e+02
c	326	13	8.8	1961	28	T51491	cDNA encoding p53 bin 2.84e+02
c	327	13	8.8	1980	17	Q99227	Heparinase-III gene. 2.84e+02
	328	13	8.8	2004	34	T86169	Human CART1 nucleotid 2.84e+02
	329	13	8.8	2004	1	N91031	Glycoprotein A antige 2.84e+02
	330	13	8.8	2010	1	N91379	Intron 1 from human f 2.84e+02
	331	13	8.8	2022	21	T16016	DNA encoding ubiquiti 2.84e+02
	332	13	8.8	2052	34	T80101	Mouse KF-1 gene open 2.84e+02
	333	13	8.8	2058	34	T80103	Human KF-1 gene open 2.84e+02
	334	13	8.8	2061	36	T85111	Maize protoporphyrino 2.84e+02
	335	13	8.8	2061	18	T11676	DNA encoding a protei 2.84e+02
	336	13	8.8	2061	38	V04307	Maize protox-2 cDNA. 2.84e+02
	337	13	8.8	2064	29	T59499	Human alpha-2b adrene 2.84e+02
	338	13	8.8	2064	3	Q14151	Human alpha 2 beta ad 2.84e+02
c	339	13	8.8	2064	3	Q14151	Human alpha 2 beta ad 2.84e+02
c	340	13	8.8	2064	29	T59499	Human alpha-2b adrene 2.84e+02
	341	13	8.8	2080	33	T80098	KF-1 gene. 2.84e+02
	342	13	8.8	2095	38	V12251	Novel protein from ra 2.84e+02
	343	13	8.8	2127	21	T12414	Ubi7 ubiquitin-lytic 2.84e+02
c	344	13	8.8	2129	16	Q93874	Bovine endothelin con 2.84e+02
c	345	13	8.8	2177	2	N60063	Partial Factor VII cD 2.84e+02
	346	13	8.8	2226	2	N92429	Sequence encoding amy 2.84e+02
c	347	13	8.8	2230	2	Q13380	Delta-1 dehydrogenase 2.84e+02
	348	13	8.8	2244	2	Q11658	3' Terminal of TSWV L 2.84e+02

c	349	13	8.8	2260	34	T88030	Brevibacterium flavum 2.84e+02
c	350	13	8.8	2293	35	T89279	TPRP-F1 gene promoter 2.84e+02
	351	13	8.8	2307	6	Q34672	P39-alpha and P39-bet 2.84e+02
c	352	13	8.8	2314	16	Q93860	Bovine endothelin con 2.84e+02
c	353	13	8.8	2384	38	V09624	D. melanogaster SSRP1 2.84e+02
c	354	13	8.8	2384	7	Q46543	Drosophila SSRP compo 2.84e+02
	355	13	8.8	2405	12	Q70827	Protein kinase (CKIga 2.84e+02
	356	13	8.8	2405	16	Q92962	Human HRR25-like case 2.84e+02
c	357	13	8.8	2422	40	V02230	Homo sapiens cDNA enc 2.84e+02
c	358	13	8.8	2422	13	Q80296	cDNA encoding Factor 2.84e+02
c	359	13	8.8	2430	20	T08125	Human syndecan-1 gene 2.84e+02
c	360	13	8.8	2438	2	N60065	Factor IX/Factor VII 2.84e+02
c	361	13	8.8	2440	24	T42151	Human endothelin conv 2.84e+02
c	362	13	8.8	2440	22	T31399	Human endothelin conv 2.84e+02
c	363	13	8.8	2450	2	N70605	Plasmid pRF1 encoding 2.84e+02
c	364	13	8.8	2483	2	N60064	Factor VII cDNA cf 1a 2.84e+02
	365	13	8.8	2484	27	T47139	Ionotropic glutamate 2.84e+02
c	366	13	8.8	2533	16	Q93877	Human placental endot 2.84e+02
	367	13	8.8	2559	40	T99583	Human platelet activa 2.84e+02
c	368	13	8.8	2578	17	Q90582	Glycosyl-phosphatidyl 2.84e+02
c	369	13	8.8	2578	3	Q22725	Bovine liver GPI-PLD 2.84e+02
c	370	13	8.8	2694	39	V09364	GMCSF/HIV-1 chimeric 2.84e+02
c	371	13	8.8	2714	4	Q23563	Proctase-A gene. 2.84e+02
c	372	13	8.8	2720	15	Q90315	Human ECE gene. 2.84e+02
c	373	13	8.8	2720	16	Q93861	Human endothelin conv 2.84e+02
	374	13	8.8	2721	21	T34617	Operon encoding LIP1, 2.84e+02
	375	13	8.8	2754	27	T58505	H218 cDNA encoding p 2.84e+02
	376	13	8.8	2813	12	Q73464	Human Egr-2. 2.84e+02
	377	13	8.8	2817	1	N91816	EGR2 clone. 2.84e+02
c	378	13	8.8	2889	37	T96149	Bovine endothelin con 2.84e+02
	379	13	8.8	2894	25	T31792	Chicken derived neuri 2.84e+02
c	380	13	8.8	2935	34	T69801	Aureobasidin sensitiv 2.84e+02
c	381	13	8.8	2941	9	Q52490	Helminth aminopeptida 2.84e+02
c	382	13	8.8	2945	16	T05127	HIV virus-1 gp160 env 2.84e+02
	383	13	8.8	3064	3	Q20323	Cynomolgus CD4 sequen 2.84e+02
	384	13	8.8	3064	2	N90642	Sequence of T4 lympho 2.84e+02
c	385	13	8.8	3084	39	V09365	HIV-1/CD4 chimeric nu 2.84e+02
	386	13	8.8	3113	2	Q13870	Cruciferin A gene. 2.84e+02
c	387	13	8.8	3134	17	Q98950	bg8 insert encoding a 2.84e+02
c	388	13	8.8	3156	2	N60128	Sequence of the envel 2.84e+02
	389	13	8.8	3261	34	T80100	Mouse KF-1 gene. 2.84e+02
c	390	13	8.8	3273	13	Q78178	Ammonia monoxygenase 2.84e+02
	391	13	8.8	3349	33	T51256	Human AD4 gene genomi 2.84e+02
c	392	13	8.8	3358	9	Q52501	Helminth aminopeptida 2.84e+02
	393	13	8.8	3423	34	T80099	Human KF-1 gene. 2.84e+02
	394	13	8.8	3425	16	T01485	Tobacco root specific 2.84e+02
	395	13	8.8	3426	12	Q67725	Nematode-inducible tr 2.84e+02
	396	13	8.8	3426	7	Q38965	Sequence of TobRB7-5A 2.84e+02
	397	13	8.8	3483	1	N70691	Sequence encoding bov 2.84e+02
	398	13	8.8	3526	34	T80106	Human KF-1 gene clone 2.84e+02
	399	13	8.8	3532	31	T72711	Human inhibitor of ap 2.84e+02

	400	13	8.8	3558	22	T27028	Atrial natriuretic pe 2.84e+02
	401	13	8.8	3562	33	T62170	Candida boidinii S2 A 2.84e+02
c	402	13	8.8	3562	33	T62170	Candida boidinii S2 A 2.84e+02
	403	13	8.8	3607	38	T89990	Cryptosporidium parvu 2.84e+02
c	404	13	8.8	3633	37	T97611	Mouse E2A-binding pro 2.84e+02
	405	13	8.8	3635	13	Q80004	Sucrose-phosphate-syn 2.84e+02
	406	13	8.8	3637	1	N91475	Human topoisomerase I 2.84e+02
	407	13	8.8	3645	16	Q91432	Variant human DNA top 2.84e+02
c	408	13	8.8	3728	10	Q58708	Mouse OSF-5 coding se 2.84e+02
	409	13	8.8	3756	15	Q99863	Mouse LTBP-3 gene. 2.84e+02
c	410	13	8.8	3813	14	Q86390	Human CSBP1 cDNA. 2.84e+02
c	411	13	8.8	3821	9	Q56609	Human PGF-2/NT-3 5' f 2.84e+02
	412	13	8.8	3841	35	T79967	Presenilin-interactin 2.84e+02
c	413	13	8.8	3861	1	Q05979	Recombinant AcNPV-HIV 2.84e+02
c	414	13	8.8	3969	28	T62655	DNA encoding protein 2.84e+02
c	415	13	8.8	4020	2	N71016	Sequence of LAV/HTLV 2.84e+02
c	416	13	8.8	4113	1	Q06644	Recombinant AcNPV-HIV 2.84e+02
	417	13	8.8	4154	13	Q79037	Human SREBP-1a coding 2.84e+02
	418	13	8.8	4164	40	V12216	Mouse retinoid metabo 2.84e+02
c	419	13	8.8	4260	32	T47528	H6 promoted cytomegal 2.84e+02
	420	13	8.8	4291	18	T04702	ALVAC C3 flanking arm 2.84e+02
c	421	13	8.8	4313	6	Q38950	Mouse multidrug resis 2.84e+02
c	422	13	8.8	4320	35	T89278	TPRP-F1 gene, includi 2.84e+02
c	423	13	8.8	4375	38	T76892	Arabidopsis floral me 2.84e+02
c	424	13	8.8	4527	24	T33899	HIV-1 SP33 genome. 2.84e+02
	425	13	8.8	4565	1	Q03704	Gene encoding plant e 2.84e+02
	426	13	8.8	4565	6	Q36024	LeEP-1 genomic clone 2.84e+02
	427	13	8.8	4643	24	T43684	Medium chain-specific 2.84e+02
	428	13	8.8	4651	14	Q72913	Human receptor-type p 2.84e+02
c	429	13	8.8	4729	1	Q04043	Recombinant DNA for a 2.84e+02
	430	13	8.8	4857	30	T69962	PIPV nucleocapsid (N) 2.84e+02
c	431	13	8.8	4951	17	T05848	Polycistronic cps gen 2.84e+02
	432	13	8.8	4978	18	T02986	Bovine phospholipase 2.84e+02
	433	13	8.8	4978	12	Q74281	PLA2-I receptor entir 2.84e+02
c	434	13	8.8	5024	14	Q55139	Staphylococcus epider 2.84e+02
c	435	13	8.8	5117	16	T06027	huDEP-1 cDNA. 2.84e+02
c	436	13	8.8	5168	13	Q79074	Cytadhesin P1 gene. 2.84e+02
c	437	13	8.8	5169	10	Q55775	Mycoplasma pneumoniae 2.84e+02
c	438	13	8.8	5169	2	Q12520	M. pneumoniae P1 cyta 2.84e+02
	439	13	8.8	5197	13	Q79038	Human SREBP-2 coding 2.84e+02
c	440	13	8.8	5234	33	T47563	Cytomegalovirus gB an 2.84e+02
c	441	13	8.8	5393	6	Q34611	Plasmid pR3. 2.84e+02
	442	13	8.8	5426	25	T18210	Turkey Herpes Virus p 2.84e+02
	443	13	8.8	5433	7	Q40945	Human DNA polymerase 2.84e+02
c	444	13	8.8	5460	20	T16508	Vector pAC3A1 contg. 2.84e+02
c	445	13	8.8	5949	23	T30053	S. pneumoniae detecti 2.84e+02
	446	13	8.8	6100	11	Q79728	Human L5/3 tumour sup 2.84e+02
	447	13	8.8	6100	30	T62440	Human L5/3 growth fac 2.84e+02
c	448	13	8.8	6412	19	Q79622	Plecked Morning Glory 2.84e+02
c	449	13	8.8	6603	11	Q62470	Phosphoenol pyruvate 2.84e+02
	450	13	8.8	6658	34	T79882	Tomato immunity 2 gen 2.84e+02

	451	13	8.8	7091	33	T47530	H6 promoted cytomegal 2.84e+02
	452	13	8.8	7291	12	Q67860	ALVAC subgenomic frag 2.84e+02
	453	13	8.8	7316	1	Q04555	Plasmid p170.2 encodi 2.84e+02
c	454	13	8.8	7323	4	Q25033	Bovine RSV strain A 5 2.84e+02
	455	13	8.8	7350	30	T69982	Canarypox virus DNA c 2.84e+02
	456	13	8.8	7351	3	Q22770	Canarypox virus BglII 2.84e+02
	457	13	8.8	7351	33	T47529	Canarypox virus DNA c 2.84e+02
	458	13	8.8	7377	1	Q05607	Plasmid p170-2 includ 2.84e+02
	459	13	8.8	7377	1	Q03005	Full length T4 cDNA o 2.84e+02
c	460	13	8.8	7399	6	Q33479	HDPACK1. 2.84e+02
	461	13	8.8	7655	15	Q86896	S. cerevisiae fkal ge 2.84e+02
c	462	13	8.8	7902	24	T09847	Human papilloma virus 2.84e+02
c	463	13	8.8	7904	35	T94723	Human papillomavirus 2.84e+02
c	464	13	8.8	7904	35	T94724	Human papillomavirus 2.84e+02
c	465	13	8.8	8201	20	T12170	pJG4-5-CDK-BP cDNA cl 2.84e+02
c	466	13	8.8	8252	12	Q69229	Host cell factor gene 2.84e+02
c	467	13	8.8	8710	17	Q89650	pSC11 PIPV E1. 2.84e+02
	468	13	8.8	8855	24	T29053	Phage T4 tail fibre g 2.84e+02
c	469	13	8.8	8959	14	Q55133	Staphylococcus aureus 2.84e+02
c	470	13	8.8	9020	17	T07652	pSC11 PIPV N. 2.84e+02
c	471	13	8.8	9088	2	N60476	Sequence of lymphaden 2.84e+02
c	472	13	8.8	9192	2	Q11943	Nucleotide sequence o 2.84e+02
c	473	13	8.8	9193	2	N60365	Sequence of LAV virus 2.84e+02
c	474	13	8.8	9213	2	N60288	Sequence of the HTLV- 2.84e+02
c	475	13	8.8	9425	7	Q45921	HTLV-III DNA (Clone B 2.84e+02
c	476	13	8.8	9425	7	Q45919	HTLV-III DNA (Clone B 2.84e+02
c	477	13	8.8	9427	7	Q45920	HTLV-III DNA (Clone B 2.84e+02
	478	13	8.8	9636	12	Q67190	P. falciparum transmi 2.84e+02
c	479	13	8.8	9709	16	Q96140	HIV-1 NL4-3 genomic D 2.84e+02
c	480	13	8.8	9709	3	Q22488	HIV-1 proviral clone 2.84e+02
c	481	13	8.8	9718	19	T14180	Attenuated HIV-1 stra 2.84e+02
c	482	13	8.8	9737	5	Q31936	ARV-2 proviral DNA fr 2.84e+02
c	483	13	8.8	9737	24	T33898	HIV-1 SF2 genome. 2.84e+02
c	484	13	8.8	9737	2	N60140	Sequence of ARV-2 (9B 2.84e+02
c	485	13	8.8	9745	2	N60240	HTLV-III virus (HIV v 2.84e+02
c	486	13	8.8	9749	7	Q45922	HTLV-III DNA (Clone R 2.84e+02
	487	13	8.8	10564	22	T03854	Human fibrinogen gamm 2.84e+02
	488	13	8.8	11225	2	N91695	Sequence of fowlpox v 2.84e+02
	489	13	8.8	13104	8	Q46852	Clone of recombinant 2.84e+02
	490	13	8.8	14985	25	T43574	Human box-dependent m 2.84e+02
	491	13	8.8	14985	40	V15693	Human Bin1 gene. 2.84e+02
c	492	13	8.8	15464	2	Q11415	Ryanodine receptor ge 2.84e+02
c	493	13	8.8	15581	39	V14355	Plasmid pNLnSG11 expr 2.84e+02
	494	13	8.8	19124	32	T72882	Plasmodium var-7 gene 2.84e+02
c	495	13	8.8	19124	32	T72882	Plasmodium var-7 gene 2.84e+02
	496	13	8.8	24025	21	T17455	Mutated BRCA1 genomic 2.84e+02
	497	13	8.8	24025	21	T17515	Mutated BRCA1 genomic 2.84e+02
	498	13	8.8	24026	21	T17513	Mutated BRCA1 genomic 2.84e+02
	499	13	8.8	24026	21	T17526	Mutated BRCA1 genomic 2.84e+02
	500	13	8.8	24026	21	T17512	Mutated BRCA1 genomic 2.84e+02
	501	13	8.8	24026	21	T17527	Mutated BRCA1 genomic 2.84e+02

	502	13	8.8	24026	21	T17519	Mutated BRCA1 genomic 2.84e+02
	503	13	8.8	24026	21	T17514	Mutated BRCA1 genomic 2.84e+02
	504	13	8.8	24026	23	T32612	BRCA1, human breast a 2.84e+02
	505	13	8.8	24026	21	T17523	Mutated BRCA1 genomic 2.84e+02
	506	13	8.8	24026	21	T17522	Mutated BRCA1 genomic 2.84e+02
	507	13	8.8	24026	21	T17516	Mutated BRCA1 genomic 2.84e+02
	508	13	8.8	24026	21	T17521	Mutated BRCA1 genomic 2.84e+02
	509	13	8.8	24026	21	T17529	Mutated BRCA1 genomic 2.84e+02
	510	13	8.8	24026	21	T17530	BRCA1 genomic sequenc 2.84e+02
	511	13	8.8	24026	18	T18325	BRCA1, human breast a 2.84e+02
	512	13	8.8	24026	21	T17524	Mutated BRCA1 genomic 2.84e+02
	513	13	8.8	24026	21	T17517	Mutated BRCA1 genomic 2.84e+02
	514	13	8.8	24026	21	T17528	Mutated BRCA1 genomic 2.84e+02
	515	13	8.8	24026	21	T17518	Mutated BRCA1 genomic 2.84e+02
	516	13	8.8	24029	21	T17520	Mutated BRCA1 genomic 2.84e+02
	517	13	8.8	24031	21	T17525	Mutated BRCA1 genomic 2.84e+02
	518	13	8.8	28994	39	V15826	Genomic DNA for inter 2.84e+02
	519	13	8.8	40352	40	V02032	MAGE-B cluster DNA se 2.84e+02
	520	13	8.8	53526	40	T94101	Human PKD1 gene. 2.84e+02
	521	13	8.8	53577	28	T18551	Human polycystic kidn 2.84e+02
	522	13	8.8	53577	40	T94108	Human PKD1 locus betw 2.84e+02
	523	13	8.8	133894	17	T13635	AcNPV genomic DNA clo 2.84e+02
c	524	13	8.8	133894	17	T13635	AcNPV genomic DNA clo 2.84e+02
c	525	12	8.2	30	21	T10902	Human cytochrome P450 9.85e+02
	526	12	8.2	36	32	T48228	Chloramphenicol acety 9.85e+02
	527	12	8.2	39	7	Q39600	CAT primer, WSB 80. 9.85e+02
c	528	12	8.2	41	7	Q41566	Oligonucleotide J. 9.85e+02
c	529	12	8.2	45	12	Q69658	Human tyrosine hydrox 9.85e+02
	530	12	8.2	50	5	Q31970	P53 binding site of c 9.85e+02
c	531	12	8.2	67	21	T24754	Human gene signature 9.85e+02
	532	12	8.2	76	18	T04487	Single stranded stem 9.85e+02
c	533	12	8.2	79	21	T24548	Human gene signature 9.85e+02
	534	12	8.2	88	20	T22383	Human gene signature 9.85e+02
	535	12	8.2	93	31	T65287	Platelet derived grow 9.85e+02
c	536	12	8.2	100	11	Q62578	Candida albicans-spec 9.85e+02
c	537	12	8.2	150	5	Q30712	Tyrosine kinase clone 9.85e+02
c	538	12	8.2	170	40	V20451	Human c-src6 oncogene 9.85e+02
c	539	12	8.2	185	1	N90230	Epidermal growth fact 9.85e+02
	540	12	8.2	189	22	T26476	Human gene signature 9.85e+02
c	541	12	8.2	191	1	N91199	Chemically synthesise 9.85e+02
c	542	12	8.2	197	5	Q33733	Upstream sequence of 9.85e+02
	543	12	8.2	201	19	T21170	Human gene signature 9.85e+02
	544	12	8.2	211	19	T22716	Human gene signature 9.85e+02
	545	12	8.2	216	10	Q76652	Human genome fragment 9.85e+02
c	546	12	8.2	223	21	T24674	Human gene signature 9.85e+02
c	547	12	8.2	232	18	T19223	Human gene signature 9.85e+02
	548	12	8.2	234	19	T10851	PECAM-1 gene promoter 9.85e+02
	549	12	8.2	239	28	T40212	Fungus responsive prp 9.85e+02
c	550	12	8.2	239	10	Q58613	Sequence of OtT1 whic 9.85e+02
	551	12	8.2	246	15	Q93946	Human MSH2 gene exon 9.85e+02
c	552	12	8.2	258	8	Q59824	Human brain Expressed 9.85e+02

c	553	12	8.2	268	8	Q61018	Human brain Expressed	9.85e+02
	554	12	8.2	269	19	T20067	Human gene signature	9.85e+02
	555	12	8.2	283	13	Q78988	Human immunoglobulin	9.85e+02
	556	12	8.2	286	22	T26304	Human gene signature	9.85e+02
c	557	12	8.2	313	22	T26683	Human gene signature	9.85e+02
c	558	12	8.2	322	40	V09102	Nucleotide sequence e	9.85e+02
	559	12	8.2	323	20	T23991	Human gene signature	9.85e+02
	560	12	8.2	329	8	Q50992	Human brain Expressed	9.85e+02
	561	12	8.2	329	20	T22157	Human gene signature	9.85e+02
	562	12	8.2	330	22	T36114	S. mutans antigen I/I	9.85e+02
	563	12	8.2	339	22	T24262	Human gene signature	9.85e+02
	564	12	8.2	346	8	Q59306	Human brain Expressed	9.85e+02
c	565	12	8.2	346	40	V15425	Human gene fragment Q	9.85e+02
	566	12	8.2	354	8	Q59745	Human brain Expressed	9.85e+02
	567	12	8.2	375	8	Q59519	Human brain Expressed	9.85e+02
c	568	12	8.2	375	8	Q60498	Human brain Expressed	9.85e+02
	569	12	8.2	380	8	Q59472	Human brain Expressed	9.85e+02
c	570	12	8.2	382	29	T39958	Promoter region of L.	9.85e+02
	571	12	8.2	384	18	T19511	Human gene signature	9.85e+02
c	572	12	8.2	393	37	T72648	HSV 863 antibody ligh	9.85e+02
c	573	12	8.2	395	12	Q72688	Cauliflower mosaic vi	9.85e+02
	574	12	8.2	405	8	Q59912	Human brain Expressed	9.85e+02
c	575	12	8.2	406	37	T99198	Rat prostaglandin D s	9.85e+02
	576	12	8.2	409	24	T34604	Probe for detecting p	9.85e+02
	577	12	8.2	409	39	V01443	Yeast EF1-alpha TEF g	9.85e+02
	578	12	8.2	413	20	T23574	Human gene signature	9.85e+02
	579	12	8.2	418	33	T85996	Arabidopsis thaliana	9.85e+02
c	580	12	8.2	421	8	Q59412	Human brain Expressed	9.85e+02
c	581	12	8.2	423	33	T80746	Type II topoisomerase	9.85e+02
c	582	12	8.2	423	33	T80737	Type II topoisomerase	9.85e+02
c	583	12	8.2	426	5	Q31009	Thyroid N-acetyl-gluc	9.85e+02
c	584	12	8.2	443	10	Q64058	HUMGLBN Fragments.	9.85e+02
c	585	12	8.2	443	14	Q85272	Foetal globin genomic	9.85e+02
c	586	12	8.2	445	8	Q60718	Human brain Expressed	9.85e+02
	587	12	8.2	453	21	T11899	Yersinia enterocoliti	9.85e+02
c	588	12	8.2	461	25	T31943	Probe PARsorI, for se	9.85e+02
	589	12	8.2	462	16	T04867	Nucleotide analogue.	9.85e+02
	590	12	8.2	474	30	T67779	H. pylori membrane pr	9.85e+02
	591	12	8.2	474	30	T77459	H. pylori transmembra	9.85e+02
c	592	12	8.2	476	11	Q64470	DNA encoding Fasciola	9.85e+02
	593	12	8.2	525	27	T42240	HIV-1 group O strain	9.85e+02
c	594	12	8.2	531	31	T43457	ATM gene exon 17.	9.85e+02
	595	12	8.2	535	13	Q78953	Human immunoglobulin	9.85e+02
c	596	12	8.2	567	8	Q50050	Decorin (N-terminal t	9.85e+02
	597	12	8.2	571	37	T79179	Drosophila immunophil	9.85e+02
c	598	12	8.2	597	30	T67770	H. pylori protein ORF	9.85e+02
c	599	12	8.2	597	30	T77450	H. pylori protein ORF	9.85e+02
c	600	12	8.2	600	30	T67853	H. pylori protein ORF	9.85e+02
	601	12	8.2	637	1	N90220	Malaria-specific DNA	9.85e+02
	602	12	8.2	650	6	Q36129	HI223 MAb heavy chain	9.85e+02
	603	12	8.2	650	8	Q47374	Antibody heavy chain	9.85e+02

	604	12	8.2	657	3	Q15266	SP-C from pC149SP-C i	9.85e+02
c	605	12	8.2	659	39	V15999	NBCCS (PTC) protein E	9.85e+02
c	606	12	8.2	659	1	N90236	Leader peptide of a p	9.85e+02
	607	12	8.2	662	2	N90107	Sequence of BamHI/Hin	9.85e+02
c	608	12	8.2	681	5	Q29259	Human calcium channel	9.85e+02
c	609	12	8.2	687	40	V24813	H. pylori ORF 14eel04	9.85e+02
c	610	12	8.2	763	2	N70145	Sequence of bovine in	9.85e+02
c	611	12	8.2	765	22	T35141	DNA encoding cytokine	9.85e+02
	612	12	8.2	773	2	N71220	Sequence of the 5' fl	9.85e+02
c	613	12	8.2	783	30	T77454	H. pylori flagella as	9.85e+02
c	614	12	8.2	787	15	Q98736	Barley Ltp1 gene prom	9.85e+02
	615	12	8.2	803	3	Q15265	SP-C from pC210SP-C i	9.85e+02
c	616	12	8.2	804	14	Q87438	P. falciparum Pfs28 g	9.85e+02
	617	12	8.2	847	22	T35102	Up-regulated senescen	9.85e+02
	618	12	8.2	848	1	Q03565	SP-C expression const	9.85e+02
c	619	12	8.2	849	8	Q50052	Decorin (N-terminal t	9.85e+02
	620	12	8.2	852	15	Q91758	Rabbit alpha-d subuni	9.85e+02
c	621	12	8.2	852	40	V24784	H. pylori ORF 06ee107	9.85e+02
c	622	12	8.2	862	40	V20539	Human T-helper 2 E26	9.85e+02
c	623	12	8.2	924	38	V06123	Viral infection gene	9.85e+02
c	624	12	8.2	924	8	Q50053	Decorin (N-terminal t	9.85e+02
	625	12	8.2	940	2	Q12863	Murine Cytotoxic Cell	9.85e+02
c	626	12	8.2	959	37	T99196	Mouse prostaglandin D	9.85e+02
	627	12	8.2	978	2	Q13847	Phibacin PSBX mutant	9.85e+02
c	628	12	8.2	979	15	Q90469	Human Kunitz-type pro	9.85e+02
c	629	12	8.2	981	6	Q34763	3-alpha-hydroxy-stero	9.85e+02
	630	12	8.2	987	34	T90554	Clostridium perfringe	9.85e+02
	631	12	8.2	987	34	T90555	Clostridium perfringe	9.85e+02
c	632	12	8.2	1002	22	T35137	DNA encoding cytokine	9.85e+02
c	633	12	8.2	1026	20	T10741	Human recombinant dec	9.85e+02
c	634	12	8.2	1026	27	T49315	Hypersensitive respon	9.85e+02
	635	12	8.2	1041	29	T60569	Cotton fibrous tissue	9.85e+02
c	636	12	8.2	1047	13	Q78152	Tuber specific class	9.85e+02
c	637	12	8.2	1056	30	T67514	H. pylori secreted or	9.85e+02
c	638	12	8.2	1089	30	T68251	H. pylori secreted or	9.85e+02
c	639	12	8.2	1092	38	T91592	Rat UBCE2A coding seq	9.85e+02
c	640	12	8.2	1119	39	V15099	N-methylamino acid ox	9.85e+02
c	641	12	8.2	1125	11	Q65603	Pre-huTFh cDNA.	9.85e+02
c	642	12	8.2	1137	30	T63234	Human ubiquitin conju	9.85e+02
	643	12	8.2	1138	2	N91630	3' sequence of gamma-	9.85e+02
c	644	12	8.2	1164	8	Q48976	Nucleotides 301-1464	9.85e+02
	645	12	8.2	1180	4	Q26503	Ribosome inactivating	9.85e+02
c	646	12	8.2	1180	14	Q79144	Hepatitis C virus gen	9.85e+02
	647	12	8.2	1196	5	Q28701	Sequence of luciferas	9.85e+02
	648	12	8.2	1196	10	Q56673	Renilla reniformis lu	9.85e+02
	649	12	8.2	1201	2	Q13849	Phibacin PSBX 1.2kb t	9.85e+02
c	650	12	8.2	1229	12	Q71536	Hornet hyaluronidase	9.85e+02
c	651	12	8.2	1230	1	N80251	H-env probe for detec	9.85e+02
c	652	12	8.2	1233	12	Q70828	Protein kinase (CK1ga	9.85e+02
	653	12	8.2	1260	36	T91125	Toxoplasma gondii ant	9.85e+02
	654	12	8.2	1276	34	T85944	Leptospira alstoni Om	9.85e+02

c	655	12	8.2	1281	32	T71585	Chaetomium thermophil	9.85e+02
c	656	12	8.2	1281	11	Q64903	EPSP synthase gene fr	9.85e+02
c	657	12	8.2	1281	11	Q64904	Mutant EPSP synthase	9.85e+02
	658	12	8.2	1319	3	Q21861	Non essential inserti	9.85e+02
c	659	12	8.2	1339	1	N90598	Human lipocortin-III.	9.85e+02
	660	12	8.2	1341	12	Q77692	Alfalfa COMT clone pC	9.85e+02
c	661	12	8.2	1347	1	N91222	DNA encoding L-sorbos	9.85e+02
	662	12	8.2	1354	2	Q12862	Murine Cytotoxic Cell	9.85e+02
	663	12	8.2	1358	37	T84705	Human renin cDNA.	9.85e+02
	664	12	8.2	1378	10	Q45241	Sequence encoding the	9.85e+02
c	665	12	8.2	1390	31	T76777	Staphylococcus aureus	9.85e+02
	666	12	8.2	1397	1	Q05681	Islets of Langerhans	9.85e+02
c	667	12	8.2	1406	11	Q64831	Human brain serotonin	9.85e+02
	668	12	8.2	1459	3	N50180	Sequence encoding pre	9.85e+02
c	669	12	8.2	1480	4	N30040	Sequence encoding bov	9.85e+02
	670	12	8.2	1481	22	T33260	RAR-alpha DNA sequenc	9.85e+02
c	671	12	8.2	1487	15	Q98737	Barley Ltp1 gene.	9.85e+02
c	672	12	8.2	1494	40	V12204	Human retinoid metabo	9.85e+02
	673	12	8.2	1520	2	N70223	Plasmid pBRV2 insert.	9.85e+02
	674	12	8.2	1521	2	N60937	Plasmid pBRV2 insert	9.85e+02
	675	12	8.2	1525	1	N81478	Sequence encoding hum	9.85e+02
	676	12	8.2	1525	1	N91086	Plasmid pP12 contg. c	9.85e+02
	677	12	8.2	1530	2	N70318	Sequence encoding por	9.85e+02
	678	12	8.2	1531	2	N60936	Plasmid pBRG4 insert	9.85e+02
	679	12	8.2	1534	2	N71319	Sequence encoding hum	9.85e+02
c	680	12	8.2	1547	40	V15419	Human macrophage/dend	9.85e+02
	681	12	8.2	1550	37	T91346	Human HO-1 cDNA.	9.85e+02
	682	12	8.2	1560	1	N91775	Human neuronal alpha-	9.85e+02
	683	12	8.2	1560	14	Q83201	Precursor of novel am	9.85e+02
	684	12	8.2	1560	40	V09146	(Alpha)-2 subunit of	9.85e+02
c	685	12	8.2	1560	40	V09146	(Alpha)-2 subunit of	9.85e+02
c	686	12	8.2	1560	36	T96813	Human neuronal calciu	9.85e+02
	687	12	8.2	1570	1	Q05684	Islets of Langerhans	9.85e+02
	688	12	8.2	1571	31	T71216	Type I, p80 IL-1-rece	9.85e+02
c	689	12	8.2	1576	2	Q10226	Clone pGEM-hRARgamma	9.85e+02
c	690	12	8.2	1584	3	N50041	Sequence of the fixD	9.85e+02
c	691	12	8.2	1584	2	N70655	FixD gene.	9.85e+02
c	692	12	8.2	1602	12	Q72217	Human adrenergic rece	9.85e+02
	693	12	8.2	1611	4	Q26814	IGFBP6 DNA.	9.85e+02
c	694	12	8.2	1611	7	Q46688	Human pp60 c-src gene	9.85e+02
	695	12	8.2	1647	3	Q21621	Human Tryptophan-2,3-	9.85e+02
c	696	12	8.2	1661	38	V03875	RTm4 gene.	9.85e+02
	697	12	8.2	1675	3	Q14634	Clone associated with	9.85e+02
c	698	12	8.2	1689	23	T31334	CC-chemokine receptor	9.85e+02
c	699	12	8.2	1689	11	Q66824	AmEPV NPH-1 gene.	9.85e+02
	700	12	8.2	1704	15	Q90217	Condensing enzyme clo	9.85e+02
	701	12	8.2	1718	23	T40760	Doc2 (brain-specific	9.85e+02
	702	12	8.2	1732	5	Q29914	Helianthinin gene fra	9.85e+02
	703	12	8.2	1733	14	Q84046	Japonicum allergen cD	9.85e+02
	704	12	8.2	1733	20	T18102	Japan cedar pollen cD	9.85e+02
c	705	12	8.2	1733	15	Q90207	Jojoba fatty acyl-CoA	9.85e+02

c	706	12	8.2	1743	21	T30683	Kaposi's sarcoma asso	9.85e+02
c	707	12	8.2	1743	20	T16813	Kaposi's sarcoma asso	9.85e+02
	708	12	8.2	1749	15	Q91211	TF55 molecular chaper	9.85e+02
	709	12	8.2	1758	7	Q42419	IDUA.	9.85e+02
c	710	12	8.2	1769	34	T85973	Human NB Phox 3' untr	9.85e+02
c	711	12	8.2	1782	30	T67856	H. pylori secreted or	9.85e+02
c	712	12	8.2	1801	20	T29602	Human EHOC-1 3' untra	9.85e+02
	713	12	8.2	1830	4	Q24334	Mutant thermostable D	9.85e+02
	714	12	8.2	1833	24	T40341	Native Btt gene encod	9.85e+02
c	715	12	8.2	1838	35	T90504	Arabidopsis thaliana	9.85e+02
c	716	12	8.2	1841	15	Q91311	Human Fas-L gene.	9.85e+02
c	717	12	8.2	1851	15	Q87839	Human calcium channel	9.85e+02
c	718	12	8.2	1851	6	Q37820	Sequence encoding spl	9.85e+02
	719	12	8.2	1860	5	Q30028	Encodes lipase II.	9.85e+02
	720	12	8.2	1878	26	T40065	Membrane co-factor pr	9.85e+02
	721	12	8.2	1881	36	T84960	Human Myt1 kinase Myt	9.85e+02
	722	12	8.2	1882	36	T72942	Phaffia carotenoid bi	9.85e+02
c	723	12	8.2	1890	15	Q89781	Human alpha 1c adrene	9.85e+02
c	724	12	8.2	1903	18	T05196	Human retinoic acid r	9.85e+02
c	725	12	8.2	1905	23	T29986	Human Fas ligand gene	9.85e+02
c	726	12	8.2	1918	5	Q31539	Lignin peroxidase gen	9.85e+02
c	727	12	8.2	1924	1	N90684	Gene encoding the 43	9.85e+02
	728	12	8.2	1932	2	N70092	Sequence encoding tox	9.85e+02
	729	12	8.2	1932	8	Q47288	Delta endotoxin gene.	9.85e+02
	730	12	8.2	1932	1	N90957	M-7 toxin gene.	9.85e+02
	731	12	8.2	1947	8	Q47837	Human interleukin 9 r	9.85e+02
c	732	12	8.2	1974	38	T49506	Transferrin receptor	9.85e+02
c	733	12	8.2	1978	12	Q72212	Truncated human alpha	9.85e+02
	734	12	8.2	1979	14	Q81335	(1-3)-beta-D-glucan s	9.85e+02
	735	12	8.2	1997	8	Q47839	Human interleukin 9 r	9.85e+02
c	736	12	8.2	1998	12	Q72213	Human alpha-1C adrene	9.85e+02
c	737	12	8.2	2004	7	Q38964	Sequence encoding hum	9.85e+02
c	738	12	8.2	2005	12	Q72211	Human alpha-1C adrene	9.85e+02
	739	12	8.2	2015	2	N70047	Encodes human decay a	9.85e+02
	740	12	8.2	2071	10	Q58121	pH 2.5 acid phosphata	9.85e+02
c	741	12	8.2	2089	21	T31378	Human alpha-1C3 adren	9.85e+02
	742	12	8.2	2096	15	Q99105	Human DAF cDNA.	9.85e+02
c	743	12	8.2	2127	2	N70657	FixD gene, including	9.85e+02
	744	12	8.2	2152	4	Q25301	cDNA encoding guinea	9.85e+02
	745	12	8.2	2160	9	Q54833	Sequence encoding hum	9.85e+02
	746	12	8.2	2166	24	T38232	Apoptosis-regulating	9.85e+02
	747	12	8.2	2169	33	T79804	Potato starch branchi	9.85e+02
	748	12	8.2	2169	33	T79783	Potato starch branchi	9.85e+02
c	749	12	8.2	2182	3	Q08510	Sequence of amylase g	9.85e+02
c	750	12	8.2	2187	40	V10651	Human oncoprotein hhc	9.85e+02
c	751	12	8.2	2187	40	V10652	Human oncoprotein hhc	9.85e+02
c	752	12	8.2	2187	40	V10647	Human oncoprotein hhc	9.85e+02
c	753	12	8.2	2187	40	V10649	Human oncoprotein hhc	9.85e+02
c	754	12	8.2	2187	40	V10650	Human oncoprotein hhc	9.85e+02
c	755	12	8.2	2187	40	V10653	Human oncoprotein hhc	9.85e+02
c	756	12	8.2	2187	40	V03330	Human oncoprotein hhc	9.85e+02

c	757	12	8.2	2187	40	V10648	Human oncoprotein hhc	9.85e+02
	758	12	8.2	2196	37	T91640	Acremonium cellulolyt	9.85e+02
	759	12	8.2	2202	33	T62383	Mouse A33 antigen.	9.85e+02
	760	12	8.2	2230	21	T11238	Neisseria meningitidi	9.85e+02
	761	12	8.2	2230	9	Q55619	N.meningitidis IM2169	9.85e+02
	762	12	8.2	2230	36	T64589	Gene sequence for HTR	9.85e+02
c	763	12	8.2	2233	11	Q62552	Candida albicans-spec	9.85e+02
	764	12	8.2	2237	37	T90287	Human glycophosphatid	9.85e+02
c	765	12	8.2	2248	38	T88990	Tumour necrosis facto	9.85e+02
c	766	12	8.2	2264	17	T15542	Borrelia afzelii 66 k	9.85e+02
	767	12	8.2	2265	4	Q24332	Mutant thermostable D	9.85e+02
c	768	12	8.2	2281	8	Q47834	Murine interleukin 9	9.85e+02
c	769	12	8.2	2291	5	Q30163	pCAM-40 clone contg.	9.85e+02
c	770	12	8.2	2291	30	T67198	Bovine brain Ca2+/cal	9.85e+02
c	771	12	8.2	2291	26	T51094	61 kD brain calcium/c	9.85e+02
	772	12	8.2	2304	37	T89959	D. melanogaster XR2C	9.85e+02
	773	12	8.2	2337	39	T94212	Thermotoga maritima e	9.85e+02
	774	12	8.2	2344	36	T91638	Castorbean ricin gene	9.85e+02
c	775	12	8.2	2350	3	Q14687	Factor XIII subunit a	9.85e+02
	776	12	8.2	2352	10	Q45333	Human TLE-1 gene.	9.85e+02
	777	12	8.2	2352	10	Q62175	Human TLE-1 gene.	9.85e+02
c	778	12	8.2	2363	38	V12196	Human neuronal nicoti	9.85e+02
c	779	12	8.2	2376	1	Q03853	Hsp70 antigen gene fr	9.85e+02
c	780	12	8.2	2384	21	T34369	Plasmid pML5 (ATCC 68	9.85e+02
c	781	12	8.2	2384	3	Q14633	Clone associated with	9.85e+02
	782	12	8.2	2387	8	Q49883	SE fimbrial antigen (	9.85e+02
	783	12	8.2	2403	4	Q24331	Mutant thermostable D	9.85e+02
c	784	12	8.2	2405	12	Q70827	Protein kinase (CKIga	9.85e+02
c	785	12	8.2	2418	4	Q27886	P.falciparum GBP130h.	9.85e+02
c	786	12	8.2	2419	1	N80104	Sequence of Trypanoso	9.85e+02
	787	12	8.2	2440	2	Q06469	Sequence encoding inc	9.85e+02
	788	12	8.2	2448	30	T63320	Yeast mitochondria CO	9.85e+02
c	789	12	8.2	2448	38	V12201	Human neuronal nicoti	9.85e+02
c	790	12	8.2	2448	28	T48240	Neuronal nicotinic ac	9.85e+02
c	791	12	8.2	2461	23	T41544	Human gas6-encoding c	9.85e+02
	792	12	8.2	2464	30	T38709	Rat farnesyl transfer	9.85e+02
c	793	12	8.2	2520	20	T09941	Anther earl gene prom	9.85e+02
c	794	12	8.2	2521	3	Q20542	Mouse retinoic acid r	9.85e+02
c	795	12	8.2	2533	39	T96031	Mouse gamma II adapti	9.85e+02
	796	12	8.2	2540	13	Q79386	Human N-methyl-D-aspa	9.85e+02
	797	12	8.2	2540	31	T70836	Human apoptosis inhib	9.85e+02
	798	12	8.2	2565	33	T62369	Human A33 antigen.	9.85e+02
	799	12	8.2	2568	9	Q54396	Feline Immunodeficien	9.85e+02
	800	12	8.2	2568	4	Q28937	Encodes Taf DNA polym	9.85e+02
c	801	12	8.2	2568	4	Q28937	Encodes Taf DNA polym	9.85e+02
	802	12	8.2	2581	2	Q13077	extA gene with gene p	9.85e+02
c	803	12	8.2	2589	1	N90574	Caprine repeat elemen	9.85e+02
c	804	12	8.2	2594	2	N70104	Sequence of yeast PHO	9.85e+02
c	805	12	8.2	2600	35	T42127	Neisseria gonorrhoeae	9.85e+02
	806	12	8.2	2615	1	N81002	Bacillus thuringiensi	9.85e+02
	807	12	8.2	2629	17	T10639	Stress activated prot	9.85e+02

	808	12	8.2	2663	39	T94120	Alternatively spliced	9.85e+02
c	809	12	8.2	2677	40	V12214	Human retinoid metabo	9.85e+02
	810	12	8.2	2723	2	N70893	MYMV strand (b).	9.85e+02
c	811	12	8.2	2740	3	Q20541	Mouse retinoic acid r	9.85e+02
	812	12	8.2	2749	15	Q86687	Human NF-AT transcrip	9.85e+02
c	813	12	8.2	2763	8	Q48044	Lactobacillus lactis	9.85e+02
c	814	12	8.2	2799	16	T05334	Peptide transport gen	9.85e+02
	815	12	8.2	2821	36	T94004	DNA encoding human KV	9.85e+02
c	816	12	8.2	2829	38	T85374	Tetrahymena 95 kD tel	9.85e+02
c	817	12	8.2	2829	25	T30868	95 kD protein coding	9.85e+02
	818	12	8.2	2839	13	Q84302	Human NF-AT120, P sub	9.85e+02
	819	12	8.2	2858	29	T64391	Pichia pastoris alpha	9.85e+02
	820	12	8.2	2858	28	T62650	DNA encoding OCH1 gly	9.85e+02
	821	12	8.2	2858	28	T59328	OCH1 gene.	9.85e+02
c	822	12	8.2	2865	31	T72717	Yeast inhibitor of ap	9.85e+02
	823	12	8.2	2871	3	Q14807	B.thuringiensis toxin	9.85e+02
c	824	12	8.2	2872	15	Q98738	Barley Ltp1 gene.	9.85e+02
	825	12	8.2	2933	3	Q21695	Plasma membrane proto	9.85e+02
	826	12	8.2	2947	15	Q87269	Human MSH2 gene.	9.85e+02
	827	12	8.2	2983	12	Q71026	CryIIIA gene which en	9.85e+02
c	828	12	8.2	2994	1	Q04713	Cancer supressing gen	9.85e+02
c	829	12	8.2	2997	17	T11641	Murine osteogenic pro	9.85e+02
	830	12	8.2	2999	14	Q81041	Friend murine leukaem	9.85e+02
	831	12	8.2	3036	5	Q29334	myl/RAR-alpha fusion.	9.85e+02
c	832	12	8.2	3060	1	Q03098	Enod2a genomic clone.	9.85e+02
	833	12	8.2	3069	2	Q11772	Sequence encoding hum	9.85e+02
c	834	12	8.2	3083	23	T36036	rchd534 gene differen	9.85e+02
	835	12	8.2	3088	11	Q66768	Tomato ADH2 gene.	9.85e+02
	836	12	8.2	3088	2	Q13727	Plasmodium falciparum	9.85e+02
c	837	12	8.2	3088	16	T05628	ADP ribosylation fact	9.85e+02
	838	12	8.2	3181	11	Q62678	Human enkephalinase c	9.85e+02
	839	12	8.2	3190	6	Q38746	CD22-beta gene.	9.85e+02
	840	12	8.2	3243	1	Q05686	Islets of Langerhans	9.85e+02
	841	12	8.2	3244	12	Q70642	L-plastin promoter an	9.85e+02
	842	12	8.2	3247	28	T63080	Active clone X1 of Ig	9.85e+02
c	843	12	8.2	3288	16	T05055	Human transforming gr	9.85e+02
c	844	12	8.2	3293	3	Q13855	Murine GCSF receptor	9.85e+02
	845	12	8.2	3294	24	T41475	Haemophilus adhesion	9.85e+02
	846	12	8.2	3303	3	Q14806	B.thuringiensis toxin	9.85e+02
c	847	12	8.2	3418	20	T15006	Type II (beta-2) rat	9.85e+02
c	848	12	8.2	3418	1	N81279	Entire nucleotide seq	9.85e+02
c	849	12	8.2	3496	28	T48237	Neuronal nicotinic ac	9.85e+02
	850	12	8.2	3500	22	T18500	Yeast calcineurin-int	9.85e+02
	851	12	8.2	3511	22	T33259	PML-RAR-alpha DNA seq	9.85e+02
	852	12	8.2	3528	16	T48816	Calgene Lambda 140 ge	9.85e+02
	853	12	8.2	3535	22	T16515	Collagen A1/BMP-2B A1	9.85e+02
c	854	12	8.2	3539	33	T47537	42K promoted cytomega	9.85e+02
	855	12	8.2	3541	22	T16516	Collagen A1/TGF-beta-	9.85e+02
	856	12	8.2	3579	15	Q84669	Human neuronal calciu	9.85e+02
c	857	12	8.2	3585	15	Q84667	Human neuronal calciu	9.85e+02
	858	12	8.2	3588	6	Q40057	Bacterial delta-6-des	9.85e+02

	859	12	8.2	3588	21	T30396	Synechocystis delta-6	9.85e+02
c	860	12	8.2	3600	15	Q84664	Human neuronal calciu	9.85e+02
c	861	12	8.2	3608	1	N91777	Rabbit seletal muscle	9.85e+02
c	862	12	8.2	3649	39	V01536	Human acylcoenzyme A:	9.85e+02
c	863	12	8.2	3650	39	T96368	Human acyl-coenzyme A	9.85e+02
	864	12	8.2	3661	28	T63081	Active clone V11 of I	9.85e+02
	865	12	8.2	3731	13	Q79394	Human NMDA receptor s	9.85e+02
	866	12	8.2	3765	6	Q31890	FMR-1 gene for detect	9.85e+02
	867	12	8.2	3794	13	Q79392	Human NMDA receptor s	9.85e+02
c	868	12	8.2	3802	30	T70227	Rabbit calcium channe	9.85e+02
	869	12	8.2	3824	2	Q11482	Sequence encoding spi	9.85e+02
c	870	12	8.2	3850	23	T37414	Prostate carcinoma tu	9.85e+02
c	871	12	8.2	3901	32	T79598	TATA-binding protein	9.85e+02
c	872	12	8.2	3902	12	Q70728	TATA-binding protein-	9.85e+02
	873	12	8.2	3905	18	T06303	Tomato pathogen resis	9.85e+02
c	874	12	8.2	3919	9	Q53413	Sequence of the pcHU	9.85e+02
c	875	12	8.2	3955	17	T01590	Rat neu promoter.	9.85e+02
c	876	12	8.2	3973	12	Q78276	T. niveum alanine-rac	9.85e+02
c	877	12	8.2	3973	12	Q78281	T. niveum alanine-rac	9.85e+02
	878	12	8.2	3981	1	N91734	Sequence of a macroph	9.85e+02
c	879	12	8.2	3983	1	N90569	Ovine Y-chromosomal D	9.85e+02
	880	12	8.2	3984	29	T58627	Protein tyrosine phos	9.85e+02
	881	12	8.2	4008	10	Q64827	Leukotoxin AppIIIA ge	9.85e+02
	882	12	8.2	4031	33	T84485	Mouse alpha-1 collage	9.85e+02
	883	12	8.2	4094	13	Q79393	Human NMDA receptor s	9.85e+02
	884	12	8.2	4157	13	Q79391	Human NMDA receptor s	9.85e+02
	885	12	8.2	4175	11	Q62129	Mouse tyrosine kinase	9.85e+02
c	886	12	8.2	4253	34	T86172	Human MLN 51 nucleoti	9.85e+02
c	887	12	8.2	4268	18	T00052	Hepatitis GB virus (H	9.85e+02
c	888	12	8.2	4286	4	Q23917	Taf DNA polymerase I	9.85e+02
	889	12	8.2	4361	13	Q79390	Human NMDA receptor s	9.85e+02
	890	12	8.2	4364	10	Q63730	Protein tyrosine kina	9.85e+02
	891	12	8.2	4383	6	Q34941	Calgene Lambda 140 ge	9.85e+02
	892	12	8.2	4383	2	Q10263	pZ130 contg. Calgene	9.85e+02
c	893	12	8.2	4464	30	T67059	AUX1 gene involved in	9.85e+02
	894	12	8.2	4465	5	Q28697	Glutamine synthetase.	9.85e+02
c	895	12	8.2	4503	31	T64783	Human oxygen regulate	9.85e+02
	896	12	8.2	4691	1	Q06414	Human polypeptide cha	9.85e+02
	897	12	8.2	4695	13	Q79378	Human N-methyl-D-aspa	9.85e+02
c	898	12	8.2	4698	19	T14692	MAGE genomic DNA.	9.85e+02
c	899	12	8.2	4699	38	T49500	Transferrin receptor	9.85e+02
c	900	12	8.2	4766	10	Q70102	Malarial PfEMP3 epito	9.85e+02
	901	12	8.2	4870	29	T60301	Human SH2-containing	9.85e+02
c	902	12	8.2	4919	5	Q29260	Human calcium channel	9.85e+02
c	903	12	8.2	4937	37	T85307	Arabidopsis thaliana	9.85e+02
	904	12	8.2	4945	35	T86717	Retroviral vector con	9.85e+02
	905	12	8.2	5022	22	Q99970	Human excitatory amin	9.85e+02
c	906	12	8.2	5033	38	T49501	Transferrin receptor	9.85e+02
c	907	12	8.2	5033	21	Q94443	Bacterial transferrin	9.85e+02
c	908	12	8.2	5144	38	T49504	Transferrin receptor	9.85e+02
	909	12	8.2	5169	18	T06001	Streptomyces wedmoren	9.85e+02

	910	12	8.2	5292	19	T13392	Hybrid vector pSF3.	9.85e+02
c	911	12	8.2	5309	2	Q10280	Complete sequence of	9.85e+02
c	912	12	8.2	5345	17	T04221	5.3 kb EcoRI fragment	9.85e+02
	913	12	8.2	5476	29	T60078	Human SMCY cDNA.	9.85e+02
c	914	12	8.2	5518	34	T73870	Cotton fibre promoter	9.85e+02
	915	12	8.2	5534	2	Q13574	EcR gene.	9.85e+02
c	916	12	8.2	5655	40	V04632	Arabidopsis thaliana	9.85e+02
	917	12	8.2	5798	32	T47565	Canarypox virus DNA c	9.85e+02
	918	12	8.2	6112	28	T48868	IamS expression vecto	9.85e+02
	919	12	8.2	6202	17	Q88161	Q8 ethylene response	9.85e+02
	920	12	8.2	6314	7	Q40341	Sequence of a cDNA cl	9.85e+02
	921	12	8.2	6340	28	T48871	IamS expression vecto	9.85e+02
	922	12	8.2	6404	7	Q42388	6.5 kb BanHI fragmant	9.85e+02
	923	12	8.2	6830	37	T89255	Mouse 22B/10B (candid	9.85e+02
	924	12	8.2	6953	19	T17116	Rhodopsin gene.	9.85e+02
c	925	12	8.2	7102	2	N60044	Sequence of rhinoviru	9.85e+02
c	926	12	8.2	7102	2	N81393	Rhinovirus 2 (HRV2) g	9.85e+02
	927	12	8.2	7228	9	Q50201	Vector p37M1-10D.	9.85e+02
c	928	12	8.2	7240	35	T79668	Partial BRCA2 cancer	9.85e+02
c	929	12	8.2	7244	40	V15705	Tomato ACC synthase L	9.85e+02
	930	12	8.2	7291	24	T41476	Haemophilus adhesion	9.85e+02
c	931	12	8.2	7475	3	Q15312	Plasmid pHPB12 encodi	9.85e+02
c	932	12	8.2	7635	15	Q84654	Human neuronal calciu	9.85e+02
c	933	12	8.2	7635	15	Q84653	Human neuronal calciu	9.85e+02
c	934	12	8.2	7653	9	Q64744	ced-3 (G6536A) gene.	9.85e+02
c	935	12	8.2	7653	9	Q64737	ced-3 (G5757A) gene.	9.85e+02
c	936	12	8.2	7653	9	Q64736	ced-3 (G2487A) gene.	9.85e+02
c	937	12	8.2	7653	9	Q64738	ced-3 (C5940T) gene.	9.85e+02
c	938	12	8.2	7653	9	Q54666	ced-3 gene.	9.85e+02
c	939	12	8.2	7653	9	Q64745	ced-3 (C7020T) gene.	9.85e+02
	940	12	8.2	7815	35	T90542	Maize male fertility	9.85e+02
	941	12	8.2	7824	28	T63073	7.8 kb fragment of pN	9.85e+02
	942	12	8.2	8316	12	Q74450	myoD retroviral expre	9.85e+02
	943	12	8.2	8323	10	Q58700	LTR clone of PB29 iso	9.85e+02
c	944	12	8.2	8457	11	Q66797	AmEPV spheroidin gene	9.85e+02
c	945	12	8.2	8654	14	Q55138	Staphylococcus epider	9.85e+02
	946	12	8.2	8700	7	Q42541	BglII/HpaII fragment	9.85e+02
c	947	12	8.2	8912	18	T00040	Hepatitis GB virus (H	9.85e+02
c	948	12	8.2	8996	31	T68754	ATM mutant 4437del175	9.85e+02
c	949	12	8.2	8996	31	T68783	ATM mutant 4443del175	9.85e+02
c	950	12	8.2	9006	31	T68753	ATM mutant 4612del165	9.85e+02
c	951	12	8.2	9032	31	T68739	ATM mutant 7789del139	9.85e+02
c	952	12	8.2	9046	16	T02792	Transcription factor	9.85e+02
c	953	12	8.2	9048	25	T43225	Brassica napus FCA ge	9.85e+02
c	954	12	8.2	9060	40	V15702	Zucchini ACC synthase	9.85e+02
c	955	12	8.2	9081	31	T68759	ATM mutant 2377del90	9.85e+02
c	956	12	8.2	9143	18	T00130	Hepatitis GB virus (H	9.85e+02
c	957	12	8.2	9165	31	T68785	ATM mutant 7278del6 c	9.85e+02
c	958	12	8.2	9166	31	T68743	ATM mutant 6573del5 c	9.85e+02
c	959	12	8.2	9167	31	T68769	ATM mutant 7240del4 c	9.85e+02
c	960	12	8.2	9168	31	T68780	ATM mutant 5435del3 c	9.85e+02

c	961	12	8.2	9169	31	T68763	ATM mutant 1563delAG 9.85e+02
c	962	12	8.2	9171	31	T68764	ATM mutant C1339T cod 9.85e+02
c	963	12	8.2	9171	31	T68781	ATM mutant A8711G cod 9.85e+02
c	964	12	8.2	9171	31	T68737	ATM mutant C8140T cod 9.85e+02
c	965	12	8.2	9171	31	T68733	ATM mutant G8307A cod 9.85e+02
c	966	12	8.2	9171	37	T43497	ATM open reading fram 9.85e+02
c	967	12	8.2	9172	31	T68732	ATM mutant 8946insA c 9.85e+02
c	968	12	8.2	9172	31	T68747	ATM mutant 5712insA c 9.85e+02
c	969	12	8.2	9385	31	T72683	Cell cycle checkpoint 9.85e+02
c	970	12	8.2	9739	27	T58549	Human Immunodeficienc 9.85e+02
c	971	12	8.2	9746	3	Q14752	HIV-1(MN-ST1) env pro 9.85e+02
	972	12	8.2	10366	40	V05828	Human huntingtin codi 9.85e+02
	973	12	8.2	10366	36	T97924	Composite sequence of 9.85e+02
c	974	12	8.2	10952	23	T33345	Human CAPL gene. 9.85e+02
	975	12	8.2	10970	12	Q73735	Retro virus vector p5 9.85e+02
c	976	12	8.2	11225	2	N91695	Sequence of fowlpox v 9.85e+02
c	977	12	8.2	11283	35	T84841	Human breast and ovar 9.85e+02
c	978	12	8.2	11288	15	Q90512	CEA clone HindIII-Sau 9.85e+02
	979	12	8.2	11478	33	T79805	Full length potato st 9.85e+02
	980	12	8.2	11478	33	T79784	Full length potato st 9.85e+02
	981	12	8.2	11817	35	T74199	Murine Lyst1 long iso 9.85e+02
	982	12	8.2	12145	30	T38833	pHK28-26 containing d 9.85e+02
c	983	12	8.2	12151	11	Q62137	Rice starch branching 9.85e+02
	984	12	8.2	12283	2	Q10353	Nucleotide sequence o 9.85e+02
c	985	12	8.2	12720	10	Q64211	L.lactis branched ami 9.85e+02
c	986	12	8.2	12778	31	T68784	ATM mutant G9170C cod 9.85e+02
	987	12	8.2	14507	32	T73568	Expression augmenting 9.85e+02
c	988	12	8.2	14556	15	Q90511	CEA genomic clone. 9.85e+02
	989	12	8.2	15664	34	T70492	Mycobacteriophage DS6 9.85e+02
	990	12	8.2	15664	29	T66130	Mycobacteriophage DS6 9.85e+02
	991	12	8.2	16382	28	T63074	IgG-Fc binding protei 9.85e+02
c	992	12	8.2	20710	20	T16806	Kaposi's sarcoma asso 9.85e+02
c	993	12	8.2	24593	3	N50226	Sequence of opine syn 9.85e+02
	994	12	8.2	24596	3	N50182	Complete nucleotide s 9.85e+02
c	995	12	8.2	29392	40	V15422	Mouse poly Ig recepto 9.85e+02
c	996	12	8.2	40352	40	V02032	MAGE-B cluster DNA se 9.85e+02
c	997	12	8.2	53526	40	T94101	Human PKD1 gene. 9.85e+02
c	998	12	8.2	53577	40	T94108	Human PKD1 locus betw 9.85e+02
c	999	12	8.2	53577	28	T18551	Human polycystic kidn 9.85e+02
c	1000	12	8.2	580073	27	T58840	Mycoplasma genitalium 9.85e+02

ALIGNMENTS

```
RESULT    1
ID   T19709 standard; cDNA to mRNA; 147 BP.
AC   T19709;
DT   05-JUL-1996  (first entry)
DE   Human gene signature HUMGS00783.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
```



```
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K, Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 454; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   147 BP;    39 A;    39 C;    30 G;    39 T;

  Query Match          100.0%;  Score 147;  DB 18;  Length 147;
  Best Local Similarity 100.0%;  Pred. No. 1.55e-114;
  Matches  147;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 gatccccagggtttctttgtcttatttatgagagaaaaaccggtcacttttgtccagcgcac 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      1 GATCCCCAGGGTTTCTTTGTCTTATTTATGGAGAAAAACCGGTCACTTTGTCCAGCGCAC 60

Db     61 tgtgagggccccccactcaggccagccctggccccccccttggtacttggaaccgaagttaca 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     61 TGTGAGGCCCCCCACTCAGGCCAGCCCTGGCCCCCCCCTTGGTACTTGGAACCGAAGTTACA 120

Db    121 gattatattaaaaataataatgtacaaa 147
          |||||||||||||||||||||||||||||
Qy    121 GATTATATTAAAAATAATAATGTACAAA 147


RESULT    2
ID   Q31693 standard; DNA; 731 BP.
```



```
AC   Q31693;
DT   15-APR-1993  (first entry)
DE   RsaI restriction fragment from Dd ras genomic clone.
KW   RsaI; restriction fragment; Dd ras; SmaI; pGEM1; promoter; vector;
KW   pAV-CAT; pEDII-CS; CS; leader peptide; CAT; Discoidin I; promoter;
KW   chloramphenicol acetyl transferase; CAT; geneticin; G418; resistance;
KW   actin 15; terminator; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   promoter        182..366
FT                   /*tag=  a
PN   WO9220806-A.
PD   26-NOV-1992.
PF   18-MAY-1992; E01112.
PR   17-MAY-1991; NL-000869.
PA   (RMFD-) RMF DICTAGENE SA.
PI   Fasel NJ, Reymond CD;
DR   WPI; 92-415784/50.
PT   Dictyostelium expression vector - used to produce Plasmodium
PT   circumsporozoite protein for malaria vaccines
PS   Disclosure; Fig 4; 20pp; English.
CC   The sequence given represents the RsaI restriction fragment from the
CC   genomic clone of Dd ras.  This sequence was inserted into the SmaI
CC   site of pGEM1 and sequenced.  Parts of this promoter sequence were
CC   placed in the vector pAV-CAT to generate a plasmid which resembled
CC   pEDII-CS (see also Q31692), with CS and leader peptides replaced by
CC   the chloramphenicol acetyl transferase gene (CAT). Discoidin I promoter
CC   by Dd ras promoter and the geneticin (G418) resistance gene under the
CC   control of the actin 15 promoter.  The terminator region was placed in
CC   the opposite orientation.
SQ   Sequence   731 BP;    317 A;    52 C;    42 G;    320 T;

  Query Match          12.2%;  Score 18;  DB 5;  Length 731;
  Best Local Similarity 100.0%;  Pred. No. 2.37e-01;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    626 tttgtacattattatttt 643
          ||||||||||||||||||
Cp    147 TTTGTACATTATTATTTT 130


RESULT    3
ID   Q61162 standard; DNA; 191 BP.
AC   Q61162;
DT   16-MAR-1994  (first entry)
DE   Human brain Expressed Sequence Tag EST01204.
KW   Gene transcription product; genetic markers; tagging; in vivo;
KW   transcription; mapping; locations; chromosomes; chromosomal; ss.
OS   Homo sapiens.
PN   WO9316178-A.
PD   19-AUG-1993.
```



```
PF   12-FEB-1993; U01294.
PR   12-FEB-1992; US-837195.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
PI   Adams MD, Moreno RF, Venter CJ;
DR   WPI; 93-272882/34.
PT   Enriched oligonucleotides and corresp. sequences - used as
PT   markers for human genes transcribed in-vivo, facilitate tagging
PT   of most human genes
PS   Example 4; Page 447; 500pp; English.
CC   The Expressed Sequence Tag was isolated from a human brain cDNA
CC   library as part of a large set of ESTs which can be used as markers
CC   for human genes transcribed in vivo. They can be used to facilitate
CC   tagging of most human genes, for mapping locations of expressed genes
CC   on chromosomes, for individual or forensic identification, for mapping
CC   locations of disease-associated genes, for identification of tissue
CC   type, and for prepn. of antisense sequences, probes and constructs.
CC   EST01204 has an "poor" coding probability as evaluated using the
CC   coding-region prediction program CRM. See also Q59041-Q61440.
SQ   Sequence   191 BP;    35 A;    40 C;    87 G;    27 T;

  Query Match          11.6%;  Score 17;  DB 8;  Length 191;
  Best Local Similarity 94.4%;  Pred. No. 1.07e+00;
  Matches   17;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db    127 actnaggccagccctggc 144
          ||| |||||||||||||||
Qy     73 ACTCAGGCCAGCCCTGGC 90


RESULT    4
ID   Q12532 standard; cDNA; 789 BP.
AC   Q12532;
DT   08-OCT-1991  (first entry)
DE   cDNA clone encoding human O6-methylguanine-DNA-methyltransferase.
KW   human O6-methylguanine-DNA-methyltransferase; Mer; repair; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             41..664
FT                   /*tag=  a
PN   J03147789-A.
PD   24-JUN-1991.
PF   02-NOV-1989; 287223.
PR   02-NOV-1989; JP-287223.
PA   (GIJU-) SHIN GIJUTSU JIGYO.
DR   WPI; 91-227679/31.
DR   P-PSDB; R12931.
PT   DNA clone for coding human O6-methyl:guanine-DNA-methyl:transferase
PT   - involves preparing cDNA library from human normal cell Mer plus
PT   cell strain then recovering cDNA from alkylation resistant trans-
PT   formant.
PS   Claim 1; Page 10; 10pp; Japanese.
```

```
CC   The cDNA clone encoding human O6-methylguanine-DNA-methyltransferase,
CC   (a specific repair enzyme) can be obtd. by first introducing a cDNA
CC   library from a normal Mer+ cell strain to a Mer- cell strain.  Then
CC   selecting a transformant resistant to an alkylating agent and
CC   recovering cDNA from the transformant.
SQ   Sequence   789 BP;    182 A;    216 C;    249 G;    142 T;

  Query Match           11.6%;  Score 17;  DB 2;  Length 789;
  Best Local Similarity 100.0%;  Pred. No. 1.07e+00;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     22 cccctttggtacttggaa 38
          ||||||||||||||||||
Qy     93 CCCCTTGGTACTTGGAA 109


RESULT    5
ID   T61680 standard; DNA; 1654 BP.
AC   T61680;
DT   19-JUN-1997  (first entry)
DE   Human alpha(1,3)-fucosyltransferase (Fuc-TVI) DNA.
KW   Alpha(1,3/1,4)-fucosyltransferase; Lewis enzyme; Fuc-TVI;
KW   glycosylation; oligosaccharide; blood group; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             129..1208
FT                   /*tag=  a
PN   WO9709421-A1.
PD   13-MAR-1997.
PF   06-SEP-1996; U13816.
PR   08-SEP-1995; US-525058.
PA   (UNMI ) UNIV MICHIGAN.
PI   Legault DJ,  Lowe JB;
DR   WPI; 97-192897/17.
DR   P-PSDB; W13643.
PT   New recombinant fucosyltransferase proteins - useful for modifying
PT   cell surface oligosaccharide structures
PS   Example 6; Page 291-292; 329pp; English.
CC   A DNA clone (T61680) codes for an enzyme (W13643) that functions as
CC   a GDP-Fuc:beta-D-Gal(1,4)-D-GlcNAc alpha(1,3)-fucosyltransferase
CC   (Fuc-TVI) which can utilise neutral type II oligosaccharide
CC   precursors but not type I glycoconjugates.  It was identified in a
CC   human lambda phage genomic DNA library using a probe obtd. from
CC   Lewis enzyme cDNA (see also T61675).  The Fuc-TVI DNA can be used
CC   to construct animal cell lines with specific capabilities with
CC   respect to post-translational modification of the oligosaccharides
CC   of expressed proteins or lipids, to produce recombinant Fuc-TVI for
CC   use in oligosaccharide prodn., to genotype individuals at this
CC   fucosyltransferase locus, and in the prodn. of chimeric
CC   fucosyltransferases (see also W14512-32).
SQ   Sequence   1654 BP;    298 A;    557 C;    459 G;    340 T;
```



```
  Query Match           11.6%;  Score 17;  DB 29;  Length 1654;
  Best Local Similarity 100.0%;  Pred. No. 1.07e+00;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1261 cctgagtgggggcctca 1277
          |||||||||||||||||
Cp     79 CCTGAGTGGGGGCCTCA 63


RESULT    6
ID   Q56905 standard; DNA; 1654 BP.
AC   Q56905;
DT   26-JUL-1994  (first entry)
DE   pCDNA1-alpha-(1-3)Fuc-TV1 genomic DNA.
KW   Glycosyltransferase; fucosyltransferase; GDP-Fuc; in vitro; cell;
KW   surface; oligosaccharide; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             128..1208
FT                   /*tag=  a
PN   WO9402616-A.
PD   03-FEB-1994.
PF   20-JUL-1993; U06703.
PR   20-JUL-1992; US-916281.
PA   (UNMI ) UNIV MICHIGAN.
PI   Lowe JB;
DR   WPI; 94-048874/06.
DR   P-PSDB; R45933.
PT   DNA fragment encoding a glycosyltransferase - can be used for in
PT   vitro reactions to modify cell surface oligosaccharide(s) e.g.
PT   blood gp. determinants, to protect against transplant rejection
PS   Claim 3; Fig 7; 249pp; English.
CC   The sequence is that encoding human GDP-Fuc: [beta-D-Gal (1-4)]-D-
CC   GlcNac alpha (1-3)-fucosyltransferase.  The enzyme produced by the
CC   DNA may be non glycosylated.  This prevents premature loss of enzyme
CC   activity.  It can also be used in in vitro reactions to modify cell
CC   surface oligosaccharide mols. e.g. blood group determinants.
CC   See also Q56906-12.
SQ   Sequence   1654 BP;    298 A;    557 C;    459 G;    340 T;

  Query Match           11.6%;  Score 17;  DB 9;  Length 1654;
  Best Local Similarity 100.0%;  Pred. No. 1.07e+00;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1261 cctgagtgggggcctca 1277
          |||||||||||||||||
Cp     79 CCTGAGTGGGGGCCTCA 63


RESULT    7
```



```
ID   Q53501 standard; DNA; 2973 BP.
AC   Q53501;
DT   22-JUN-1994  (first entry)
DE   Sequence of bean malic enzyme gene promoter together with leader
DE   sequence and transcription and translation initiation regions down
DE   to the first intron of the coding region.
KW   Malic enzyme gene promoter; clone gPv ME1 1-4.1 HH (NCIMB 40498); ss.
OS   Phaseolus vulgaris var. Processor.
FH   Key             Location/Qualifiers
FT   promoter        1..2638
FT                   /*tag=  a
FT                   /note= "bases 1 to approx. 2640 are claimed"
PN   WO9324639-A.
PD   09-DEC-1993.
PF   27-MAY-1993; G01099.
PR   29-MAY-1992; GB-011440.
PA   (ZENE ) ZENECA LTD.
PI   Schaaf J,  Walter MH;
DR   WPI; 93-405830/50.
PT   DNA construct for transforming plant cells - contains malic
PT   enzyme promoter used for inducible exogenous gene expression
PS   Claim 3; Figure 2; 23pp; English.
CC   A library was constructed from bean genomic DNA  and screened with a
CC   pME1 cDNA fragment of 1.4kb. One positive clone was isolated and
CC   characterised - gPvME1. The promoter sequence of the malic enzyme
CC   gene is given in Q53501. gPvME1 1-4.1HH, a plasmid subclone of
CC   gPVME1 comprising the promoter and part of the coding region, was
CC   used to isolate a 2821 bp HindIII-BclII fragment contg. approx. 2.65
CC   kb of the PvME1 promoter and leader sequences down to the BclI site
CC   immediately upstream of the ATG start codon. This fragment was
CC   inserted into the binary plant transformation vector pBI 101.4
CC   between the HindIII and the BamHI site of the polylinker, which has
CC   an overhang compatibility with BclI cut DNA. This construct was
CC   named pPvME1 Prom-GUS1. The promoter sequence of a malic enzyme
CC   gene contained within the insert in clone gPvME1 1-4.1HH which
CC   has been deposited in Aberdeen, Scotland (NCIMB 40498) has been
CC   claimed.
SQ   Sequence   2973 BP;    1077 A;    413 C;    424 G;    1059 T;

  Query Match           11.6%;  Score 17;  DB 9;  Length 2973;
  Best Local Similarity 100.0%;  Pred. No. 1.07e+00;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1694 attattattttaatata 1710
          |||||||||||||||||
Cp    140 ATTATTATTTTAATATA 124


RESULT    8
ID   Q43661 standard; DNA; 4558 BP.
AC   Q43661;
```



```
DT   13-OCT-1993  (first entry)
DE   Acetobacter cdg2 operon.
KW   Cyclic diguanylate; diguanylate phosphodiesterase;
KW   diguanylate cyclase; cellulose production; ss.
OS   Acetobacter xylinum.
FH   Key             Location/Qualifiers
FT   cds             98..2354
FT                   /*tag=  a
FT                   /label= pdeA2
FT                   /note= "a diguanylate phosphodiesterase gene;
FT                   the amino acid sequence R38153 cannot be
FT                   deduced directly from this open reading
FT                   frame due to a number of apparent
FT                   (typographical ?) insertions and deletions
FT                   in the printed sequence"
FT   misc_difference 326..327
FT                   /*tag=  b
FT                   /codon= seq: GG; aa: Val
FT                   /note= "inserting a T between nucleotides 326 and
FT                   327 will restore the reading frame"
FT   misc_difference 730..732
FT                   /*tag=  c
FT                   /codon= seq: TAG; aa: Met
FT   misc_difference 739..740
FT                   /*tag=  d
FT                   /codon= seq: GC; aa: Arg
FT                   /note= "inserting a C between nucleotides 738 and
FT                   739 will restore the reading frame"
FT   misc_difference 789..791
FT                   /*tag=  e
FT                   /codon= seq: ATC; aa: Met
FT   misc_difference 792..794
FT                   /*tag=  f
FT                   /codon= seq: GTG; aa: Leu
FT   misc_difference 834..835
FT                   /*tag=  g
FT                   /codon= seq: GC; aa: Gly
FT                   /note= "inserting a G between nucleotides 834 and
FT                   835 will restore the reading frame"
FT   misc_difference 908
FT                   /*tag=  h
FT                   /codon= seq:CAGG; aa: Gln
FT                   /note= "delete G(908) to maintain reading frame"
FT   misc_difference 918..921
FT                   /*tag=  i
FT                   /codon= seq: GCTG; aa: Val
FT                   /note= "delete C(919) to maintain reading frame"
FT   misc_difference 1258..1260
FT                   /*tag=  j
FT                   /codon= seq: GAA; aa: Asp
FT   misc_difference 1720..1721
```

```
FT                   /*tag=  k
FT                   /codon= seq: CC; aa: Ala
FT                   /note= "inserting a G between nucleotides 1719 and
FT                   1720 will restore reading frame"
FT   cds             2463..4187
FT                   /*tag=  l
FT                   /label= dgc2
FT                   /note= "a diguanylate cyclase gene"
PN   WO9311244-A.
PD   10-JUN-1993.
PF   14-OCT-1992; UO8756.
PR   29-NOV-1991; US-800218.
PA   (WEYE ) WEYERHAEUSER CO.
PI   Ben-Bassat A,  Benziman M,  Calhoon RD,  Gelfand DH;
PI   Tal R,  Wong HC;
DR   WPI; 93-197062/24.
DR   P-PSDB; R38153, R38154.
PT   Polynucleotide sequence from Acetobacter cdg operon - encodes
PT   cyclic di:guanosine mono:phosphate degradation enzymes e.g.
PT   3-phosphodiesterase isozyme
PS   Claim 3; Page 67-70; 98pp; English.
CC   A gene bank of Acetobacter 1306-11 was constructed in the broad host
CC   range, mobilisable cosmid pKT230C085. The average insert size was ca.
CC   30kb. The bank was screened with a 53bp PCR amplified probe whose
CC   sequence was expected to be an exact match for a diguanylate
CC   cyclase encoding gene. Three unique cosmid classes were identified
CC   corresponding to three different cdg operons. The cdg2 operon
CC   comprises at least 2 genes. Manipulating the expression levels of
CC   the different cdg2 genes is contemplated, particularly in cellulose
CC   producing bacteria.
SQ   Sequence   4558 BP;    850 A;    1379 C;    1467 G;    862 T;

  Query Match           11.6%;  Score 17;  DB 7;  Length 4558;
  Best Local Similarity 100.0%;  Pred. No. 1.07e+00;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   4106 ccagggtttctttgtct 4122
          |||||||||||||||||
Qy      6 CCAGGGTTTCTTTGTCT 22


RESULT    9
ID   V17547 standard; DNA; 608 BP.
AC   V17547;
DT   22-JUN-1998  (first entry)
DE   Peptidyl prolyl cis-trans isomerase coding sequence.
KW   Peptidyl prolyl cis-trans isomerase; PPIase; immunosuppressor; ss.
OS   Methanococcus thermolithotrophicus.
FH   Key             Location/Qualifiers
FT   CDS             82..546
FT                   /*tag=  a
```



```
PN   J10000091-A.
PD   06-JAN-1998.
PF   13-JUN-1996; 152546.
PR   13-JUN-1996; JP-152546.
PA   (KAIY-) KAIYO BIOTECHNOLOGY KENKYUSHO KK.
DR   WPI; 98-113594/11.
DR   P-PSDB; W41489.
PT   Peptidyl prolyl cis-trans isomerase - useful for, e.g. regeneration
PT   of denatured proteins
PS   Claim 1; Page 5; 8pp; Japanese.
CC   This sequence encodes the peptidyl prolyl cis-trans isomerase (PPIase)
CC   of the invention. PPIase may be used for, e.g. regeneration of denatured
CC   proteins, stabilisation of protein reagents, production of recombinant
CC   proteins, and detection of new immunosuppressors and physiologically
CC   active compounds.
SQ   Sequence   608 BP;    233 A;    80 C;    123 G;    172 T;

  Query Match           10.9%;  Score 16;  DB 40;  Length 608;
  Best Local Similarity 100.0%;  Pred. No. 4.67e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   447 aaccgaagttacagat 462
         ||||||||||||||||
Qy   108 AACCGAAGTTACAGAT 123


RESULT   10
ID   Q47167 standard; DNA; 1517 BP.
AC   Q47167;
DT   21-JAN-1994  (first entry)
DE   HPV1 L1 ORF.
KW   L1; capsid protein; bovine; human; papillomavirus; baculovirus;
KW   transfer vector; promoter; capsid protein; transformation; vaccine;
KW   neutralising antibody; vertebrate; ss.
OS   Human papillomavirus.
PN   US8032869-A.
PD   15-JUL-1993.
PF   03-SEP-1992; 941371.
PR   03-SEP-1992; US-941371.
PR   16-MAR-1993; US-032869.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
PI   Kirnbauer R,  Lowy DR,  Schiller JT;
DR   WPI; 93-249995/31.
DR   P-PSDB; R38808.
PT   Recombinant papilloma virus capsid proteins - for vaccines
PT   against papilloma virus and for diagnosis of virus infection
PS   Example 1; Page 35-37; 45pp; English.
CC   The sequences given in Q47166-67 encode the L1 capsid proteins from
CC   bovine and human papillomavirus respectively.  These sequences may be
CC   inserted into a baculovirus transfer vector and operatively
CC   expressed by a promoter of the vector, and the capsid protein
```



```
CC   produced by transformed cells.  The capsid protiens may be used in
CC   vaccines to induce high-titre neutralising antibody response in
CC   vertebrates.
SQ   Sequence   1517 BP;    487 A;    291 C;    286 G;    453 T;

  Query Match           10.9%;  Score 16;  DB 7;  Length 1517;
  Best Local Similarity 100.0%;  Pred. No. 4.67e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   695 cagattatattaaaat 710
         ||||||||||||||||
Qy   119 CAGATTATATTAAAAT 134


RESULT   11
ID   V12162 standard; DNA; 1517 BP.
AC   V12162;
DT   05-MAY-1998  (first entry)
DE   Human papillomavirus prototype L1 gene.
KW   Human; papillomavirus; bovine; L1 conformational protein; antibody;
KW   detection; diagnosis; ss.
OS   Human papillomavirus.
FH   Key             Location/Qualifiers
FT   CDS             1..1517
FT                   /*tag=  a
FT                   /product= "L1"
FT                   /note= "no stop codon given"
PN   US5709996-A.
PD   20-JAN-1998.
PF   07-JUN-1995; 472673.
PR   16-MAR-1993; US-032869.
PR   03-SEP-1992; US-941371.
PR   07-JUN-1995; US-472673.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI   Kirnbauer R, Lowy DR, Schiller JT;
DR   WPI; 98-109811/10.
DR   P-PSDB; W44143.
PT   Assay for papilloma virus - using antibody directed against
PT   recombinant L1 polypeptide
PS   Disclosure; Column 17-20; 34pp; English.
CC   An method has been developed for detecting papilloma virus in a specimen
CC   from a mammal. The method comprises contacting the sample with an
CC   antibody and detecting any antibody binding. The antibody is a
CC   polyclonal or monoclonal antibody produced by transforming a eukaryotic
CC   host cell with a genetic construct containing a papillomavirus L1 gene
CC   so that the cell expresses self-assembled papillomavirus-like particles
CC   containing at least 1 papillomavirus L1 conformational epitope. The
CC   present sequence encodes prototype human papillomavirus L1 polypeptide.
CC   The method is for the diagnosis of papillomavirus infections in mammals.
SQ   Sequence   1517 BP;    487 A;    293 C;    284 G;    453 T;
```



```
  Query Match           10.9%;  Score 16;  DB 38;  Length 1517;
  Best Local Similarity 100.0%;  Pred. No. 4.67e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   695 cagattatattaaaat 710
         ||||||||||||||||
Qy   119 CAGATTATATTAAAAT 134


RESULT   12
ID   Q47166 standard; DNA; 1517 BP.
AC   Q47166;
DT   21-JAN-1994  (first entry)
DE   BPV1 L1 ORF.
KW   L1; capsid protein; bovine; human; papillomavirus; baculovirus;
KW   transfer vector; promoter; capsid protein; transformation; vaccine;
KW   neutralising antibody; vertebrate; ss.
OS   Bovine papillomavirus.
PN   US8032869-A.
PD   15-JUL-1993.
PF   03-SEP-1992; 941371.
PR   03-SEP-1992; US-941371.
PR   16-MAR-1993; US-032869.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
PI   Kirnbauer R,  Lowy DR,  Schiller JT;
DR   WPI; 93-249995/31.
DR   P-PSDB; R38807.
PT   Recombinant papilloma virus capsid proteins - for vaccines
PT   against papilloma virus and for diagnosis of virus infection
PS   Example 1; Page 32-34; 45pp; English.
CC   The sequences given in Q47166-67 encode the L1 capsid proteins from
CC   bovine and human papillomavirus respectively.  These sequences may be
CC   inserted into a baculovirus transfer vector and operatively
CC   expressed by a promoter of the vector, and the capsid protein
CC   produced by transformed cells.  The capsid protiens may be used in
CC   vaccines to induce high-titre neutralising antibody response in
CC   vertebrates.
SQ   Sequence   1517 BP;    487 A;    293 C;    284 G;    453 T;

  Query Match           10.9%;  Score 16;  DB 7;  Length 1517;
  Best Local Similarity 100.0%;  Pred. No. 4.67e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   695 cagattatattaaaat 710
         ||||||||||||||||
Qy   119 CAGATTATATTAAAAT 134


RESULT   13
ID   V09855 standard; DNA; 1517 BP.
AC   V09855;
```

```
DT   26-MAY-1998  (first entry)
DE   HPV16 L1 capsid protein DNA.
KW   Capsid protein; L1; HPV16; vaccine; prevention; treatment;
KW   self-assembly; viral protein; capsomer; capsid; antigenic epitope; ss.
OS   Human papillomavirus.
FH   Key             Location/Qualifiers
FT   CDS             1..1517
FT                   /*tag= a
FT                   /product= L1
FT                   /note= "partial capsid protein sequence"
PN   US5716620-A.
PD   10-FEB-1998.
PF   07-JUN-1995; 475783.
PR   16-MAR-1993; US-032869.
PR   03-SEP-1992; US-941371.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI   Kirnbauer R, Lowy DR, Schiller JT;
DR   WPI; 98-158363/14.
DR   P-PSDB; W39903.
PT   Vaccine against human papilloma virus - comprises HPV16 L1
PT   polypeptide
PS   Disclosure; Column 17-20; 20pp; English.
CC   This sequence encodes the L1 capsid protein from Human Papillomavirus
CC   strain 16 (HPV16). A recombinant form of this viral protein which is
CC   capable of self-assembly into capsomer structures and viral capsids that
CC   comprise conformational antigenic epitopes can be used as a vaccine for
CC   the prevention or treatment of papillomavirus infections in vertebrates.
CC   The vaccine comprises a unit dose of a composition containing a self
CC   assembled HPV16 with at least 1 papillomavirus L1 conformational epitope.
SQ   Sequence   1517 BP;    487 A;    293 C;    284 G;    453 T;

  Query Match          10.9%;  Score 16;  DB 39;  Length 1517;
  Best Local Similarity 100.0%;  Pred. No. 4.67e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    695 cagattatattaaaat 710
          ||||||||||||||||
Qy    119 CAGATTATATTAAAAT 134


RESULT   14
ID   V09856 standard; DNA; 1517 BP.
AC   V09856;
DT   26-MAY-1998  (first entry)
DE   HPV16 recombinant L1 capsid protein DNA.
KW   Capsid protein; L1; HPV16; vaccine; prevention; treatment;
KW   self-assembly; viral protein; capsomer; capsid; antigenic epitope; ss.
OS   Human papillomavirus.
FH   Key             Location/Qualifiers
FT   CDS             1..1517
FT                   /*tag= a
```



```
FT                   /product= L1
FT                   /note= "partial recombinant capsid protein sequence"
PN   US5716620-A.
PD   10-FEB-1998.
PF   07-JUN-1995; 475783.
PR   16-MAR-1993; US-032869.
PR   03-SEP-1992; US-941371.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI   Kirnbauer R, Lowy DR, Schiller JT;
DR   WPI; 98-158363/14.
DR   P-PSDB; W39904.
PT   Vaccine against human papilloma virus - comprises HPV16 L1
PT   polypeptide
PS   Claim 2; Column 21-24; 20pp; English.
CC   This sequence encodes a recombinant L1 capsid protein from Human
CC   Papillomavirus strain 16 (HPV16). This recombinant form of viral
CC   protein is capable of self-assembly into capsomer structures and viral
CC   capsids that comprise conformational antigenic epitopes can be used as a
CC   vaccine for the prevention or treatment of papillomavirus infections in
CC   vertebrates. The vaccine comprises a unit dose of a composition
CC   containing a self assembled HPV16 with at least 1 papillomavirus L1
CC   conformational epitope.
SQ   Sequence   1517 BP;    487 A;    291 C;    286 G;    453 T;

  Query Match          10.9%;  Score 16;  DB 39;  Length 1517;
  Best Local Similarity 100.0%;  Pred. No. 4.67e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    695 cagattatattaaaat 710
          ||||||||||||||||
Qy    119 CAGATTATATTAAAAT 134


RESULT   15
ID   V12161 standard; DNA; 1517 BP.
AC   V12161;
DT   05-MAY-1998  (first entry)
DE   Human papillomavirus wild-type L1 gene.
KW   Human; papillomavirus; bovine; L1 conformational protein; antibody;
KW   detection; diagnosis; ss.
OS   Human papillomavirus.
FH   Key             Location/Qualifiers
FT   CDS             1..1517
FT                   /*tag= a
FT                   /product= "L1"
FT                   /note= "no stop codon given"
PN   US5709996-A.
PD   20-JAN-1998.
PF   07-JUN-1995; 472673.
PR   16-MAR-1993; US-032869.
PR   03-SEP-1992; US-941371.
```



```
PR   07-JUN-1995; US-472673.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI   Kirnbauer R, Lowy DR, Schiller JT;
DR   WPI; 98-109811/10.
DR   P-PSDB; W44142.
PT   Assay for papilloma virus - using antibody directed against
PT   recombinant L1 polypeptide
PS   Claim 1; Column 21-24; 34pp; English.
CC   An method has been developed for detecting papilloma virus in a specimen
CC   from a mammal. The method comprises contacting the sample with an
CC   antibody and detecting any antibody binding. The antibody is a
CC   polyclonal or monoclonal antibody produced by transforming a eukaryotic
CC   host cell with a genetic construct containing a papillomavirus L1 gene
CC   so that the cell expresses self-assembled papillomavirus-like particles
CC   containing at least 1 papillomavirus L1 conformational epitope. The
CC   present sequence encodes wild-type human papillomavirus L1 polypeptide.
CC   The method is for the diagnosis of papillomavirus infections in mammals.
SQ   Sequence   1517 BP;    487 A;    291 C;    286 G;    453 T;

  Query Match          10.9%;  Score 16;  DB 38;  Length 1517;
  Best Local Similarity 100.0%;  Pred. No. 4.67e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    695 cagattatattaaaat 710
          ||||||||||||||||
Qy    119 CAGATTATATTAAAAT 134


RESULT   16
ID   Q98697 standard; DNA; 2526 BP.
AC   Q98697;
DT   17-JAN-1996  (first entry)
DE   Yeast aspartyl protease 3 gene.
KW   Serum albumin; HSA; aspartyl protease-3; YAP3 gene; ss.
OS   Saccharomyces cerevisiae.
FH   Key             Location/Qualifiers
FT   cds             541..2250
FT                   /*tag= a
PN   WO9523857-A1.
PD   08-SEP-1995.
PF   01-MAR-1995; G00434.
PR   05-MAR-1994; GB-004270.
PA   (DELS ) DELTA BIOTECHNOLOGY LTD.
PI   Gilbert SC, Kerry-Williams SM;
DR   WPI; 95-320572/41.
PT   Yeast with reduced levels of aspartyl protease 3 proteolytic
PT   activity - used to secrete human albumin without prodn. of the 45
PT   kD fragment
PS   Example 2; Page 37-38; 50pp; English.
CC   The YAP3 gene encoding yeast aspartyl protease 3 was mutated by
CC   gene disruption.  Primer pairs given in Q98687-88 and Q98689-90
```



```
CC   were used to amplify the 5' and 3' ends, respectively, of the YAP3
CC   gene.  Amplified products were used to construct yeast strains
CC   lacking Yap3p activity, allowing improved prodn. of recombinant HSA.
SQ   Sequence   2526 BP;    693 A;    546 C;    495 G;    792 T;

  Query Match          10.9%;  Score 16;  DB 15;  Length 2526;
  Best Local Similarity 100.0%;  Pred. No. 4.67e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    330 cttatttatggagaaa 345
          ||||||||||||||||
Qy     21 CTTATTTATGGAGAAA 36


RESULT   17
ID   T72814 standard; cDNA; 3817 BP.
AC   T72814;
DT   08-NOV-1997  (first entry)
DE   Human SPT5 cDNA.
KW   SPT5; multiple function regulatory protein; MFRP; cell cycle;
KW   gastrulation; carcinogenesis; tumorigenicity; breast cancer;
KW   tumour suppressor; neoplasm; cardiogenesis; organogenesis;
KW   gonad development; gene therapy; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             265..3531
FT                   /*tag= a
PN   WO9703087-A1.
PD   30-JAN-1997.
PF   12-JUL-1996; U11702.
PR   13-JUL-1995; US-001113.
PA   (UNMI ) UNIV MICHIGAN.
PI   Chiang P,  Kurnit DM,  Song W;
DR   WPI; 97-118981/11.
DR   P-PSDB; W19786.
PT   New family of multiple function regulatory proteins - regulate cell
PT   cycle progression, gastrulation, carcinogenesis, tumorigenicity,
PT   gonad development, etc.
PS   Claim 3; Fig 14a; 98pp; English.
CC   This cDNA sequence codes for human SPT5 (W19784), a novel multiple
CC   function regulatory protein (MFRP) that shows homology to yeast
CC   SPT5.  It was isolated from a human lung cDNA library using a human
CC   expressed sequence tag that showed homology to yeast SPT5.  The
CC   human SPT5 gene maps close to the ryanodine receptor at 19q13.
CC   SPT5 polynucleotides can be used to produce SPT5 polypeptides in
CC   host cells, or introduced in vitro, ex vivo or in vivo into a cell
CC   to regulate a cellular function associated with MFRP activity, such
CC   as cell cycle progression, control of gastrulation, carcinogenesis
CC   (such as breast cancer), tumorigenicity, inhibition or reversal of
CC   a neoplastic phenotype, progression through the cell cycle, gonad
CC   development, cardiogenesis or organogenesis.  They can also be used
```

```
CC   to restore wild-type MFRP function, such as tumour suppression, and
CC   as probes.
SQ   Sequence   3817 BP;    914 A;    1048 C;    1150 G;    705 T;

  Query Match          10.9%;  Score 16;  DB 33;  Length 3817;
  Best Local Similarity 100.0%;  Pred. No. 4.67e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2080 ggccccactcaggcc 2095
          ||||||||||||||||
Qy     66 GGCCCCCACTCAGGCC 81


RESULT   18
ID   V02209 standard; cDNA; 3877 BP.
AC   V02209;
DT   27-MAR-1998  (first entry)
DE   cDNA of leaderless protein FGF-1.
KW   Leaderless protein; FGF-1; FGF-2; inhibitor; cardiac glycoside;
KW   aglycone; treatment; carcinoma; diabetes; ss.
OS   Homo sapiens.
PN   WO9728808-A1.
PD   14-AUG-1997.
PF   12-FEB-1997; U02237.
PR   12-FEB-1996; US-599895.
PA   (SCRI ) SCRIPPS RES INST.
PI   Florkiewicz RZ;
DR   WPI; 97-415065/38.
PT   Inhibition of export of leaderless protein from cells - using
PT   cardiac glycoside or its aglycone, e.g. ouabain or digoxin
PS   Claim 8; Pages 24-26; 61pp; English.
CC   This is the cDNA sequence of the leaderless protein FGF-1. The export of
CC   such leaderless proteins from a cell can be inhibited by a method which
CC   comprises contacting the cell with a cardiac glycoside or with an
CC   aglycone derivative of a cardiac glycoside. Preferably the glycoside is
CC   digoxin, strophanthin K, digitoxin, lanatoside A, ouabain, digitoxin,
CC   oleandrin or acovenoside A, and the aglycone is strophanthidin,
CC   digoxigenin, digitoxigenin or usarigenin. The method is useful for
CC   inhibiting export of leaderless proteins like FGF-1, FGF-2, IL-1 alpha,
CC   IL-1 beta , PD-ECGF, CNTF, thymosin, parathymosin, factor XIIIa, vas
CC   deferens protein, sciatic nerve growth promoting protein, L-14 lectin,
CC   transglutaminase, thioredoxin-like protein, HIV tat and int-2. Inhibition
CC   of export of FGF is useful for treating FGF-mediated pathophysiological
CC   conditions (e.g. melanoma, ovarian carcinoma, teratocarcinoma and
CC   neuroblastoma). It is useful for inhibiting proliferation of cells
CC   bearing an FGF-receptor, and for treating complications of diabetes.
SQ   Sequence   3877 BP;    1117 A;    762 C;    817 G;    1181 T;

  Query Match          10.9%;  Score 16;  DB 37;  Length 3877;
  Best Local Similarity 100.0%;  Pred. No. 4.67e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```



```
Db   3639 ttattattttaatata 3654
          ||||||||||||||||
Cp    139 TTATTATTTTAATATA 124


RESULT   19
ID   N81502 standard; DNA; 3878 BP.
AC   N81502;
DT   14-SEP-1990  (first entry)
DE   DNA encoding human basic fibroblast growth factor
KW   Human basic fibroblast growth factor; cell propagation acceleration;
KW   low toxicity; burn treatment; wound treatment; thrombosis treatment;
KW   arteriosclerosis treatment; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             340..807
FT                   /*tag=  a
FT                   /product="peptide"
FT   mat_peptide     367..807
FT                   /*tag=  b
FT   misc_feature    367..804
FT                   /*tag=  c
FT                   /note="claimed sequence in claim 4"
PN   J63226287-A.
PD   20-SEP-1988.
PF   11-MAR-1987; 056249.
PR   09-OCT-1986; JP-241053.
PR   11-MAR-1987; JP-231428.
PA   (TAKE) Takeda Chemical Ind Kk.
DR   WPI; 88-305161/43.
DR   P-PSDB; P81083.
PT   Polypeptide, DNA for human basic fibroblast growth factor -
PT   comprises speific aminoacid sequence, giving cell progogotion
PT   accelerating activity, low toxicity
PS   Disclosure; Fig 1-1 to 1-5, pages 562-6; 30pp; Japanese.
CC   The human basic fibroblast growth factor has cell propagation
CC   accelerating activity and low toxicity.  The growth factor is used as a
CC   medicine for burns, wounds, thrombosis or arteriosclerosis.  A large amt.
CC   of the growth factor is produced using recombinant DNA which contains
CC   the base sequence coding for the growth factor.  The growth factor is
CC   used in the form of powder, injections, tablets, capsules, liq. or
CC   ointment.
SQ   Sequence   3878 BP;    1117 A;    763 C;    817 G;    1181 T;

  Query Match          10.9%;  Score 16;  DB 1;  Length 3878;
  Best Local Similarity 100.0%;  Pred. No. 4.67e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   3639 ttattattttaatata 3654
          ||||||||||||||||
```



```
Cp    139 TTATTATTTTAATATA 124


RESULT   20
ID   N70494 standard; cDNA; 3878 BP.
AC   N70494;
DT   05-JUN-1991  (first entry)
DE   Sequence encoding human basic fibroblast growth factor (hbFGF).
KW   Fibroblast growth promoter; mesoderm cell growth promoter;
KW   wound healing; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             340..366
FT                   /*tag=  a
FT   mat_peptide     367..807
FT                   /*tag=  b
FT                   /note= "DNA claimed"
FT   polya_signal    886..891
FT                   /*tag=  c
FT   polya_site      3878
FT                   /*tag=  d
PN   EP-237966-A.
PD   23-SEP-1987.
PF   12-MAR-1987; 103601.
PR   14-MAR-1986; JP-057819.
PR   09-APR-1986; JP-082699.
PR   29-SEP-1986; JP-231428.
PR   09-OCT-1986; JP-241053.
PA   (TAKE ) TAKEDA CHEMICAL IND KK.
PI   Kurokawa T, Sasada R, Iwane M, Igarashi K;
DR   WPI; 87-265363/38.
DR   P-PSDB; P70301.
PT   Human basic fibroblast growth factor - produced by recombinant
PT   DNA techniques, useful for healing wounds, prophylaxis,
PT   thrombosis and arteriosclerosis treatment, etc.
PS   Disclosure; Fig 1; 38pp; English.
CC   hbFGF is produced using cDNA prepd. from RNA isolated from W138 or
CC   IMR90 human fibroblasts. hbFGF promotes growth of fibroblasts and
CC   other mesoderm-derived cells and is useful for promoting healing of
CC   wounds (e.g. burns), for prophylaxis and treatment of thrombosis and
CC   arteriosclerosis, and as a promoter for cell culture.
SQ   Sequence   3878 BP;    1117 A;    762 C;    817 G;    1182 T;

  Query Match          10.9%;  Score 16;  DB 2;  Length 3878;
  Best Local Similarity 100.0%;  Pred. No. 4.67e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   3639 ttattattttaatata 3654
          ||||||||||||||||
Cp    139 TTATTATTTTAATATA 124
```



```
RESULT   21
ID   Q55147 standard; DNA; 5247 BP.
AC   Q55147;
DT   22-SEP-1995  (first entry)
DE   Pseudomonas aeruginosa probe P2-17.
KW   Probe; S.aureus; S.epidermis; E.faecalis; P.aeruginosa; E.coli;
KW   K.pneumoniae; E.cloacae; clinical sample; ds.
OS   Pseudomonas aeruginosa.
PN   WO9401583-A.
PD   20-JAN-1994.
PF   07-JUL-1993; J00936.
PR   07-JUL-1992; JP-179719.
PA   (FUSO ) FUSO PHARM IND LTD.
PA   (OHNO/) OHNO T.
PI   Eda S,  Matsuhisa A,  Ohno T,  Uehara H;
DR   WPI; 94-035086/04.
PT   Probe for identifying bacteria causing infectious disease -
PT   consists of a DNA fragment obtained by HindIII cleavage of the
PT   pathogenic bacterial genomic DNA
PS   Claim 6; Page 56-59; 100pp; Japanese.
CC   The nucleotide sequence of a 5247 bp probe obtained by digestion of
CC   Pseudomonas aeruginosa genomic DNA with the restriction enzyme HindIII.
CC   The probes (Q55133-48 and Q55752-58) represent other probes derived,
CC   by HindIII digestion of the genomes of S.aureus, S.epidermis,
CC   Enterococcus faecalis, Pseudomonas aeruginosa, E.coli, Klebsiella
CC   pneumoniae or Enterobacter cloacae.  The probes can be used to detect
CC   their respective microorganisms in clinical samples.
SQ   Sequence   5247 BP;    1034 A;    1625 C;    1690 G;    897 T;

  Query Match          10.9%;  Score 16;  DB 14;  Length 5247;
  Best Local Similarity 100.0%;  Pred. No. 4.67e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     56 ggccagggctggcctg 71
          ||||||||||||||||
Cp     91 GGCCAGGGCTGGCCTG 76


RESULT   22
ID   Q05883 standard; DNA; 6767 BP.
AC   Q05883;
DT   11-JAN-1991  (first entry)
DE   Plasmid clone pUC-SK1.
KW   Basic fibroblast growth factor; tissue regeneration; infarction; ss.
FH   Key             Location/Qualifiers
FT   cds             302..934
FT                   /*tag=  a
FT                   /label=basic fibroblast growth factor
PN   FR2642086-A.
PD   27-JUL-1990.
```

```
PF   26-JAN-1989; 000973.
PR   26-JAN-1989; FR-000973.
PA   (SNFI ) SANOFI SA.
PI   Caput D, Ferrara P, Kaghad M;
DR   WPI; 90-277408/37.
DR   P-PSDB; R06685.
PT   New recombinant gene encoding basic fibroblast growth factor - in
PT   new high mol. wt. form, useful e.g. for stimulating tissue
PT   regenerating or treating infarction
PS   Disclosure; fig 3; 43pp; French.
CC   This recombinant gene encodes a basic fibroblast growth factor (bFGF)
CC   which stimulates growth of mesodermal and neuroectodermal cells.
CC   It is thus potentially useful e.g. for regenerating damaged tissues,
CC   and for treating myocardial infarctions, Parkinsons disease and
CC   Alzheimers disease.  It can be produced on a large scale using rec-
CC   ombinant DNA methods without risk of contamination.  There are a
CC   further 3 potential initiation codons allowing expression of larger
CC   forms of the protein.  See also Q05884.
SQ   Sequence   6767 BP;    2047 A;    1201 C;    1334 G;    2185 T;

  Query Match          10.9%;  Score 16;  DB 1;  Length 6767;
  Best Local Similarity 100.0%;  Pred. No. 4.67e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   3764 ttattattttaatata 3779
         ||||||||||||||||
Cp    139 TTATTATTTTAATATA 124


RESULT   23
ID   T09847 standard; DNA; 7902 BP.
AC   T09847;
DT   02-JAN-1997  (first entry)
DE   Human papilloma virus 16 genomic DNA.
KW   Human papilloma virus; HPV; ribozyme; cleavage; E6; E7;
KW   keratinocyte; infection; genital warts; neoplasia; cervical cancer;
KW   warts; ds.
OS   Human papilloma virus 16.
PN   WO9531552-A2.
PD   23-NOV-1995.
PF   15-MAY-1995; U06016.
PR   13-MAY-1994; US-242665.
PA   (UYNI-) UNIV NORTHERN ILLINOIS.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI   Dipaolo J,  Hampel A,  Siwkowski A;
DR   WPI; 96-010937/01.
PT   Synthetic ribozyme(s) for human papilloma virus (HPV) - useful in
PT   treatment of cervical cancer and in detection of HPV
PS   Disclosure; Page 25-29; 60pp; English.
CC   Synthetic ribozymes directed against human papilloma virus (HPV) 16
CC   may be used to cleave HPV or to detect HPV in human tissue.  The
```



```
CC   ribozymes cleave HPV to lower the production of the E6 and E7
CC   proteins which play a key role in keratinocyte transformation.  They
CC   can be used for the treatment of HPV infections which are associated
CC   with genital warts and genital neoplasms.  In particular they can be
CC   used for the treatment of cervical cancer optionally with other
CC   agents such as cisplatin.  Ribozymes are described in T40306-T40315.
SQ   Sequence   7902 BP;    2601 A;    1376 C;    1507 G;    2418 T;

  Query Match          10.9%;  Score 16;  DB 24;  Length 7902;
  Best Local Similarity 100.0%;  Pred. No. 4.67e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   6329 cagattatattaaaat 6344
         ||||||||||||||||
Qy    119 CAGATTATATTAAAAT 134


RESULT   24
ID   T94724 standard; DNA; 7904 BP.
AC   T94724;
DT   16-FEB-1998  (first entry)
DE   Human papillomavirus type 16 E6 variant nucleotide sequence.
KW   Complete genome; circular; human papillomavirus type 16; HPV16 E6;
KW   cervical dysplasia; cervical cancer; cervical smear; ds.
OS   Human papillomavirus type 16.
FH   Key             Location/Qualifiers
FT   TATA_signal     17..23
FT                   /*tag=  a
FT   TATA_signal     65..71
FT                   /*tag=  b
FT   CDS             83..559
FT                   /*tag=  c
FT                   /product= E6_transforming_protein
FT                   /note= "E6 ORF from 65 to 559; putative"
FT   mutation        350
FT                   /*tag=  d
FT                   /note= "Mutated from the normal T in the reference
FT                        sequence (T94723)"
FT   CDS             562..858
FT                   /*tag=  e
FT                   /product= E7_transforming_protein
FT                   /note= "E7 ORF 544 to 858; putative"
FT   CDS             865..2813
FT                   /*tag=  f
FT                   /product= E1_replication_protein
FT                   /note= "E1 interrupted ORF from 859 to 2813; putative"
FT   CDS             2755..3852
FT                   /*tag=  g
FT                   /product= E2_regulatory_protein
FT                   /note= "E2 ORF from 2725 to 3852; putative"
FT   CDS             3332..3619
```



```
FT                   /*tag=  h
FT                   /product= E4
FT                   /note= "E4 ORF from 3332 to 3619; putative"
FT   CDS             3863..4099
FT                   /*tag=  i
FT                   /product= E5
FT                   /note= "E5 ORF from 3863 to 4099; putative"
FT   polyA_signal    4213..4218
FT                   /*tag=  j
FT                   /note= "Putative"
FT   CDS             4235..5656
FT                   /*tag=  k
FT                   /product= L2_minor_capsid_protein
FT                   /note= "L2 ORF from 4133 to 5656; putative"
FT   TATA_signal     4289..4295
FT                   /*tag=  l
FT   CDS             5559..7154
FT                   /*tag=  m
FT                   /product= L1_major_capsid_protein
FT                   /note= "L1 ORF from 5526 to 7154; putative"
FT   polyA_signal    7260..7265
FT                   /*tag=  n
PN   US5679509-A.
PD   21-OCT-1997.
PF   28-SEP-1993; 127906.
PR   30-SEP-1994; US-316239.
PR   28-SEP-1993; US-127906.
PA   (UYNE-) UNIV NEW MEXICO STATE.
PI   Parmenter CA,  Wheeler CM;
DR   WPI; 97-525714/48.
DR   P-PSDB; W35742.
PT   Evaluating risk of cervical dysplasia or cervical cancer - by
PT   detecting variant form of human papilloma virus 16
PS   Claim 1; Column 15-24; 33pp; English.
CC   Methods have been developed for distinguishing a subset of human
CC   papilloma virus (HPV) that is associated with an increased risk of
CC   developing cervical dysplasia or cervical cancer. The methods involve:
CC   (1) preparing a cervical sample to expose any HPV-16 E6 gene in the
CC   sample and determining if the base at position 350 of the E6 gene
CC   (see T94723 and T94724 for comparison) is T or G, where the presence of
CC   G at position 350 is associated with an increased risk of developing
CC   cervical dysplasia or cervical cancer; and (2) preparing a cervical
CC   sample to expose any HPV-16 E6 protein in the sample and determining
CC   if the amino acid at position 83 of the protein (see position 90 in
CC   W35741 and w35742 for comparison) is Val or Leu, where the presence of
CC   Val at position 83 that is associated with an increased risk of
CC   developing cervical dysplasia or cervical cancer. The present sequence
CC   represents the variant nucleotide sequence for HPV-16 E6. The 350G
CC   variant correlates well with Pap scores: 350T:350G ratios among 45
CC   HPV16 samples were 10:4 for negative Pap scores; 4:2 for CIN I, 1:6 for
CC   CIN II; 2:9 for CIN III; 0:3 for cancer.
```



```
SQ   Sequence   7904 BP;    2601 A;    1377 C;    1510 G;    2416 T;

  Query Match          10.9%;  Score 16;  DB 35;  Length 7904;
  Best Local Similarity 100.0%;  Pred. No. 4.67e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   6331 cagattatattaaaat 6346
         ||||||||||||||||
Qy    119 CAGATTATATTAAAAT 134


RESULT   25
ID   T94723 standard; DNA; 7904 BP.
AC   T94723;
DT   16-FEB-1998  (first entry)
DE   Human papillomavirus type 16 E6 complete genome.
KW   Complete genome; circular; human papillomavirus type 16; HPV16 E6;
KW   cervical dysplasia; cervical cancer; cervical smear; ds.
OS   Human papillomavirus type 16.
FH   Key             Location/Qualifiers
FT   TATA_signal     17..23
FT                   /*tag=  a
FT   TATA_signal     65..71
FT                   /*tag=  b
FT   CDS             83..559
FT                   /*tag=  c
FT                   /product= E6_transforming_protein
FT                   /note= "E6 ORF from 65 to 559; putative"
FT   mutation        350
FT                   /*tag=  d
FT                   /note= "Mutated from the normal T (as in this case)
FT                        to G in the variant"
FT   CDS             562..858
FT                   /*tag=  e
FT                   /product= E7_transforming_protein
FT                   /note= "E7 ORF 544 to 858; putative"
FT   CDS             865..2813
FT                   /*tag=  f
FT                   /product= E1_replication_protein
FT                   /note= "E1 interrupted ORF from 859 to 2813; putative"
FT   CDS             2755..3852
FT                   /*tag=  g
FT                   /product= E2_regulatory_protein
FT                   /note= "E2 ORF from 2725 to 3852; putative"
FT   CDS             3332..3619
FT                   /*tag=  h
FT                   /product= E4
FT                   /note= "E4 ORF from 3332 to 3619; putative"
FT   CDS             3863..4099
FT                   /*tag=  i
FT                   /product= E5
```

```
FT                   /note= "E5 ORF from 3863 to 4099; putative"
FT   polyA_signal    4213..4218
FT                   /*tag=  j
FT                   /note= "Putative"
FT   CDS             4235..5656
FT                   /*tag=  k
FT                   /product= L2_minor_capsid_protein
FT                   /note= "L2 ORF from 4133 to 5656; putative"
FT   TATA_signal     4289..4295
FT                   /*tag=  l
FT   CDS             5559..7154
FT                   /*tag=  m
FT                   /product= L1_major_capsid_protein
FT                   /note= "L1 ORF from 5526 to 7154; putative"
FT   polyA_signal    7260..7265
FT                   /*tag=  n
PN   US5679509-A.
PD   21-OCT-1997.
PF   28-SEP-1993; 127906.
PR   30-SEP-1994; US-316239.
PR   28-SEP-1993; US-127906.
PA   (UYNE-) UNIV NEW MEXICO STATE.
PI   Parmenter CA,  Wheeler CM;
DR   WPI; 97-525714/48.
DR   P-PSDB; W35741.
PT   Evaluating risk of cervical dysplasia or cervical cancer - by
PT   detecting variant form of human papilloma virus 16
PS   Claim 1; Column 9-16; 33pp; English.
CC   Methods have been developed for distinguishing a subset of human
CC   papilloma virus (HPV) that is associated with an increased risk of
CC   developing cervical dysplasia or cervical cancer. The methods involve:
CC   (1) preparing a cervical sample to expose any HPV-16 E6 gene in the
CC   sample and determining if the base at position 350 of the E6 gene
CC   (see T94723 and T94724 for comparison) is T or G, where the presence of
CC   G at position 350 is associated with an increased risk of developing
CC   cervical dysplasia or cervical cancer; and (2) preparing a cervical
CC   sample to expose any HPV-16 E6 protein in the sample and determining
CC   if the amino acid at position 83 of the protein (see position 90 in
CC   W35741 and w35742 for comparison) is Val or Leu, where the presence of
CC   Val at position 83 that is associated with an increased risk of
CC   developing cervical dysplasia or cervical cancer. The present sequence
CC   represents the reference nucleotide sequence for HPV-16 E6. The 350G
CC   variant correlates well with Pap scores: 350T:350G ratios among 45
CC   HPV16 samples were 10:4 for negative Pap scores; 4:2 for CIN I, 1:6 for
CC   CIN II; 2:9 for CIN III; 0:3 for cancer.
SQ   Sequence   7904 BP;    2601 A;    1377 C;    1509 G;    2417 T;

  Query Match           10.9%;  Score 16;  DB 35;  Length 7904;
  Best Local Similarity 100.0%;  Pred. No. 4.67e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```



```
Db   6331 cagattatattaaaat 6346
          ||||||||||||||||
Qy    119 CAGATTATATTAAAAT 134


RESULT   26
ID   T22689 standard; cDNA to mRNA; 256 BP.
AC   T22689;
DT   15-AUG-1996  (first entry)
DE   Human gene signature HUMGS04345.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 1191; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   256 BP;    79 A;    45 C;    34 G;    92 T;

  Query Match           10.2%;  Score 15;  DB 20;  Length 256;
  Best Local Similarity  93.8%;  Pred. No. 1.94e+01;
  Matches   15;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db   220 atnttaaaatataaat 235
         || ||||||||||||||
Qy   125 ATATTAAAATAATAAT 140
```



```
RESULT   27
ID   T20016 standard; cDNA to mRNA; 304 BP.
AC   T20016;
DT   17-JUL-1996  (first entry)
DE   Human gene signature HUMGS01154.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 537; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   304 BP;    69 A;    93 C;    51 G;    88 T;

  Query Match           10.2%;  Score 15;  DB 19;  Length 304;
  Best Local Similarity 100.0%;  Pred. No. 1.94e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   186 gggggccagggctgg 200
          ||||||||||||||||
Cp    94 GGGGGCCAGGGCTGG 80


RESULT   28
```



```
ID   T80680 standard; DNA; 423 BP.
AC   T80680;
DT   11-NOV-1997  (first entry)
DE   Type II topoisomerase database reference sequence SEQ ID NO:93.
KW   Detection; identification; TopoII; contamination; food; human;
KW   animal; plant; soil; water; PCR; polymerase chain reaction; ds.
OS   Shigella dysenteriae.
PN   US5645994-A.
PD   08-JUL-1997.
PF   05-JUL-1990; 548138.
PR   06-JUN-1995; US-470179.
PR   05-JUL-1990; US-548138.
PR   13-AUG-1993; US-106482.
PA   (UTAH ) UNIV UTAH RES FOUND.
PI   Huang WM;
DR   WPI; 97-362925/33.
PT   Detection and identification of organisms - using DNA primers to
PT   amplify signature segment of organism's type II topoisomerase
PS   Claim 6; Column 99-102; 114pp; English.
CC   A method has been produced for selectively amplifying DNA segments of
CC   one or more species of organisms in a sample. The method involves: (a)
CC   providing a database containing reference sequences, comprising a
CC   subunit sequence of a signature region of a macromolecule selected from
CC   a type II topoisomerase (TopoII), or a homologue, where each reference
CC   sequence is specific to a different species of a chosen group, and the
CC   macromolecule comprises 1st and 2nd conserved regions adjacently
CC   flanking the signature region; and (b) making an extract of DNA
CC   molecules, and selectively amplifying DNA segments of the signature
CC   region using a universal primer composition, comprising a primer
CC   constructed to bind a DNA encoding the macromolecule, to produce
CC   amplified DNA segments. The present sequence represents a DNA fragment
CC   containing a signature segment which is used in the database as a
CC   reference sequence for Shigella dysenteriae. The method can be
CC   used to identify all of the different organisms present in a single
CC   sample without using multiple probes. It can accurately distinguish
CC   among similar and related species, and can be used with, e.g.
CC   contaminated food products, tissue or fluid samples from diseased
CC   humans, animal or plants, soil samples and water samples from any
CC   source.
SQ   Sequence   423 BP;    106 A;    107 C;    110 G;    100 T;

  Query Match           10.2%;  Score 15;  DB 33;  Length 423;
  Best Local Similarity 100.0%;  Pred. No. 1.94e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   206 gtaacttcggttcca 220
          ||||||||||||||||
Cp   119 GTAACTTCGGTTCCA 105


RESULT   29
```

```
ID   T80710 standard; DNA; 423 BP.
AC   T80710;
DT   11-NOV-1997  (first entry)
DE   Type II topoisomerase database reference sequence SEQ ID NO:129.
KW   Detection; identification; TopoII; contamination; food; human;
KW   animal; plant; soil; water; PCR; polymerase chain reaction; ds.
OS   Chloroflexus aurantiacus.
PN   US5645994-A.
PD   08-JUL-1997.
PF   05-JUL-1990; 548138.
PR   06-JUN-1995; US-470179.
PR   05-JUL-1990; US-548138.
PR   13-AUG-1993; US-106482.
PA   (UTAH ) UNIV UTAH RES FOUND.
PI   Huang WM;
DR   WPI; 97-362925/33.
PT   Detection and identification of organisms - using DNA primers to
PT   amplify signature segment of organism's type II topoisomerase
PS   Claim 6; Column 129-132; 114pp; English.
CC   A method has been produced for selectively amplifying DNA segments of
CC   one or more species of organisms in a sample. The method involves: (a)
CC   providing a database containing reference sequences, comprising a
CC   subunit sequence of a signature region of a macromolecule selected from
CC   a type II topoisomerase (TopoII), or a homologue, where each reference
CC   sequence is specific to a different species of a chosen group, and the
CC   macromolecule comprises 1st and 2nd conserved regions adjacently
CC   flanking the signature region; and (b) making an extract of DNA
CC   molecules, and selectively amplifying DNA segments of the signature
CC   region using a universal primer composition, comprising a primer
CC   constructed to bind a DNA encoding the macromolecule, to produce
CC   amplified DNA segments. The present sequence represents a DNA fragment
CC   containing a signature segment which is used in the database as a
CC   reference sequence for Chloroflexus aurantiacus. The method can be
CC   used to identify all of the different organisms present in a single
CC   sample without using multiple probes. It can accurately distinguish
CC   among similar and related species, and can be used with, e.g.
CC   contaminated food products, tissue or fluid samples from diseased
CC   humans, animal or plants, soil samples and water samples from any
CC   source.
SQ   Sequence   423 BP;    87 A;    120 C;    124 G;    92 T;

  Query Match          10.2%;  Score 15;  DB 33;  Length 423;
  Best Local Similarity 100.0%;  Pred. No. 1.94e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    206 gtaacttcggttcca 220
          |||||||||||||||
Cp    119 GTAACTTCGGTTCCA 105


RESULT   30
```



```
ID   Q52274 standard; DNA; 744 BP.
AC   Q52274;
DT   01-JUN-1994  (first entry)
DE   Soybean seed storage protein glycinin gene terminator.
KW   Terminator; glycinin; plasmid pSUM-GY1; ds.
OS   Glycine max.
PN   EP-571741-A.
PD   01-DEC-1993.
PF   29-MAR-1993; 105138.
PR   27-MAR-1992; JP-070647.
PA   (SUMO ) SUMITOMO CHEM CO LTD.
PI   Iida A,  Nagasawa A,  Oheda K;
DR   WPI; 93-378708/48.
PT   Expression plasmid - comprises soybean seed storage protein
PT   glycinin gene, for gene expression in e.g. seeds
PS   Claim 2; Page 7; 16pp; English.
CC   Expression plasmid pSUM-GY1 (FERM BP-4131) comprising the
CC   terminator and the glycinin promoter (Q52273) is useful for
CC   expressing a foreign gene in a transgenic plant.
SQ   Sequence   744 BP;    252 A;    117 C;    121 G;    254 T;

  Query Match          10.2%;  Score 15;  DB 9;  Length 744;
  Best Local Similarity 100.0%;  Pred. No. 1.94e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     73 taaaataataatgta 87
          |||||||||||||||
Qy    129 TAAAATAATAATGTA 143


RESULT   31
ID   T68963 standard; cDNA; 758 BP.
AC   T68963;
DT   05-AUG-1997  (first entry)
DE   Cowpox virus p35 chemokine inhibitor cDNA.
KW   Chemokine inhibitor; chemotaxis-stimulating cytokine; inflammation;
KW   antiinflammatory; ss.
OS   Cowpox virus.
FH   Key             Location/Qualifiers
FT   cds             18..758
FT                   /*tag=  a
FT   signal_peptide  18..65
FT                   /*tag=  b
FT   mat_peptide     66..755
FT                   /*tag=  c
PN   WO9711714-A1.
PD   03-APR-1997.
PF   26-SEP-1996; U15385.
PR   29-SEP-1995; US-537324.
PR   20-DEC-1995; US-575715.
PA   (IMMV ) IMMUNEX CORP.
```



```
PI   Smith CA;
DR   WPI; 97-212669/19.
DR   P-PSDB; W17058.
PT   Use of viral p35 polypeptide(s) for inhibiting chemokines - used
PT   for treating chemokine-mediated conditions such as inflammation and
PT   in binding assays
PS   Example 1; Page 26-27; 45pp; English.
CC   cDNA clones (T68963-65) respectively code for the p35 polypeptides
CC   (W17058-60) of cowpox virus, vaccinia virus Lister strain and
CC   variola (smallpox) virus.  p53 is a secreted protein that can be
CC   purified from the culture supernatant of cells infected with these
CC   viruses.  It binds to a number of chemotaxis-stimulating cytokines
CC   (chemokines) and can be used to treat chemokine-mediated conditions
CC   such as inflammation, or in binding assays.  Recombinant p35
CC   polypeptides can be produced in engineered host cells utilising
CC   expression vectors incorporating the isolated p35 DNA sequences.
SQ   Sequence   758 BP;    248 A;    170 C;    163 G;    177 T;

  Query Match          10.2%;  Score 15;  DB 30;  Length 758;
  Best Local Similarity 100.0%;  Pred. No. 1.94e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    206 aaccgaagttacaga 220
          |||||||||||||||
Qy    108 AACCGAAGTTACAGA 122


RESULT   32
ID   Q78979 standard; DNA; 819 BP.
AC   Q78979;
DT   04-AUG-1995  (first entry)
DE   Human immunoglobulin variable heavy chain #41.
KW   Primer; PCR; amplify; human; immunoglobulin; variable; heavy chain;
KW   cosmid; placenta; vector; pJB81; E.coli; mammalian; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             189..639
FT                   /*tag=  a
FT                   /product= human immunoglobulin variable heavy chain
FT   intron          235..337
FT                   /*tag=  b
FT   misc_signal     487..489
FT                   /*tag=  c
FT                   /note= "non-functional terminator codon"
FT   misc_signal     604..606
FT                   /*tag=  d
FT                   /note= "non-functional terminator codon"
FT   misc_signal     640..642
FT                   /*tag=  e
FT                   /note= "miscellaneous signal, does not conform to
FT                   terminator or splice site sequence"
```



```
PN   WO9426895-A.
PD   24-NOV-1994.
PF   10-MAY-1993; J00603.
PR   10-MAY-1993; WO-J00603.
PA   (NISB ) JAPAN TOBACCO INC.
PI   Honjo T,  Matsuda F;
DR   WPI; 95-000791/01.
DR   P-PSDB; R66330.
PT   DNA fragment comprising human immunoglobulin Vh genes - for the
PT   production of human immunoglobulin in mammalian hosts
PS   Claim 48; Page 83-84; 130pp; Japanese.
CC   A series of genes (Q78939-79002) encoding human immunoglobulin variable
CC   heavy chains. The genes were isolated and cloned from a series of cosmid
CC   constructs: Y202; Y103; Y21; Y6;Y24; 3-31; M84; M118 and M131, by PCR
CC   amplification using primers Q78917-38.  The genes are subdivided into 5
CC   families of Vh genes.  The fragments cover a region of 800 kb.  The DNA
CC   fragments were isolated from high molecular weight DNA from human
CC   placenta.  The DNA was partially digested with TaqI restriction enzyme.
CC   The fragments were separated by gel elecrophoresis and 35-45 kb fractions
CC   were collected.  The fragments were ligated with ClaI-digested cosmid
CC   vector pJB81.  The ligation products were in vitro packed and infected
CC   into E.coli 490A.  The fragments were then subcloned by colony
CC   hybridisation.  The Vh genes and the DNA fragments encoding them are
CC   useful in producing human immunoglobulin in mammalian hosts.
SQ   Sequence   819 BP;    190 A;    192 C;    241 G;    193 T;

  Query Match          10.2%;  Score 15;  DB 13;  Length 819;
  Best Local Similarity 100.0%;  Pred. No. 1.94e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    315 ccagggtttctttgt 329
          |||||||||||||||
Qy      6 CCAGGGTTTCTTTGT 20


RESULT   33
ID   T28972 standard; cDNA; 1374 BP.
AC   T28972;
DT   21-AUG-1996  (first entry)
DE   Mouse CRF receptor genomic clone mCRF-RB1.
KW   Corticotropin releasing factor receptor; CRF-R; corticoliberin;
KW   signal transduction; ds.
OS   Mus sp.
FH   Key             Location/Qualifiers
FT   cds             79..1374
FT                   /*tag=  a
PN   WO9617934-A2.
PD   13-JUN-1996.
PF   06-DEC-1995; U15909.
PR   09-DEC-1994; US-353537.
PR   17-JAN-1995; US-374009.
```

```
PA   (SALK ) SALK INST BIOLOGICAL STUDIES.
PI   Chen R, Donaldson CJ, Lewis KA, Perrin MH, Sawchenko P;
PI   Vale WW;
DR   WPI; 96-287179/29.
DR   P-PSDB; R97293.
PT   Isolated corticotropin-releasing factor receptor (CRF-R) - used to
PT   develop prods. for modulating signal transduction activity mediated
PT   by CRF-R
PS   Claim 6; Page 83-85; 102pp; English.
CC   A genomic clone (T28972) codes for mouse corticotropin
CC   releasing factor receptor mCRF-RB1 (R97293). It was isolated
CC   isolated from a mouse heart cDNA library using a probe obtd.
CC   by PCR amplification (see also T28973-74) of an exon of mCRF-RB1
CC   The mCRF-RB1 cDNA can be used as a probe for detecting the presence
CC   of CRF receptor-encoding nucleic acids in samples or for the
CC   identification of additional CRF receptors, or can be used for prodn.
CC   of recombinant CRF receptor, or in gene therapy to target a vector to
CC   specific cell types.
SQ   Sequence   1374 BP;    279 A;    421 C;    344 G;    330 T;

  Query Match          10.2%;  Score 15;  DB 20;  Length 1374;
  Best Local Similarity 100.0%;  Pred. No. 1.94e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1188 ccagggtttctttgt 1202
         |||||||||||||||
Qy      6 CCAGGGTTTCTTTGT 20


RESULT   34
ID   Q74063 standard; cDNA to mRNA; 1892 BP.
AC   Q74063;
DT   29-JAN-1996  (first entry)
DE   The mouse beta-actin gene.
KW   Beta-actin; primer; mRNA; specificity; pharmaceutical; ss.
OS   Mus sp.
PN   JO7123984-A.
PD   16-MAY-1995.
PF   05-NOV-1993; 275852.
PR   05-NOV-1993; JP-275852.
PA   (HITB ) HITACHI CHEM CO LTD.
DR   WPI; 95-211627/28.
PT   A primer for the detection and the determn. of a specific messenger
PT   RNA - can detect and determine specific mRNA(s) with high
PT   reliability
PS   Example 27; Page 29-30; 35pp; Japanese.
CC   Q74063 is the mouse beta-actin gene. This gene is amplified by the
CC   primers Q74045 and Q74046. The primers are used specifically for the
CC   detection and isolation of this sequence. They have the advantage of
CC   high sensitivity and reliability and are useful in the pharmaceutical
CC   industry.
```



```
SQ   Sequence   1892 BP;    418 A;    506 C;    459 G;    509 T;

  Query Match          10.2%;  Score 15;  DB 16;  Length 1892;
  Best Local Similarity 100.0%;  Pred. No. 1.94e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1830 gccagggctggcctg 1844
         |||||||||||||||
Cp     90 GCCAGGGCTGGCCTG 76


RESULT   35
ID   T28190 standard; DNA; 2020 BP.
AC   T28190;
DT   17-DEC-1996  (first entry)
DE   Feline herpesvirus glycoprotein E gene.
KW   Feline herpesvirus; FHV; deletion; insertion; glycoprotein E;
KW   gp E; unique short region; attenuated virus; vaccine; protection;
KW   cat; vector; delivery; antigen; ds.
OS   Feline herpesvirus.
FH   Key             Location/Qualifiers
FT   cds             413..2011
FT                   /*tag=  a
PN   WO9613575-A1.
PD   09-MAY-1996.
PF   26-OCT-1995; U13975.
PR   26-OCT-1994; US-329883.
PA   (SYTR ) SYNTRO CORP.
PI   Cochran MD, McDonnell MW;
DR   WPI; 96-239489/24.
DR   P-PSDB; R96588.
PT   Recombinant feline herpes virus attenuated by alteration of the gE
PT   gene - and related homology vectors, useful in vaccines, also as
PT   vectors for delivering antigens or therapeutic agents to mammals and
PT   birds
PS   Example 1; Pages 76-78; 162pp; English.
CC   A novel feline herpesvirus (FHV) contains a deletion or insertion
CC   in the glycoprotein E (gp E) gene, i.e. the present sequence, in
CC   the unique short region of the viral genome. Altering the gp E
CC   gene, so that it no longer produces functional gp E, attenuates the
CC   virus. The attenuated virus is useful in vaccines to protect cats
CC   against FHV (and opt. against other diseases if appropriate DNA is
CC   inserted), and as a vector for delivering vaccinating antigens or
CC   therapeutic agents (e.g. antisense mols., riboxymes, interferon
CC   inducers, hormones, lymphokines, etc.) to other mammals (including
CC   humans) or birds.
SQ   Sequence   2020 BP;    631 A;    439 C;    390 G;    560 T;

  Query Match          10.2%;  Score 15;  DB 23;  Length 2020;
  Best Local Similarity 100.0%;  Pred. No. 1.94e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```



```
Db    316 agattatattaaaat 330
          |||||||||||||||
Qy    120 AGATTATATTAAAAT 134


RESULT   36
ID   T49318 standard; cDNA; 2122 BP.
AC   T49318;
DT   30-MAR-1997  (first entry)
DE   Wnt-10b gene.
KW   Human; Wnt-10b; fetal bone marrow stroma cell; growth factor;
KW   antibody; antisense; hematopoietic stem cell cultulre;
KW   gene therapy; ex vivo; diagnostic; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   5'utr           1..193
FT                   /*tag=  a
FT   cds             194..1363
FT                   /*tag=  b
FT                   /product= Wnt-10b protein
FT   signal_peptide  194..256
FT                   /*tag=  c
FT   mat_peptide     257..1360
FT                   /*tag=  d
FT                   /product= Mature Wnt-10b (claim 19)
FT   3'utr           1364..2122
FT                   /*tag=  e
PN   WO9640910-A1.
PD   19-DEC-1996.
PF   06-JUN-1996; E02455.
PR   07-JUN-1995; US-485449.
PA   (SANO ) SANDOZ LTD.
PA   (SANO ) SANDOZ PATENT GMBH.
PA   (SANO ) SANDOZ-ERFINDUNGEN VERW GMBH.
PA   (SYST-) SYSTEMIX INC.
PI   Van Den Berg DJ;
DR   WPI; 97-052317/05.
DR   P-PSDB; W08928.
PT   Expression vectors and methods for cloning N-terminal signal
PT   sequences - used for selection, expression and isolation of new
PT   signal sequences
PS   Claim 3; Page 28-30; 46pp; English.
CC   The sequence encodes human growth factor Wnt-10b, and has been
CC   isolated from a fetal bone marrow stroma cell cDNA library by PCR
CC   using consensus primers corresponding to conserved sequences within
CC   the Wnt gene family, followed by use of the amplified fragment as a
CC   probe and chromosome walking for isolation of the full-length
CC   sequence. A truncated sequence, Wnt-10b-delta, is given in T49319
CC   (claim 3). A plasmid containing the Wnt-10b cDNA has been deposited
CC   as ATCC 97208. Antibodies and antisense sequences corresponding to
```



```
CC   Wnt-10b may be used as pharmaceuticals, and Wnt-10b may be used in
CC   a culture medium for ex vivo haematopoietic stem cell culture for
CC   use e.g. in gene therapy. Fragments of the gene may be used as
CC   diagnostic probes or primers.
SQ   Sequence   2122 BP;    415 A;    604 C;    635 G;    468 T;

  Query Match          10.2%;  Score 15;  DB 27;  Length 2122;
  Best Local Similarity 100.0%;  Pred. No. 1.94e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    750 ccagccctggccccc 764
          |||||||||||||||
Qy     80 CCAGCCCTGGCCCCC 94


RESULT   37
ID   Q25698 standard; cDNA; 2256 BP.
AC   Q25698;
DT   28-DEC-1992  (first entry)
DE   Sequence encoding pre-pro stromelysin-3.
KW   Metalloprotease; stromelysin-3; cancer; diagnosis; tumour;
KW   detection; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   signal_peptide  10..102
FT                   /*tag=  a
FT                   /note= "could be 10..108, rather than 10..102"
FT   cds             103..476
FT                   /*tag=  b
FT                   /note= "could be 109..476 rather than 103..476"
FT   polya_signal    2228..2233
FT                   /*tag=  c
PN   WO9209701-A.
PD   11-JUN-1992.
PF   21-NOV-1991; F00924.
PR   21-NOV-1990; GB-025326.
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PI   Basset P, Bellocq JP, Chambon P;
DR   WPI; 92-217079/26.
DR   P-PSDB; R24862.
PT   Diagnosis and treatment of invasive tumours - especially breast
PT   cancer, using nucleotide probes and antibodies specific for
PT   stromelysin-3
PS   Example; Fig 2; 70pp; French.
CC   A cDNA bank was established from polyA-RNA of a fibroadenoma and
CC   screened with probes derived from polyA-RNA of metastatic tumours.
CC   Clones binding strongly were selected; one of these was found to be
CC   expressed at a high level only in malignant cancers of the breast,
CC   pharynx, head, neck and skin., which are characterised by
CC   deterioration of the extracellular matrix. This clone, which encodes
CC   stromolysin-3 (ST3), was sequenced (Q25698). A probe corresponding
```

```
CC   to human cDNA encoding ST3 was used to screen a mouse placenta cDNA
CC   library. This led to the isolation of cDNA with the sequence in
CC   Q35699. A comparison of the mouse and human ST3 sequences revealed
CC   89% homology in the amino acids of the mature protein; and
CC   approximately 55% homology in the pre- and pro-domains.
SQ   Sequence   2256 BP;    402 A;    714 C;    676 G;    464 T;

  Query Match          10.2%;  Score 15;  DB 4;  Length 2256;
  Best Local Similarity 100.0%;  Pred. No. 1.94e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    781 ggccagccctggccc 795
          |||||||||||||||
Qy     78 GGCCAGCCCTGGCCC 92


RESULT   38
ID   T12938 standard; DNA; 2316 BP.
AC   T12938;
DT   17-SEP-1996  (first entry)
DE   Prophage P1-derived plasmid, encodes modified repA protein.
KW   RepA protein; mutein; recombination; high copy number plasmid;
KW   plasmid partition; stable plasmid maintenance; bacteriophage P1;
KW   incompatibility region; temperature sensitive; drug resistance factor;
KW   Rts1; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   cds             182..574
FT                   /*tag=  a
FT                   /product= RepA_(modified)
FT   cds             891..1751
FT                   /*tag=  b
FT                   /product= IncC_(modified)
FT   repeat_unit     8..27
FT                   /*tag=  c
FT                   /rpt_type= DIRECT
FT                   /note= "repeat A"
FT   repeat_unit     246..262
FT                   /*tag=  d
FT                   /rpt_type= DIRECT
FT                   /note= "repeat A"
FT   repeat_unit     289..308
FT                   /*tag=  e
FT                   /rpt_type= DIRECT
FT                   /note= "repeat W"
FT   repeat_unit     345..362
FT                   /*tag=  f
FT                   /rpt_type= DIRECT
FT                   /note= "repeat W"
FT   repeat_unit     296..308
FT                   /*tag=  g
```



```
FT                   /rpt_type= DIRECT
FT                   /note= "repeat Z"
FT   repeat_unit     644..666
FT                   /*tag=  h
FT                   /rpt_type= DIRECT
FT                   /note= "repeat Z"
FT   repeat_unit     668..687
FT                   /*tag=  i
FT                   /rpt_type= DIRECT
FT                   /note= "repeat Z"
FT   repeat_unit     338..358
FT                   /*tag=  j
FT                   /rpt_type= DIRECT
FT                   /note= "repeat X"
FT   repeat_unit     891..919
FT                   /*tag=  k
FT                   /rpt_type= DIRECT
FT                   /note= "repeat X"
FT   repeat_unit     634..649
FT                   /*tag=  l
FT                   /rpt_type= DIRECT
FT                   /note= "repeat Y"
FT   repeat_unit     667..685
FT                   /*tag=  m
FT                   /rpt_type= DIRECT
FT                   /note= "repeat Y"
FT   repeat_unit     732..750
FT                   /*tag=  n
FT                   /rpt_type= DIRECT
FT                   /note= "repeat 14"
FT   repeat_unit     753..771
FT                   /*tag=  o
FT                   /rpt_type= DIRECT
FT                   /note= "repeat 13"
FT   repeat_unit     774..792
FT                   /*tag=  p
FT                   /rpt_type= DIRECT
FT                   /note= "repeat 12"
FT   repeat_unit     795..813
FT                   /*tag=  q
FT                   /rpt_type= DIRECT
FT                   /note= "repeat 11"
FT   repeat_unit     817..835
FT                   /*tag=  r
FT                   /rpt_type= DIRECT
FT                   /note= "repeat 10"
FT   repeat_unit     1750..1769
FT                   /*tag=  s
FT                   /rpt_type= DIRECT
FT                   /note= "repeat 9"
FT   repeat_unit     1782..1799
```



```
FT                   /*tag=  t
FT                   /rpt_type= DIRECT
FT                   /note= "repeat 8"
FT   repeat_unit     1823..1843
FT                   /*tag=  u
FT                   /rpt_type= DIRECT
FT                   /note= "repeat 7"
FT   repeat_unit     1865..1883
FT                   /*tag=  v
FT                   /rpt_type= DIRECT
FT                   /note= "repeat 6"
FT   repeat_unit     1896..1914
FT                   /*tag=  w
FT                   /rpt_type= DIRECT
FT                   /note= "repeat 5"
FT   repeat_unit     1927..1945
FT                   /*tag=  x
FT                   /rpt_type= DIRECT
FT                   /note= "repeat 4"
FT   repeat_unit     1958..1975
FT                   /*tag=  y
FT                   /rpt_type= DIRECT
FT                   /note= "repeat 3"
FT   repeat_unit     1988..2006
FT                   /*tag=  z
FT                   /rpt_type= DIRECT
FT                   /note= "repeat 2"
FT   repeat_unit     2019..2037
FT                   /*tag=  aa
FT                   /rpt_type= DIRECT
FT                   /note= "repeat 1"
FT   repeat_unit     816..837
FT                   /*tag=  ab
FT                   /rpt_type= DIRECT
FT                   /note= "repeat C"
FT   repeat_unit     1193..1212
FT                   /*tag=  ac
FT                   /rpt_type= DIRECT
FT                   /note= "repeat C"
FT   repeat_unit     812..827
FT                   /*tag=  ad
FT                   /rpt_type= DIRECT
FT                   /note= "repeat B"
FT   repeat_unit     1200..1220
FT                   /*tag=  ae
FT                   /rpt_type= DIRECT
FT                   /note= "repeat B"
FT   repeat_unit     1124..1142
FT                   /*tag=  af
FT                   /rpt_type= DIRECT
FT                   /note= "repeat D"
```



```
FT   repeat_unit     1209..1226
FT                   /*tag=  ag
FT                   /rpt_type= DIRECT
FT                   /note= "repeat D"
FT   repeat_unit     1538..1553
FT                   /*tag=  ah
FT                   /rpt_type= DIRECT
FT                   /note= "repeat E"
FT   repeat_unit     1785..1804
FT                   /*tag=  ai
FT                   /rpt_type= DIRECT
FT                   /note= "repeat E"
FT   repeat_unit     2256..2275
FT                   /*tag=  aj
FT                   /rpt_type= DIRECT
FT                   /note= "repeat F"
FT   repeat_unit     2277..2297
FT                   /*tag=  ak
FT                   /rpt_type= DIRECT
FT                   /note= "repeat F"
FT   rbs             880..882
FT                   /*tag=  al
FT                   /standard_name= Shine-Dalgarno
FT   misc_difference 882..885
FT                   /*tag=  am
FT                   /note= "sequence shown in Fig.10 has 5 adenine
FT                   residues here"
PN   J08051984-A.
PD   27-FEB-1996.
PF   19-MAY-1995; 144172.
PR   20-MAY-1994; US-246403.
PA   (KAJI/) KAJI A.
DR   WPI; 96-174567/18.
DR   P-PSDB; R88917, W02651.
PT   Plasmid pOKcop comprising mutant repA coding sequence - maintained
PT   at high copy number in E.coli host cells and useful for recombinant
PT   prodn. of proteins
PS   Claim 3; Page 33-35; 51pp; Japanese.
CC   The present sequence is from modified P1 plasmid which codes for
CC   RepA and IncC proteins. Plasmids with modified repA genes can be
CC   maintained stably at higher copy numbers than plasmids with
CC   unmodified repA genes.
SQ   Sequence   2316 BP;    646 A;    525 C;    534 G;    611 T;

  Query Match          10.2%;  Score 15;  DB 21;  Length 2316;
  Best Local Similarity 100.0%;  Pred. No. 1.94e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2245 tacattattatttta 2259
          |||||||||||||||
Cp    143 TACATTATTATTTA 129
```

```
RESULT   39
ID   Q66637 standard; cDNA; 2333 BP.
AC   Q66637;
DT   19-JAN-1995  (first entry)
DE   Human Activin receptor-like kinase 4 (hALK-4) cDNA.
KW   serine threonine kinases; activin receptors; Act-R; superfamily;
KW   transforming growth factor; TGF; diagnostics; detection; therapy;
KW   rheumatoid arthritis; glomerular nephritis; fibrosis; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             1..1518
FT                   /*tag= a
FT                   /product= activin receptor-like kinase 4
FT                   /note= "truncated version"
FT   3'utr           1519..2333
FT                   /*tag= b
FT                   /note= "no poly A tail, suggests cDNA internally
FT                   primed"
PN   WO9411502-A.
PD   26-MAY-1994.
PF   17-NOV-1993; G02367.
PR   17-NOV-1992; GB-024057.
PR   08-MAR-1993; GB-004677.
PR   08-MAR-1993; GB-004680.
PR   28-MAY-1993; GB-011047.
PR   02-JUL-1993; GB-013763.
PR   03-AUG-1993; GB-016099.
PR   15-OCT-1993; GB-021344.
PA   (LUDW-) LUDWIG INST CANCER RES.
PI   Dijke P,  Franzen P,  Heldin C,  Miyazono K,  Yamashita H;
DR   WPI; 94-183503/22.
DR   P-PSDB; R55369.
PT   Activin receptor-like kinase(s) with serine/threonine kinase
PT   domains - have activin/TGF beta-type I receptor function and can
PT   be used in diagnosis or therapy of rheumatoid arthritis,
PT   glomerular nephritis, fibrosis, etc.
PS   Claim 21; Page 53-55; 97pp; English.
CC   The inventors have identified a new family of receptor kinases
CC   called activin receptor-like kinases (ALK).  Their discovery was
CC   based on the realisation that receptor serine/threonine kinases
CC   form a new receptor family, which may include the type II receptors
CC   for other proteins in the transforming growth factor(TGF) beta
CC   superfamily. The activin receptor type II sequences from mouse and
CC   the daf1 gene product of C.elegans have high sequence similarity
CC   and were used to design degenerate primers to clone related cDNA's
CC   (see Q66643-49). Six distinct putative receptor serine/threonine
CC   kinases (ALK 1-6, of which human ALK 1-5 are Q66634-38 respectively,
CC   and the mouse ALK-1,3,4,and 6 are Q66639-42 respectively)
CC   were identified. Human ALK 4 cDNA was identified as having an ORF
```



```
CC   encoding  a truncated extracellular domain, the complete extracellular
CC   domain was obtained by RT-PCR using primers Q66648-49. Products of
CC   the invention can be used in therapy, eg. to modulate conditions
CC   associated with activin or TGF beta activity. These conditions include
CC   fibrosis, eg. liver cirrhosis and pulmonary fibrosis, cancer,
CC   rheumatoid arthritis and glomeronephritis.
SQ   Sequence   2333 BP;    469 A;    649 C;    667 G;    548 T;

  Query Match          10.2%;  Score 15;  DB 11;  Length 2333;
  Best Local Similarity 100.0%;  Pred. No. 1.94e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    591 ctttgtccagcgcgcac 605
          ||||||||||||||||
Qy     46 CTTTGTCCAGCGGCAC 60


RESULT   40
ID   N90703 standard; DNA; 2662 BP.
AC   N90703;
DT   25-JAN-1990  (first entry)
DE   Rhoptry membrane antigen-1 gene.
KW   Malaria; rhoptry membrane antigen-1; antibodies.
OS   Plasmodium falciparum D10.
FH   Key             Location/Qualifiers
FT   cds               CDS             634..2202
FT                     /*tag=   a
PN   WO8907645-A.
PD   24-AUG-1989.
PF   10-FEB-1989; AU0056.
PR   12-FEB-1988; AU-006743.
PA   (SARA) Saramane Pty Ltd.
PI   Peterson MG, Crewther PE, Smythe JA, Marshall VM, Silva A;
DR   WPI; 89-263714/36.
DR   P-PSDB; P91632.
PT   Rhoptry membrane antigen of Plasmodium falciparum
PT   - used for producing antibodies and in immunisation,
PT   diagnostic and treatment methods for malaria.
PS   Claim 8; fig 1; 46pp; English.
CC   The RMA-1 gene can generate an immune response to malaria, and
CC   antibodies which can inhibit growth of the parasite. RMA-1 initially
CC   has mol. wt. 80 kD.
SQ   Sequence   2662 BP;   1134 A;    307 C;    387 G;    834 T;

  Query Match          10.2%;  Score 15;  DB 1;  Length 2662;
  Best Local Similarity 100.0%;  Pred. No. 1.94e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    253 ttatttaatataaat 267
          ||||||||||||||||
Cp    136 TTATTTTAATATAAT 122
```



```
RESULT   41
ID   Q55148 standard; DNA; 2812 BP.
AC   Q55148;
DT   25-SEP-1995  (first entry)
DE   Pseudomonas aeruginosa probe P4-5.
KW   Probe; S.aureus; S.epidermis; E.faecalis; P.aeruginosa; E.coli;
KW   K.pneumoniae; E.cloacae; clinical sample; ds.
OS   Pseudomonas aeruginosa.
PN   WO9401583-A.
PD   20-JAN-1994.
PF   07-JUL-1993; J00936.
PR   07-JUL-1992; JP-179719.
PA   (FUSO ) FUSO PHARM IND LTD.
PA   (OHNO/) OHNO T.
PI   Eda S,  Matsuhisa A,  Ohno T,  Uehara H;
DR   WPI; 94-035086/04.
PT   Probe for identifying bacteria causing infectious disease -
PT   consists of a DNA fragment obtained by HindIII cleavage of the
PT   pathogenic bacterial genomic DNA
PS   Claim 6; Page 60-61; 100pp; Japanese.
CC   The nucleotide sequence of a 2812 bp probe obtained by digestion of
CC   Pseudomonas aeruginosa genomic DNA with the restriction enzyme HindIII.
CC   The probes (Q55133-48 and Q55752-58) represent other probes derived,
CC   by HindIII digestion of the genomes of S.aureus, S.epidermis,
CC   Enterococcus faecalis, Pseudomonas aeruginosa, E.coli, Klebsiella
CC   pneumoniae or Enterobacter cloacae.  The probes can be used to detect
CC   their respective microorganisms in clinical samples.
SQ   Sequence   2812 BP;    695 A;    577 C;    756 G;    784 T;

  Query Match          10.2%;  Score 15;  DB 14;  Length 2812;
  Best Local Similarity 100.0%;  Pred. No. 1.94e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2088 aggccagccctggcc 2102
          ||||||||||||||
Qy     77 AGGCCAGCCCTGGCC 91


RESULT   42
ID   T18535 standard; DNA; 3315 BP.
AC   T18535;
DT   05-JUL-1996  (first entry)
DE   Receptor tyrosine kinase HER4 gene (5' sequence truncation).
KW   HER4/p180erbB4; HER4; receptor tyrosine kinase;
KW   epidermal growth factor receptor; cancer; diagnosis; therapy; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             156..1781
FT                   /*tag= a
```



```
PN   WO9612019-A2.
PD   25-APR-1996.
PF   10-OCT-1995; U13524.
PR   14-OCT-1994; US-323442.
PA   (BRIM ) BRISTOL-MYERS SQUIBB CO.
PI   Culouscou J,  Hellstrom I,  Hellstrom KE,  Plowman GD;
PI   Shoyab M,  Siegall C;
DR   WPI; 96-222005/22.
DR   P-PSDB; R91736.
PT   DNA encoding receptor tyrosine kinase, HER4 - related to human
PT   epidermal growth factor receptor, used for diagnosis and therapy of
PT   human cancers
PS   Disclosure; Page 129-132; 203pp; English.
CC   A DNA sequence (T18535) isolated from a human breast MDA-MB-453
CC   includes the 3' portion of a novel prototype receptor tyrosine
CC   kinase HER4 sequence (T18532); nucleotide 156 aligns with position
CC   2333 of the prototype.  The first 155 nucleotides differ from HER4
CC   and may represent the 5' untranslated region of a transcript derived
CC   from a cryptic promoter within an intfon of the HER4 gene.  The
CC   product (R91736) of the truncated sequence lacks the N-terminal
CC   767 amino acids of prototype HER4 (R91733).
SQ   Sequence   3315 BP;   1004 A;    684 C;    702 G;    925 T;

  Query Match          10.2%;  Score 15;  DB 19;  Length 3315;
  Best Local Similarity 100.0%;  Pred. No. 1.94e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2661 tttttctccataaat 2675
          ||||||||||||||
Cp     38 TTTTTCTCCATAAAT 24


RESULT   43
ID   Q64897 standard; cDNA; 3321 BP.
AC   Q64897;
DT   11-JAN-1995  (first entry)
DE   HER4 N-terminal truncated with AP domain.
KW   Human epidermal growth factor receptor; HER; EGF; tyrosine kinase;
KW   cancer; neuronal tissue; muscle tissue; neoplasm;
KW   carcinoma; primer; probe; PCR; autophosphorylation domain; AP; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             156..1781
FT                   /*tag= a
PN   EP-599274-A.
PD   01-JUN-1994.
PF   23-NOV-1993; 118837.
PR   24-NOV-1992; US-981165.
PA   (BRIM ) BRISTOL-MYERS SQUIBB CO.
PI   Culouscou J,  Plowman GD,  Shoyab M;
DR   WPI; 94-169599/21.
```

```
DR   P-PSDB; R54842.
PT   New recombinant nucleic acid expressing HER4 - a new receptor
PT   tyrosine kinase expressed in some cancer cells, and related
PT   vectors, antibodies, ligands etc, for diagnosis and treatment of
PT   cancers
PS   Claim 1; Fig 2B; 104pp; English.
CC   This sequence contains the 3'-portion of the HER4 sequence where
CC   nucleotide position 156 of the truncated sequence aligns with
CC   position 2335 of the complete HER4 sequence (Q64896) (just
CC   downstream from the region encoding the ATP-binding site of the HER4
CC   kinase).  The first 155 nucleotides of the truncated sequence are
CC   unique from HER4 and may represent the 5'-UTR of a transcript
CC   derived from a cryptic promoter within an intron of the HER4 gene.
CC   HER4 is the fourth member of the EGFR-family of tyrosine kinases and
CC   is expressed in some human cancers and in some tissues of neuronal
CC   or muscle origin.  HER4 polynucleotides, opt. labelled, are useful
CC   in assays (e.g. of HER4 mRNA to detect certain neoplasms, esp.
CC   breast carcinoma) and as primers in PCR or as probes.
SQ   Sequence   3321 BP;    1006 A;    685 C;    705 G;    925 T;

  Query Match           10.2%;  Score 15;  DB 11;  Length 3321;
  Best Local Similarity 100.0%;  Pred. No. 1.94e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2667 tttttctccataaaat 2681
          ||||||||||||||||
Cp     38 TTTTTCTCCATAAAT 24


RESULT   44
ID   Q15130 standard; DNA; 3360 BP.
AC   Q15130;
DT   16-MAR-1992  (first entry)
DE   DNA contg. Sau3AI methylase and endonuclease genes.
KW   Recombinant; ss.
OS   Staphylococcus carnosus.
FH   Key             Location/Qualifiers
FT   cds             448..1917
FT                   /*tag= a
FT                   /note= "Sau3AI endonuclease"
FT   cds             2017..3255
FT                   /*tag= b
FT                   /note= "Sau3AI methylase"
FT   misc_feature    448..450
FT                   /*tag= c
FT                   /codon= TTG
FT                   /note= "amino acid encoded = Met"
PN   EP-460673-A.
PD   11-DEC-1991.
PF   06-JUN-1991; 109302.
PR   08-JUN-1990; DE-018441.
```



```
PA   (BOEF ) BOEHRINGER MANNHEIM GMBH.
PI   Gotz F, Seeber S;
DR   WPI; 91-363167/50.
DR   P-PSDB; R15503, R15647.
PT   New DNA encoding Sau3AI methylase and endonuclease enzymes -
PT   useful in recombinant DNA technology, and corresp. recombinant
PT   vectors and transformed microorganisms
PS   Claim 1; Page 12; 24pp; German.
CC   The DNA sequence contains coding regions for a protein (P1) with
CC   Sau3AI methylase activity and a protein (P2) with Sau3AI endonuclease
CC   activity. Sau3AI endonuclease is much used in recombinant DNA
CC   technology and can now be prepd. more simply using non-pathogenic
CC   cells. Cloning of the Sau3AI restriction/methylase system is now
CC   possible by using a host organism closely related to Staphylococcus
CC   aureus 3A but lacking the system itself.
SQ   Sequence   3360 BP;    1296 A;    398 C;    634 G;    1032 T;

  Query Match           10.2%;  Score 15;  DB 3;  Length 3360;
  Best Local Similarity 100.0%;  Pred. No. 1.94e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   3344 tattattttaatata 3358
          ||||||||||||||||
Cp    138 TATTATTTTAATATA 124


RESULT   45
ID   Q73652 standard; DNA; 3836 BP.
AC   Q73652;
DT   21-JUN-1995  (first entry)
DE   HSV-2 strain 25766 in the region of the gH glycoprotein.
KW   Herpes simplex; lesions; humoral; cell mediated; immunity; ss.
OS   Herpes simplex virus type 2.
FH   Key             Location/Qualifiers
FT   cds             720..3233
FT                   /*tag= a
PN   WO9421807-A.
PD   29-SEP-1994.
PF   21-MAR-1994; G00572.
PR   19-MAR-1993; GB-005710.
PR   06-DEC-1993; GB-024964.
PA   (CANT-) CANTAB PHARM RES LTD.
PI   Boursnell MEG,  Inglis SC,  Minson AC;
DR   WPI; 94-317027/39.
DR   P-PSDB; R60763.
PT   Pharmaceutical contg. non-retroviral virus unable to produce
PT   infectious particles - inducing humoral and cellular immune
PT   responses for protection or treatment, partic. of herpes simplex
PT   infection
PS   Disclosure; Fig 18; 96pp; English.
CC   The sequence is that of the Herpes simplex virus type 2 glycoprotein
```



```
CC   gH gene.  Clones defective in this gene can be used to treat or
CC   prevent HSV infection, and has the advantage of inducing both a
CC   humoral and cell mediated response.
CC   See also Q73653-9.
SQ   Sequence   3836 BP;    544 A;    1377 C;    1255 G;    651 T;

  Query Match           10.2%;  Score 15;  DB 13;  Length 3836;
  Best Local Similarity 100.0%;  Pred. No. 1.94e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    491 ggggccagggctggc 505
          ||||||||||||||||
Cp     93 GGGGCCAGGGCTGGC 79


RESULT   46
ID   Q74062 standard; DNA; 4100 BP.
AC   Q74062;
DT   29-JAN-1996  (first entry)
DE   The rat beta-actin gene.
KW   Beta-actin; primer; mRNA; specificity; pharmaceutical; ss.
OS   Rattus rattus.
PN   J07123984-A.
PD   16-MAY-1995.
PF   05-NOV-1993; 275852.
PR   05-NOV-1992; JP-275852.
PA   (HITB ) HITACHI CHEM CO LTD.
DR   WPI; 95-211627/28.
PT   A primer for the detection and the determn. of a specific messenger
PT   RNA - can detect and determine specific mRNA(s) with high
PT   reliability
PS   Example 27; Page 28-29; 35pp; Japanese.
CC   Q74062 is the rat beta-actin gene. This gene is amplified by the
CC   primers Q74045 and Q74046. The primers are used specifically for the
CC   detection and isolation of this sequence. They have the advantage of
CC   high sensitivity and reliability and are useful in the pharmaceutical
CC   industry.
SQ   Sequence   4100 BP;    765 A;    1111 C;    1126 G;    1097 T;

  Query Match           10.2%;  Score 15;  DB 16;  Length 4100;
  Best Local Similarity 100.0%;  Pred. No. 1.94e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   3743 gccagggctggcctg 3757
          ||||||||||||||||
Cp     90 GCCAGGGCTGGCCTG 76


RESULT   47
ID   Q99462 standard; DNA; 5198 BP.
AC   Q99462;
```



```
DT   30-DEC-1995  (first entry)
DE   Maize HM1 genomic DNA.
KW   HM1 gene; Cochliobolus carbonum Nelson race 1; fungus;
KW   fungal disease toxin; resistance; ss.
OS   Zea mays.
PN   WO9507989-A.
PD   23-MAR-1995.
PF   19-SEP-1994; U10497.
PR   17-SEP-1993; US-123761.
PA   (PION-) PIONEER HI-BRED INT INC.
PI   Briggs SP,  Johal GS;
DR   WPI; 95-131357/17.
PT   Disease resistance gene Hm1 from maize - is used to confer
PT   resistance to Cochliobolus carbonum.
PS   Claim 2; Page 20-22; 26pp; English.
CC   A DNA sequence is claimed having 90% translational homology to
CC   genomic or cDNA sequence of maize HM1 gene (Q99462 & Q99463). The
CC   gene HM1 controls resistance to Cochliobolus carbonum Nelson race 1.
CC   The HM1 gene can be used in conjunction with HC-toxin in a
CC   selectable marker system for use in maize transformation.
CC   This gene can also be inserted into the genomes of maize varieties
CC   which lack the gene to produce transformants that are resistance
CC   to disease caused by C. carbonum. The MH1 resistance gene is one of
CC   a family of homologous disease resistance genes in maize and other
CC   crops. The invention encompasses methods for identification of
CC   homologous disease resistance genes by using this gene or a
CC   fragment of it as a probe.
SQ   Sequence   5198 BP;    1375 A;    1159 C;    1136 G;    1528 T;

  Query Match           10.2%;  Score 15;  DB 15;  Length 5198;
  Best Local Similarity 100.0%;  Pred. No. 1.94e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1011 tatttatggagaaaa 1025
          ||||||||||||||||
Qy     23 TATTTATGGAGAAAA 37


RESULT   48
ID   Q68433 standard; DNA; 5198 BP.
AC   Q68433;
DT   26-JAN-1995  (first entry)
DE   Maize Hm1 gene.
KW   Hm1 gene; HC toxin; race-specific resistance; disease-resistance;
KW   Cochliobolus carbonum; selectable marker; transformation;
KW   transposon tagging; ds.
OS   Zea mays.
PN   WO9413825-A.
PD   23-JUN-1994.
PF   14-DEC-1993; U12146.
PR   15-DEC-1992; US-995658.
```

```
PA   (PION-) PIONEER HI-BRED INT INC.
PI   Briggs SP,  Johal GS;
DR   WPI; 94-217898/26.
PT   Hm1 gene conferring race-resistance to Cochliobolus carbonum to
PT   maize - for use as a selectable marker for transformed maize
PT   cells
PS   Disclosure; Page 14-16; 19pp; English.
CC   Transposon mutagenesis was used to tag, clone and characterize the
CC   maize Hm1 gene.  Genomic and cDNA sequences of the Hm1 gene are
CC   provided in Q68433 and Q68434, respectively.
SQ   Sequence   5198 BP;    1376 A;    1160 C;    1135 G;    1527 T;

  Query Match           10.2%;  Score 15;  DB 11;  Length 5198;
  Best Local Similarity 100.0%;  Pred. No. 1.94e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1011 tatttatggagaaaa 1025
          ||||||||||||||||
Qy     23 TATTTATGGAGAAAA 37


RESULT   49
ID   Q64896 standard; cDNA; 5501 BP.
AC   Q64896;
DT   11-JAN-1995  (first entry)
DE   HER4.
KW   Human epidermal growth factor receptor; HER; EGF; tyrosine kinase;
KW   cancer; neuronal tissue; muscle tissue; neoplasm;
KW   carcinoma; primer; probe; PCR; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             34..3960
FT                   /*tag= a
PN   EP-599274-A.
PD   01-JUN-1994.
PF   23-NOV-1993; 118837.
PR   24-NOV-1992; US-981165.
PA   (BRIM ) BRISTOL-MYERS SQUIBB CO.
PI   Culouscou J,  Plowman GD,  Shoyab M;
DR   WPI; 94-169599/21.
DR   P-PSDB; R54841.
PT   New recombinant nucleic acid expressing HER4 - a new receptor
PT   tyrosine kinase expressed in some cancer cells, and related
PT   vectors, antibodies, ligands etc, for diagnosis and treatment of
PT   cancers
PS   Claim 4; Fig 1; 104pp; English.
CC   HER4 is the fourth member of the EGFR-family of tyrosine kinases and
CC   is expressed in some human cancers and in some tissues of neuronal
CC   or muscle origin.  HER4 polynucleotides, opt. labelled, are useful
CC   in assays (e.g. of HER4 mRNA to detect certain neoplasms, esp.
CC   breast carcinoma) and as primers in PCR or as probes.
```



```
SQ   Sequence   5501 BP;    1605 A;    1173 C;    1230 G;    1493 T;

  Query Match           10.2%;  Score 15;  DB 11;  Length 5501;
  Best Local Similarity 100.0%;  Pred. No. 1.94e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   4846 tttttctccataaat 4860
          ||||||||||||||||
Cp     38 TTTTTCTCCATAAAT 24


RESULT   50
ID   T18532 standard; DNA; 5501 BP.
AC   T18532;
DT   05-JUL-1996  (first entry)
DE   Receptor tyrosine kinase HER4 gene.
KW   HER4/p180erbB4; HER4; receptor tyrosine kinase;
KW   epidermal growth factor receptor; cancer; diagnosis; therapy; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             34..3960
FT                   /*tag= a
FT   signal_peptide  34..108
FT                   /*tag= b
FT   mat_peptide     109..3957
FT                   /*tag= c
PN   WO9612019-A2.
PD   25-APR-1996.
PF   10-OCT-1995; U13524.
PR   14-OCT-1994; US-323442.
PA   (BRIM ) BRISTOL-MYERS SQUIBB CO.
PI   Culouscou J,  Hellstrom I,  Hellstrom KE,  Plowman GD;
PI   Shoyab M,  Siegall C;
DR   WPI; 96-222005/22.
DR   P-PSDB; R91733.
PT   DNA encoding receptor tyrosine kinase, HER4 - related to human
PT   epidermal growth factor receptor, used for diagnosis and therapy of
PT   human cancers
PS   Claim 1; Page 111-116; 203pp; English.
CC   A DNA sequence (T18532) codes for a novel receptor tyrosine kinase,
CC   designated HER4/p180erbB4 or HER4 (R91733), which is related to
CC   epidermal growth factor receptor and which is expressed in several
CC   human cancers and in certain tissues of neuronal and muscular origin.
CC   It was obtd. by assembly of cDNA clones that had been obtd. from
CC   human breast MDA-MB-453 and human heart cell libraries by PCR
CC   amplification using primers (see T18536-42) based on conserved exons
CC   from kinase domains of HER2, HER3, EGFR and Xmrk.  Related sequences
CC   (see T18533 and T18535) were isolated from the MDA-MB-453 library.
CC   HER4 polynucleotides can be used in the prepn. of HER4 expression
CC   vectors, primers and probes to detect and/or clone HER4, and
CC   diagnostic reagents.
```



```
SQ   Sequence   5501 BP;    1606 A;    1173 C;    1229 G;    1493 T;

  Query Match           10.2%;  Score 15;  DB 19;  Length 5501;
  Best Local Similarity 100.0%;  Pred. No. 1.94e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   4846 tttttctccataaat 4860
          ||||||||||||||||
Cp     38 TTTTTCTCCATAAAT 24


Search completed: Tue Dec 22 20:06:01 1998
Job time : 148 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Tue Dec 22 19:48:32 1998; MasPar time 248.67 Seconds
1302.740 Million cell updates/sec

Tabular output not generated.

Title: >US-08-530-112A-709
Description: (1-147) from US08530112A.seq
Perfect Score: 147
N.A. Sequence: 1 GATCCCCAGGGTTTCTTTGT.....TTAAAATAATAATGTACAAA 147
Comp: CTAGGGGTCCCAAAGAAACA.....AATTTTATTATTACATGTTT

Scoring table: TABLE jmetric
Gap 60

Nmatch STD : Dbase 0; Query 0

Searched: 567134 seqs, 1101898692 bases x 2

Post-processing: Minimum Match 0%
Listing first 1000 summaries

Database: emb155
1:em_ba 2:em_fun 3:em_htg 4:em_huml 5:em_hum2 6:em_in
7:em_om 8:em_or 9:em_ov 10:em_pat 11:em_ph 12:em_pl
13:em_ro 14:em_vi
Database: genbank107
15:gb_ba1 16:gb_ba2 17:gb_htg 18:gb_in 19:gb_om 20:gb_ov
21:gb_pat 22:gb_ph 23:gb_pl1 24:gb_pl2 25:gb_pri
26:gb_pr2 27:gb_pr3 28:gb_ro 29:gb_st 30:gb_sts 31:gb_sy
32:gb_un 33:gb_vi

Statistics: Mean 7.787; Variance 2.896; scale 2.689

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
c	1	18	12.2	310	18	AMU32452	Apodemia mormo NADH de	3.50e+00
c	2	18	12.2	311	18	BPU32454	Battus philenor NADH d	3.50e+00
c	3	18	12.2	311	18	HMU25877	Hemileuca maia NADH de	3.50e+00
c	4	18	12.2	311	18	ISU32461	Incisalia sp. NADH deh	3.50e+00
c	5	18	12.2	311	18	PCU32464	Parnassius clodius NAD	3.50e+00
c	6	18	12.2	311	18	ISU32462	Ithomia sp. NADH dehyd	3.50e+00
c	7	18	12.2	311	18	EPU32458	Enodia portlandia NADH	3.50e+00
c	8	18	12.2	311	18	AAU32451	Anaea andria NADH dehy	3.50e+00
c	9	18	12.2	311	18	DSU25876	Diatraea saccharalis N	3.50e+00
c	10	18	12.2	311	18	ACU25881	Agrius cingulata NADH	3.50e+00
c	11	18	12.2	311	18	HCU25873	Hyalophora cecropia NA	3.50e+00
c	12	18	12.2	311	18	PCU25880	Pyrgus communis NADH d	3.50e+00
c	13	18	12.2	731	21	A30910	D.discoideum DNA for r	3.50e+00
c	14	18	12.2	731	18	DDDDRAS	D.discoideum Dd ras ge	3.50e+00
	15	18	12.2	1119	23	S46989	chalcone synthase [soy	3.50e+00
c	16	18	12.2	1676	18	DDRASX	D.discoideum ras gene.	3.50e+00
c	17	18	12.2	2256	18	SPFMTTGRG	Spodoptera frugiperda	3.50e+00
	18	18	12.2	2688	23	SOYCHS7A	Glycine max chalcone s	3.50e+00
	19	18	12.2	6393	18	CEC49G9	Caenorhabditis elegans	3.50e+00
c	20	18	12.2	40267	25	HSU159B9	Human DNA sequence fro	3.50e+00
c	21	18	12.2	46255	18	CELC28H8	Caenorhabditis elegans	3.50e+00
	22	18	12.2	117864	17	CEY38G4	Caenorhabditis elegans	3.50e+00
c	23	18	12.2	143092	18	CEY17G7B	Caenorhabditis elegans	3.50e+00
	24	18	12.2	176011	17	AC005073	*** SEQUENCING IN PROG	3.50e+00
	25	18	12.2	200167	26	AC002480	Human BAC clone GS055X	3.50e+00
c	26	18	12.2	207139	17	CEY17G7	Caenorhabditis elegans	3.50e+00
	27	18	12.2	226063	27	AC005406	, complete sequence.	3.50e+00
	28	18	12.2	256404	17	AC005071	*** SEQUENCING IN PROG	3.50e+00
	29	17	11.6	338	19	CFDOPRX01	C.familiaris mRNA DOPC	1.71e+01
c	30	17	11.6	405	18	PARMTDI7B	paramecium species 7,3	1.71e+01
c	31	17	11.6	434	25	S52967S4	alpha-(1.3/1,4) fucosy	1.71e+01
	32	17	11.6	769	25	HSMETTRAN	H.sapiens mRNA for met	1.71e+01
	33	17	11.6	791	21	E02953	cDNA sequence encoding	1.71e+01
c	34	17	11.6	915	27	HSU27332	Human alpha (1,3) fuco	1.71e+01
	35	17	11.6	952	19	OCU04276	Oryctolagus cuniculus	1.71e+01
c	36	17	11.6	1212	26	PTY14035	Pan troglodytes FUT6 g	1.71e+01
c	37	17	11.6	1287	25	HUMFUT41A	Homo sapiens galactosi	1.71e+01
c	38	17	11.6	1468	20	XELXP42	X.laevis myelin basic	1.71e+01
c	39	17	11.6	1654	21	I34194	Sequence 13 from paten	1.71e+01
c	40	17	11.6	2145	27	HSU27336	Human alpha (1,3) fuco	1.71e+01
c	41	17	11.6	2270	27	HSU27334	Human alpha (1,3) fuco	1.71e+01
c	42	17	11.6	2612	27	HSU27337	Human alpha (1,3) fuco	1.71e+01

	43	17	11.6	2682	23	ATHCRBAA	A.thaliana 12S storage	1.71e+01
	44	17	11.6	2682	23	ATCRB	Arabidopsis CRB gene f	1.71e+01
c	45	17	11.6	2973	21	A36402	Sequence 1 from Patent	1.71e+01
c	46	17	11.6	2997	20	XLMPK1G	X.laevis MPK1 mRNA for	1.71e+01
c	47	17	11.6	3002	27	HSU27335	Human alpha (1,3) fuco	1.71e+01
c	48	17	11.6	3085	18	DDC60169	Dictyostelium discoide	1.71e+01
	49	17	11.6	3110	26	HUMDIHM	Human gene for myelope	1.71e+01
c	50	17	11.6	3127	27	HSU27333	Human alpha (1,3) fuco	1.71e+01
c	51	17	11.6	3554	23	SCYGR259C	S.cerevisiae chromosom	1.71e+01
	52	17	11.6	4560	16	AF052518	Acetobacter xylinus cd	1.71e+01
c	53	17	11.6	5175	23	SCYGR258C	S.cerevisiae chromosom	1.71e+01
c	54	17	11.6	5361	18	PFD16955	Plasmodium falciparum	1.71e+01
c	55	17	11.6	7153	23	PVME1G	P.vulgaris PvME1 gene.	1.71e+01
c	56	17	11.6	7876	31	SYNAGABA	Synthetic Gossypium hi	1.71e+01
c	57	17	11.6	11628	23	SCYGRGENS	S.cerevisiae YGR gene	1.71e+01
	58	17	11.6	23822	26	HUMRCP	Homo sapiens gene for	1.71e+01
	59	17	11.6	29468	27	HS431H6	Human DNA sequence ***	1.71e+01
	60	17	11.6	32239	17	AC005717	*** SEQUENCING IN PROG	1.71e+01
	61	17	11.6	34321	18	CE2C247	Caenorhabditis elegans	1.71e+01
	62	17	11.6	38586	18	CELF55C12	Caenorhabditis elegans	1.71e+01
c	63	17	11.6	39147	18	CELC35A11	Caenorhabditis elegans	1.71e+01
c	64	17	11.6	41151	18	CEC48B4	Caenorhabditis elegans	1.71e+01
	65	17	11.6	63711	17	CET02F9	Caenorhabditis elegans	1.71e+01
c	66	17	11.6	73864	17	HS496H19	Human DNA sequence ***	1.71e+01
	67	17	11.6	85372	17	AC005508	*** SEQUENCING IN PROG	1.71e+01
c	68	17	11.6	93762	17	AC004335	*** SEQUENCING IN PROG	1.71e+01
	69	17	11.6	112863	17	HS482L3	Human DNA sequence ***	1.71e+01
c	70	17	11.6	116884	17	PFMAL1P1	Plasmodium falciparum	1.71e+01
c	71	17	11.6	117729	17	AC003027	*** SEQUENCING IN PROG	1.71e+01
	72	17	11.6	118721	17	AC005010	*** SEQUENCING IN PROG	1.71e+01
	73	17	11.6	122940	17	AC004709	*** SEQUENCING IN PROG	1.71e+01
c	74	17	11.6	123193	17	AC005060	*** SEQUENCING IN PROG	1.71e+01
	75	17	11.6	123360	24	ATAC003040	Arabidopsis thaliana c	1.71e+01
	76	17	11.6	127816	17	AC004822	*** SEQUENCING IN PROG	1.71e+01
	77	17	11.6	128136	17	HS889N15	Human DNA sequence ***	1.71e+01
c	78	17	11.6	129169	17	AC005506	*** SEQUENCING IN PROG	1.71e+01
c	79	17	11.6	129252	26	HSU91327	Human chromosome 12p15	1.71e+01
	80	17	11.6	142600	26	AC005371	Homo sapiens chromosom	1.71e+01
	81	17	11.6	147615	17	HS453C12	Human DNA sequence ***	1.71e+01
	82	17	11.6	152184	17	CEY70G10	Caenorhabditis elegans	1.71e+01
c	83	17	11.6	164536	17	HS134N8	Human DNA sequence ***	1.71e+01
c	84	17	11.6	166001	17	AC005225	*** SEQUENCING IN PROG	1.71e+01
	85	17	11.6	175120	26	AC004687	Homo sapiens chromosom	1.71e+01
c	86	17	11.6	176939	26	HS30P20	Human DNA sequence fro	1.71e+01
c	87	17	11.6	180388	25	HUMRETBLAS	Human retinoblastoma s	1.71e+01
	88	17	11.6	197273	26	AC002461	Human BAC clone RG204I	1.71e+01
c	89	17	11.6	198552	23	ATAP22	Arabidopsis thaliana D	1.71e+01
	90	17	11.6	202861	23	ATPCA2	Arabidopsis thaliana D	1.71e+01
c	91	17	11.6	204596	17	AC005069	*** SEQUENCING IN PROG	1.71e+01
	92	17	11.6	230023	17	HS522P13	Human DNA sequence ***	1.71e+01
	93	17	11.6	244254	26	HSAC001228	244Kb Contig from Huma	1.71e+01

	94	16	10.9	220	26	HSY15930	Homo sapiens CTNS gene	7.85e+01
c	95	16	10.9	257	33	AF024358	HIV-1 patient 8 clone	7.85e+01
c	96	16	10.9	311	18	MSU25872	Macrosoma semiermis NA	7.85e+01
c	97	16	10.9	311	18	DPU32457	Danaus plexippus NADH	7.85e+01
	98	16	10.9	355	18	RBU95917	Rhipicentor bicornis 1	7.85e+01
	99	16	10.9	360	25	HSGLUD110	H.sapiens GLUD1 gene f	7.85e+01
c	100	16	10.9	418	15	PSCRSRNAA	Plasmid incI-1 minirep	7.85e+01
c	101	16	10.9	428	18	ACU20866	Angiostrongylus canton	7.85e+01
c	102	16	10.9	446	18	DNYMTRGDA	Mitochondrion Dermanys	7.85e+01
	103	16	10.9	465	15	D89881	Methanococcus thermoli	7.85e+01
	104	16	10.9	594	18	SGATPDP	S.graminum atpD gene e	7.85e+01
	105	16	10.9	865	23	YSCMTCO1	Yeast (S.cerevisiae) m	7.85e+01
	106	16	10.9	865	23	MISC23	Yeast mitochondrial ox	7.85e+01
c	107	16	10.9	924	27	HUMFGFRE	Human fibroblast growt	7.85e+01
c	108	16	10.9	1047	23	DDIUDPGP1	D.discoideum uridine d	7.85e+01
	109	16	10.9	1175	18	EMHDRP	Entamoeba invadens dev	7.85e+01
c	110	16	10.9	1253	25	HUMHTA1DR	Human 5-hydroxytryptam	7.85e+01
c	111	16	10.9	1299	25	HUMHTR1DP	Homo sapiens serotonin	7.85e+01
	112	16	10.9	1445	24	PUECHS	Pueraria lobata mRNA f	7.85e+01
c	113	16	10.9	1469	23	SCYOR251C	S.cerevisiae chromosom	7.85e+01
	114	16	10.9	1484	33	HPU34190	Human papillomavirus t	7.85e+01
	115	16	10.9	1484	33	HPU34188	Human papillomavirus t	7.85e+01
	116	16	10.9	1484	33	HPU34185	Human papillomavirus t	7.85e+01
	117	16	10.9	1484	33	HPU34180	Human papillomavirus t	7.85e+01
	118	16	10.9	1484	33	HPU34182	Human papillomavirus t	7.85e+01
	119	16	10.9	1484	33	HPU34183	Human papillomavirus t	7.85e+01
	120	16	10.9	1484	33	HPU34173	Human papillomavirus t	7.85e+01
	121	16	10.9	1484	33	HPU34193	Human papillomavirus t	7.85e+01
	122	16	10.9	1484	33	HPU34192	Human papillomavirus t	7.85e+01
	123	16	10.9	1484	33	HPU34179	Human papillomavirus t	7.85e+01
	124	16	10.9	1484	33	HPU34178	Human papillomavirus t	7.85e+01
	125	16	10.9	1484	33	HPU34181	Human papillomavirus t	7.85e+01
	126	16	10.9	1484	33	HPU34177	Human papillomavirus t	7.85e+01
	127	16	10.9	1484	33	HPU34172	Human papillomavirus t	7.85e+01
	128	16	10.9	1484	33	HPU34187	Human papillomavirus t	7.85e+01
	129	16	10.9	1484	33	HPU34186	Human papillomavirus t	7.85e+01
	130	16	10.9	1484	33	HPU34175	Human papillomavirus t	7.85e+01
	131	16	10.9	1484	33	HPU34191	Human papillomavirus t	7.85e+01
	132	16	10.9	1484	33	HPU34176	Human papillomavirus t	7.85e+01
	133	16	10.9	1484	33	HPU34184	Human papillomavirus t	7.85e+01
	134	16	10.9	1484	33	HPU34189	Human papillomavirus t	7.85e+01
	135	16	10.9	1484	33	HPU34171	Human papillomavirus t	7.85e+01
	136	16	10.9	1484	33	HPU34170	Human papillomavirus t	7.85e+01
	137	16	10.9	1484	33	HPU34165	Human papillomavirus t	7.85e+01
	138	16	10.9	1484	33	HPU34169	Human papillomavirus t	7.85e+01
	139	16	10.9	1484	33	HPU34168	Human papillomavirus t	7.85e+01
	140	16	10.9	1484	33	HPU34166	Human papillomavirus t	7.85e+01
	141	16	10.9	1484	33	HPU34167	Human papillomavirus t	7.85e+01
	142	16	10.9	1484	33	HPU34174	Human papillomavirus t	7.85e+01
	143	16	10.9	1486	33	AF003031	Human papillomavirus t	7.85e+01
	144	16	10.9	1517	21	I80879	Sequence 2 from patent	7.85e+01

	145	16	10.9	1517	21	I80878	Sequence 1 from patent 7.85e+01
	146	16	10.9	1517	21	I13523	Sequence 2 from patent 7.85e+01
	147	16	10.9	1517	21	I13522	Sequence 1 from patent 7.85e+01
	148	16	10.9	1518	33	AF084952	Human papillomavirus t 7.85e+01
	149	16	10.9	1518	33	AF043287	Human papillomavirus t 7.85e+01
	150	16	10.9	1518	33	AF043286	Human papillomavirus t 7.85e+01
	151	16	10.9	1557	33	AF005730	Marburg virus strain M 7.85e+01
c	152	16	10.9	1578	26	HSPGFR34	H.sapiens PGFR3 gene, 7.85e+01
	153	16	10.9	1814	25	HS285ALU	H.sapiens spcDNA, AluV 7.85e+01
c	154	16	10.9	1818	23	SCYOR250C	S.cerevisiae chromosom 7.85e+01
c	155	16	10.9	1845	15	CIBRCRA	Plasmid Colicin Ib-P9 7.85e+01
	156	16	10.9	1907	15	ECOHEMBE	Escherichia coli 5-ami 7.85e+01
	157	16	10.9	2100	26	AF023676	Homo sapiens lamin B r 7.85e+01
	158	16	10.9	2150	25	HSCAII	Human gene fragment fo 7.85e+01
c	159	16	10.9	2165	23	CHSPERBCL	S.persica chloroplast 7.85e+01
c	160	16	10.9	2202	15	CLOMP	Clostridium perfringen 7.85e+01
	161	16	10.9	2325	26	HSAJ730	Homo sapiens mRNA for 7.85e+01
c	162	16	10.9	2374	25	HUMHDRTPS	Homo sapiens 5-hydroxy 7.85e+01
	163	16	10.9	2416	25	HUMTOPPG1	Human topoisomerase I 7.85e+01
	164	16	10.9	2519	23	YSCAAA	Saccharomyces cerevisi 7.85e+01
	165	16	10.9	2526	21	A46393	Sequence 15 from Paten 7.85e+01
c	166	16	10.9	2568	23	SCYLR120C	S.cerevisiae chromosom 7.85e+01
	167	16	10.9	2674	15	CLOSERTRNA	Clostridium perfringen 7.85e+01
c	168	16	10.9	2733	23	SCYLR121C	S.cerevisiae chromosom 7.85e+01
c	169	16	10.9	2839	33	HIVU85918	HIV-1 strain FRMNR clo 7.85e+01
c	170	16	10.9	2863	28	RNTHYCSG	Rat thy-1 gene for cel 7.85e+01
c	171	16	10.9	2863	28	RNTHY1G	Rat gene for Thy-1 ant 7.85e+01
c	172	16	10.9	2884	33	HIVU85913	HIV-1 strain FRCNP clo 7.85e+01
	173	16	10.9	2912	23	HRAHRPCA	A.rusticana peroxidase 7.85e+01
	174	16	10.9	2920	33	HPU37217	Human papillomavirus t 7.85e+01
	175	16	10.9	2925	25	HUMCAIIA01	H.sapiens carbonic anh 7.85e+01
	176	16	10.9	3260	19	OAR16533	Ovis aries IGF-II gene 7.85e+01
	177	16	10.9	3268	28	MMJ225124	Mus musculus mRNA for 7.85e+01
	178	16	10.9	3302	33	REOBRVP1	Bovine rotavirus mRNA 7.85e+01
	179	16	10.9	3600	27	HSU56402	Human chromatin struct 7.85e+01
	180	16	10.9	3698	26	HSSUPT5H	H.sapiens mRNA for sup 7.85e+01
	181	16	10.9	3712	26	AB000516	Homo sapiens mRNA for 7.85e+01
c	182	16	10.9	3877	21	E01662	cDNA encoding human ba 7.85e+01
c	183	16	10.9	3877	25	HUMFGFB	Human basic fibroblast 7.85e+01
c	184	16	10.9	3927	23	MIWMCOX	W.makrii mitochondrial 7.85e+01
	185	16	10.9	4080	19	OAINIGFII3	Ovis aries insulin-lik 7.85e+01
c	186	16	10.9	4346	23	CMU37774	Cucurbita maxima 1-ami 7.85e+01
	187	16	10.9	4348	26	AF048704	Homo sapiens lamin B r 7.85e+01
	188	16	10.9	4550	33	AF001600	Human papillomavirus t 7.85e+01
c	189	16	10.9	4690	23	PVU34754	Phaseolus vulgaris cel 7.85e+01
c	190	16	10.9	4860	18	CEVF39H2L	Caenorhabditis elegans 7.85e+01
c	191	16	10.9	6757	25	HUMGFB	Human basic fibroblast 7.85e+01
	192	16	10.9	7305	23	PTGWIN62B	P.trichocarpa chitinas 7.85e+01
	193	16	10.9	7348	33	PAPPH63	Human papillomavirus t 7.85e+01
	194	16	10.9	7904	21	I70143	Sequence 2 from patent 7.85e+01
	195	16	10.9	7904	21	I70142	Sequence 1 from patent 7.85e+01

	196	16	10.9	7904	21	I72345	Sequence 1 from patent 7.85e+01
	197	16	10.9	7904	33	PPH16	Human papillomavirus t 7.85e+01
	198	16	10.9	7905	33	HPU89348	Human papillomavirus t 7.85e+01
	199	16	10.9	9927	15	U67485	Methanococcus jannasch 7.85e+01
	200	16	10.9	10168	23	YSCMTCYO1	Yeast (S.cerevisiae) m 7.85e+01
	201	16	10.9	10168	23	MISC13	oxi3 locus in yeast mi 7.85e+01
c	202	16	10.9	10281	15	ECAE000143	Escherichia coli K-12 7.85e+01
c	203	16	10.9	13251	18	CELEGAP9	Caenorhabditis elegans 7.85e+01
	204	16	10.9	13271	15	AE001168	Borrelia burgdorferi (7.85e+01
c	205	16	10.9	14326	15	U67558	Methanococcus jannasch 7.85e+01
c	206	16	10.9	16048	15	ECOPS	Escherichia coli hemB 7.85e+01
	207	16	10.9	17109	17	CET16B3	*** SEQUENCING IN PROG 7.85e+01
	208	16	10.9	19298	19	SSPPK	S.scrofa ppk98 gene. 7.85e+01
	209	16	10.9	20793	18	CET14D7	Caenorhabditis elegans 7.85e+01
	210	16	10.9	21153	23	YSCMTCG13	Saccharomyces cerevisi 7.85e+01
c	211	16	10.9	21619	26	HSIGFBP	H.sapiens IGFBP-4 gene 7.85e+01
	212	16	10.9	21984	28	MMAJ3128	Mus musculus hapl gene 7.85e+01
	213	16	10.9	22237	18	CELF40B5	Caenorhabditis elegans 7.85e+01
c	214	16	10.9	23223	23	YSCL9233	Saccharomyces cerevisi 7.85e+01
c	215	16	10.9	26521	18	CEW01D2	Caenorhabditis elegans 7.85e+01
c	216	16	10.9	26676	18	AC005134	Drosophila melanogaste 7.85e+01
	217	16	10.9	28762	18	CE2C376	Caenorhabditis elegans 7.85e+01
c	218	16	10.9	28984	26	AB002059	Homo sapiens DNA for H 7.85e+01
c	219	16	10.9	29023	18	CELF21H11	Caenorhabditis elegans 7.85e+01
	220	16	10.9	31111	20	CHMMYHE	Chicken embryonic myos 7.85e+01
	221	16	10.9	31714	17	AC005119	*** SEQUENCING IN PROG 7.85e+01
c	222	16	10.9	31744	18	CEF13A7	Caenorhabditis elegans 7.85e+01
c	223	16	10.9	33319	18	CET10H4	Caenorhabditis elegans 7.85e+01
c	224	16	10.9	35214	18	CELC42D4	Caenorhabditis elegans 7.85e+01
c	225	16	10.9	37639	23	SCCHXIILA	S.cerevisiae DNA from 7.85e+01
c	226	16	10.9	38487	18	CELC01G8	Caenorhabditis elegans 7.85e+01
	227	16	10.9	38744	18	CEC56A3	Caenorhabditis elegans 7.85e+01
c	228	16	10.9	39043	26	AC000031	Homo sapiens Chromosom 7.85e+01
c	229	16	10.9	40172	18	CELC10F3	Caenorhabditis elegans 7.85e+01
c	230	16	10.9	40643	18	CELC33E10	Caenorhabditis elegans 7.85e+01
c	231	16	10.9	41122	26	HSAC000109	Human Cosmid g0771a222 7.85e+01
	232	16	10.9	41303	26	AC005175	Homo sapiens chromosom 7.85e+01
	233	16	10.9	41680	28	AF049091	Mus musculus H19 and m 7.85e+01
	234	16	10.9	43365	17	AC004396	*** SEQUENCING IN PROG 7.85e+01
c	235	16	10.9	45508	25	HSAC000110	Human Cosmid g0771a233 7.85e+01
	236	16	10.9	45672	18	DMC22E5	Drosophila melanogaste 7.85e+01
	237	16	10.9	45878	24	AB016883	Arabidopsis thaliana g 7.85e+01
c	238	16	10.9	45947	18	CELF46C8	Caenorhabditis elegans 7.85e+01
	239	16	10.9	55359	23	SCD9509	Saccharomyces cerevisi 7.85e+01
c	240	16	10.9	57304	27	AC004802	Homo sapiens 12p13.3 R 7.85e+01
c	241	16	10.9	63495	18	AC004244	Drosophila melanogaste 7.85e+01
	242	16	10.9	71803	17	AC005504	*** SEQUENCING IN PROG 7.85e+01
c	243	16	10.9	72698	23	AB007645	Arabidopsis thaliana g 7.85e+01
c	244	16	10.9	78239	23	AB012241	Arabidopsis thaliana g 7.85e+01
	245	16	10.9	79150	18	AC002513	Drosophila melanogaste 7.85e+01
c	246	16	10.9	79355	23	AB012247	Arabidopsis thaliana g 7.85e+01

	247	16	10.9	79815	26	AC005212	Homo sapiens chromosom 7.85e+01
	248	16	10.9	80686	17	AC005444	*** SEQUENCING IN PROG 7.85e+01
	249	16	10.9	80837	18	AC002514	Drosophila melanogaste 7.85e+01
c	250	16	10.9	81826	25	HSU78045	Human collagenase and 7.85e+01
	251	16	10.9	83922	24	T9A4	Arabidopsis thaliana B 7.85e+01
	252	16	10.9	84132	26	AC005348	Homo sapiens chromosom 7.85e+01
c	253	16	10.9	85139	18	AC001657	Drosophila melanogaste 7.85e+01
	254	16	10.9	87434	23	AB010698	Arabidopsis thaliana g 7.85e+01
c	255	16	10.9	90736	26	HS299D3	Human DNA sequence fro 7.85e+01
	256	16	10.9	91184	17	HS633O19	Human DNA sequence *** 7.85e+01
c	257	16	10.9	92802	17	AC004710	*** SEQUENCING IN PROG 7.85e+01
	258	16	10.9	93008	17	AC005120	*** SEQUENCING IN PROG 7.85e+01
	259	16	10.9	94386	26	HS452M16	Human DNA sequence *** 7.85e+01
	260	16	10.9	95565	26	AC000379	Human Chromosome 15 pa 7.85e+01
c	261	16	10.9	97979	26	AC004739	Homo sapiens PAC clone 7.85e+01
c	262	16	10.9	98029	26	AC004082	Homo sapiens PAC clone 7.85e+01
	263	16	10.9	99383	23	AC000103	Genomic sequence for A 7.85e+01
	264	16	10.9	99512	26	AC004606	Homo sapiens chromosom 7.85e+01
c	265	16	10.9	100000	26	AP000041	Homo sapiens genomic D 7.85e+01
	266	16	10.9	100997	26	HS262D12	Homo sapiens DNA seque 7.85e+01
c	267	16	10.9	101098	26	AC002427	Human BAC clone GS011E 7.85e+01
	268	16	10.9	104386	26	D87009	Human (lambda) DNA for 7.85e+01
c	269	16	10.9	104710	26	AC002488	Homo sapiens chromosom 7.85e+01
c	270	16	10.9	107388	17	HS98Z21	Human DNA sequence *** 7.85e+01
	271	16	10.9	108056	17	AC005170	*** SEQUENCING IN PROG 7.85e+01
c	272	16	10.9	110000	17	CEY53F4	Caenorhabditis elegans 7.85e+01
c	273	16	10.9	112884	17	AC004974	*** SEQUENCING IN PROG 7.85e+01
	274	16	10.9	113916	27	HUAC003007	Human Chromosome 16 BA 7.85e+01
	275	16	10.9	119184	15	D90909	Synechocystis sp. PCC6 7.85e+01
	276	16	10.9	119704	23	OSCHLPLXX	O.sinensis complete ch 7.85e+01
c	277	16	10.9	121166	26	AC004613	Homo sapiens PAC clone 7.85e+01
c	278	16	10.9	122940	17	AC004709	*** SEQUENCING IN PROG 7.85e+01
	279	16	10.9	123521	27	HS75G13	Homo sapiens chromosom 7.85e+01
c	280	16	10.9	124752	17	AC004986	*** SEQUENCING IN PROG 7.85e+01
c	281	16	10.9	124990	26	HS434P1	Human DNA sequence fro 7.85e+01
	282	16	10.9	128169	26	HS849L7	Human DNA sequence fro 7.85e+01
c	283	16	10.9	128824	15	ECU73857	Escherichia coli chrom 7.85e+01
c	284	16	10.9	128888	26	AC003106	Homo sapiens Xp22 GS-5 7.85e+01
c	285	16	10.9	129271	24	AC005687	Arabidopsis thaliana B 7.85e+01
	286	16	10.9	130467	27	HS798A17	Human DNA sequence *** 7.85e+01
c	287	16	10.9	130467	27	HS798A17	Human DNA sequence *** 7.85e+01
c	288	16	10.9	134743	26	HUAC002300	Homo sapiens Chromosom 7.85e+01
	289	16	10.9	136299	26	HS359N14	Human DNA sequence fro 7.85e+01
	290	16	10.9	136328	26	AC004126	Human Chromosome 11q12 7.85e+01
	291	16	10.9	136805	17	AC005507	*** SEQUENCING IN PROG 7.85e+01
c	292	16	10.9	136805	17	AC005507	*** SEQUENCING IN PROG 7.85e+01
	293	16	10.9	137785	26	HS345P10	Human DNA sequence fro 7.85e+01
	294	16	10.9	138706	17	AC004946	*** SEQUENCING IN PROG 7.85e+01
c	295	16	10.9	138965	26	AC003998	Human PAC clone DJ0161 7.85e+01
c	296	16	10.9	139458	17	HS1119A7	Human DNA sequence *** 7.85e+01
	297	16	10.9	140554	28	AC000399	Genomic sequence from 7.85e+01

c	298	16	10.9	140896	17	HS163G9	Human DNA sequence *** 7.85e+01
c	299	16	10.9	146828	17	AC004968	*** SEQUENCING IN PROG 7.85e+01
	300	16	10.9	147104	27	AC002472	Homo sapiens Chromosom 7.85e+01
	301	16	10.9	148600	26	AC002367	Homo sapiens Xp22 PAC 7.85e+01
	302	16	10.9	149963	17	HS466P17	Human DNA sequence *** 7.85e+01
c	303	16	10.9	152798	26	AC004057	Homo sapiens chromosom 7.85e+01
c	304	16	10.9	155022	27	AC002395	Homo sapiens 12q24 PAC 7.85e+01
	305	16	10.9	157248	26	AC002312	Human Chromosome 15 pa 7.85e+01
	306	16	10.9	157412	17	HS376D21	Human DNA sequence *** 7.85e+01
c	307	16	10.9	160531	17	HS127F18	Human DNA sequence *** 7.85e+01
c	308	16	10.9	160986	17	AC005505	*** SEQUENCING IN PROG 7.85e+01
	309	16	10.9	161997	17	AC005026	*** SEQUENCING IN PROG 7.85e+01
	310	16	10.9	162075	26	HS127D3	Homo sapiens DNA seque 7.85e+01
c	311	16	10.9	164214	17	HS163O16	Human DNA sequence *** 7.85e+01
c	312	16	10.9	166715	26	HS309O1	Human genomic DNA sequ 7.85e+01
	313	16	10.9	167101	26	AC002094	Genomic sequence from 7.85e+01
c	314	16	10.9	170000	17	AC005509	*** SEQUENCING IN PROG 7.85e+01
	315	16	10.9	170101	33	ASU18466	African swine fever vi 7.85e+01
	316	16	10.9	173275	17	AC005377	*** SEQUENCING IN PROG 7.85e+01
c	317	16	10.9	174582	26	HUAC003010	Homo sapiens Chromosom 7.85e+01
	318	16	10.9	177426	17	HS436K10	Human DNA sequence *** 7.85e+01
c	319	16	10.9	178698	17	HS67K17	Human DNA sequence *** 7.85e+01
	320	16	10.9	181795	26	AC003685	Homo sapiens Xp22 BAC 7.85e+01
c	321	16	10.9	181995	17	HS596C15	Human DNA sequence *** 7.85e+01
c	322	16	10.9	182123	17	AC002399	*** SEQUENCING IN PROG 7.85e+01
c	323	16	10.9	187051	17	AC004688	*** SEQUENCING IN PROG 7.85e+01
c	324	16	10.9	189357	17	HS596H12	Human DNA sequence *** 7.85e+01
c	325	16	10.9	197391	17	AC005140	*** SEQUENCING IN PROG 7.85e+01
c	326	16	10.9	199082	17	AC005632	*** SEQUENCING IN PROG 7.85e+01
c	327	16	10.9	199490	17	AC005081	*** SEQUENCING IN PROG 7.85e+01
c	328	16	10.9	205698	17	AC005773	*** SEQUENCING IN PROG 7.85e+01
	329	16	10.9	206419	23	ATAP21	Arabidopsis thaliana D 7.85e+01
	330	16	10.9	207931	17	HS101G11	Human DNA sequence *** 7.85e+01
	331	16	10.9	223408	27	AC005610	Homo sapiens chromosom 7.85e+01
c	332	16	10.9	224645	17	AC004153	*** SEQUENCING IN PROG 7.85e+01
c	333	16	10.9	229269	17	AC004953	*** SEQUENCING IN PROG 7.85e+01
	334	16	10.9	235134	17	AC004849	*** SEQUENCING IN PROG 7.85e+01
	335	16	10.9	236157	17	AC005703	*** SEQUENCING IN PROG 7.85e+01
c	336	16	10.9	243820	17	CEY75B12	Caenorhabditis elegans 7.85e+01
	337	16	10.9	251364	28	MMA000664	Mus musculus TCR beta 7.85e+01
c	338	16	10.9	256404	17	AC005071	*** SEQUENCING IN PROG 7.85e+01
	339	16	10.9	317311	17	AC004085	*** SEQUENCING IN PROG 7.85e+01
c	340	16	10.9	323792	17	AC004965	*** SEQUENCING IN PROG 7.85e+01
c	341	16	10.9	326663	26	AF011889	Human Xq28 cosmids U12 7.85e+01
c	342	15	10.2	79	28	S60502	cytoskeletal beta-acti 3.36e+02
c	343	15	10.2	79	28	MUSCBA04	Mouse gene for cytoske 3.36e+02
c	344	15	10.2	102	26	HS2PT074	H.sapiens telomeric DN 3.36e+02
c	345	15	10.2	152	30	G35271	orf9d Cryptosporidium 3.36e+02
c	346	15	10.2	163	30	G38826	TA35 Plasmodium falcip 3.36e+02
	347	15	10.2	170	18	EHEXRIDRI	E.histolytica extrachr 3.36e+02
c	348	15	10.2	170	18	S67112	rDNA episome (DraI sho 3.36e+02

	349	15	10.2	216	18	SMZMTGMDL	Strymon melinus mitoch 3.36e+02
c	350	15	10.2	218	23	YSCMTARSB	Yeast (S.cerevisiae) m 3.36e+02
c	351	15	10.2	245	30	HUMC5628	Human chromosome 5 LAN 3.36e+02
c	352	15	10.2	245	15	AF065131	Shigella flexneri DNA 3.36e+02
c	353	15	10.2	249	18	SSU43579	Strongyloides stercora 3.36e+02
c	354	15	10.2	255	30	G38665	SHGC-64057 Human Homo 3.36e+02
c	355	15	10.2	266	18	SSU43578	Strongyloides stercora 3.36e+02
	356	15	10.2	268	18	TETCBSAI	T.themophila micronucl 3.36e+02
c	357	15	10.2	273	18	SSU43576	Strongyloides stercora 3.36e+02
c	358	15	10.2	284	25	HS49C3F	H.sapiens CpG island D 3.36e+02
c	359	15	10.2	285	22	PP1PAR1	Bacteriophage miniplas 3.36e+02
c	360	15	10.2	295	30	GJ8558	SHGC-64279 Human Homo 3.36e+02
	361	15	10.2	297	28	MMJ004891	Mus musculus JunB prom 3.36e+02
c	362	15	10.2	299	25	HS49B10F	H.sapiens CpG island D 3.36e+02
	363	15	10.2	311	30	G04780	human STS WI-5164. 3.36e+02
	364	15	10.2	321	18	PTUD3620	Paramecium tetraurelia 3.36e+02
c	365	15	10.2	322	23	MISC20	Yeast mitochondrial ge 3.36e+02
c	366	15	10.2	322	23	YSCMTTGSX	Yeast (S.cerevisiae) m 3.36e+02
c	367	15	10.2	337	23	YSCMTTGS	Yeast (S.cerevisiae) m 3.36e+02
c	368	15	10.2	354	18	ASU36732	Adelges sp. 12S riboso 3.36e+02
	369	15	10.2	356	18	DVU95869	Dermacentor variabilis 3.36e+02
	370	15	10.2	357	25	HSNETEX6	H.sapiens HG NET gene 3.36e+02
c	371	15	10.2	379	18	AF010540	Plasmodium falciparum 3.36e+02
	372	15	10.2	385	18	AF020298	Melanoplus sanguinipes 3.36e+02
	373	15	10.2	385	30	HS024XB8	H.sapiens DNA segment 3.36e+02
	374	15	10.2	400	30	G11265	human STS SHGC-9613. 3.36e+02
c	375	15	10.2	423	21	I51914	Sequence 129 from pate 3.36e+02
c	376	15	10.2	423	21	I51882	Sequence 93 from paten 3.36e+02
	377	15	10.2	473	18	DMRH4PR	Drosophila melanogaste 3.36e+02
c	378	15	10.2	473	30	G36600	SHGC-53344 Human Homo 3.36e+02
	379	15	10.2	485	15	BUHARGSC	Buchnera aphidicola Ar 3.36e+02
c	380	15	10.2	504	23	PTAJ5292	Populus tremula x Popu 3.36e+02
c	381	15	10.2	563	18	TPHISIN	Tetrahymena paravorax 3.36e+02
c	382	15	10.2	589	16	AF052254	Escherichia coli DNA g 3.36e+02
	383	15	10.2	591	28	AF032459	Mus musculus BimEL mRN 3.36e+02
c	384	15	10.2	592	18	DPCH2A1	D.polychroa microsatel 3.36e+02
c	385	15	10.2	623	25	HUMZNF35P1	Homo sapiens zinc fing 3.36e+02
	386	15	10.2	659	25	HSPA18F7	H.sapiens flow-sorted 3.36e+02
c	387	15	10.2	678	30	HUMUT6065	Human STS UT6065. 3.36e+02
c	388	15	10.2	703	23	YSCMTV11	Yeast (S.cerevisiae) m 3.36e+02
	389	15	10.2	711	23	ATBIBBIND	A.thaliana gibberellin 3.36e+02
	390	15	10.2	744	21	E07851	Glycinin gene terminat 3.36e+02
	391	15	10.2	759	26	AG001178	Homo sapiens genomic D 3.36e+02
c	392	15	10.2	761	18	DROSALAE	D.simulans homeotic ge 3.36e+02
c	393	15	10.2	774	15	CSMT4116S1	Colwellia sp. MT41 16S 3.36e+02
	394	15	10.2	787	16	HPY10008	Helicobacter pylori (s 3.36e+02
c	395	15	10.2	793	25	HUMPROF	Human profilin mRNA, c 3.36e+02
c	396	15	10.2	802	18	EAARSORI	Entamoeba histolytica 3.36e+02
	397	15	10.2	819	25	HUMIGH341P	Human immunoglobulin h 3.36e+02
	398	15	10.2	837	15	MMGRPEPH	M.matei archael gene f 3.36e+02
	399	15	10.2	838	18	LOU19291	Loligo opalescens clon 3.36e+02

c	400	15	10.2	842	25	HUMGIP3A1	Human Gi3 protein alph 3.36e+02
	401	15	10.2	910	26	HSWNT12GE	Homo sapiens mRNA for 3.36e+02
	402	15	10.2	911	18	LOU19293	Loligo opalescens clon 3.36e+02
	403	15	10.2	947	16	HPY14394	Helicobacter pylori fu 3.36e+02
	404	15	10.2	947	15	BSSB25	B.subtilis SB25 ribR g 3.36e+02
	405	15	10.2	1012	26	HS19PTO09	H.sapiens telomeric DN 3.36e+02
	406	15	10.2	1072	20	XLIGMJHE	Xenopus laevis DNA fra 3.36e+02
	407	15	10.2	1212	18	ETTUFAGEN	E.tenella plastid DNA 3.36e+02
	408	15	10.2	1286	18	AF072373	Glossina morsitans mor 3.36e+02
c	409	15	10.2	1291	18	DDU13150	Dictyostelium discoide 3.36e+02
	410	15	10.2	1296	28	MMU17858	Mus musculus CRF recep 3.36e+02
	411	15	10.2	1373	23	PSGSR1	Pea mRNA for glutamine 3.36e+02
	412	15	10.2	1390	18	TRIALPHA	Trichomonas vaginalis 3.36e+02
c	413	15	10.2	1451	26	HSY08375	H.sapiens gene encodin 3.36e+02
	414	15	10.2	1452	23	SPMFM2	S.pombe MFm2 gene. 3.36e+02
	415	15	10.2	1518	25	HSU14722	Human activin type I r 3.36e+02
	416	15	10.2	1557	28	MMU21729	Mus musculus sauvagine 3.36e+02
c	417	15	10.2	1570	18	HIU11645	Haematobia irritans cl 3.36e+02
	418	15	10.2	1591	15	BACSEPRAAA	Bacillus sp. serine pr 3.36e+02
c	419	15	10.2	1601	33	HEHS07	Herpes simplex virus g 3.36e+02
	420	15	10.2	1630	18	DDMYBBD	D.discoideum DNA for M 3.36e+02
c	421	15	10.2	1656	33	HEHSV2TK	Herpes simplex virus t 3.36e+02
c	422	15	10.2	1658	33	HS2TK1A	Herpes simplex virus t 3.36e+02
c	423	15	10.2	1658	33	HS2TK4A	Herpes simplex virus t 3.36e+02
c	424	15	10.2	1659	33	HS2TK2A	Herpes simplex virus t 3.36e+02
c	425	15	10.2	1659	33	HS2TK3A	Herpes simplex virus t 3.36e+02
	426	15	10.2	1665	25	HUMSKR23A	Human serine/threonine 3.36e+02
c	427	15	10.2	1665	18	MILTRRNG	Leishmania tarentolae 3.36e+02
	428	15	10.2	1692	18	DDU61989	Dictyostelium discoide 3.36e+02
c	429	15	10.2	1706	18	LEIACTIN	Leishmania major (clon 3.36e+02
	430	15	10.2	1717	23	DDISP70B	Dictyostelium discoide 3.36e+02
	431	15	10.2	1719	25	HUMMSCA	Human GT mitochondrial 3.36e+02
c	432	15	10.2	1739	28	MMAJ92	Mus musculus gene enco 3.36e+02
	433	15	10.2	1743	23	SOYGLYBSU	Soybean glycinin A-la- 3.36e+02
	434	15	10.2	1746	21	E02463	cDNA encoding glycinin 3.36e+02
	435	15	10.2	1746	23	GMGLY1A	Soybean mRNA for glyci 3.36e+02
	436	15	10.2	1780	25	HUMSKR22A	Human serine/threonine 3.36e+02
	437	15	10.2	1848	15	S76079	aerolysin precursor=au 3.36e+02
c	438	15	10.2	1857	28	MUSACTMEL	Mouse A-X actin mRNA, 3.36e+02
c	439	15	10.2	1880	28	AF004368	Cricetulus griseus UDP 3.36e+02
c	440	15	10.2	1891	18	DDU73685	Dictyostelium discoide 3.36e+02
c	441	15	10.2	1892	28	MMACTBR	Mouse cytoskeletal mRN 3.36e+02
	442	15	10.2	1895	19	OCU19815	Oryctolagus cuniculus 3.36e+02
c	443	15	10.2	1899	19	D89477	Porcine mRNA for short 3.36e+02
	444	15	10.2	1901	28	MMU74079	Mus musculus protein c 3.36e+02
	445	15	10.2	1948	18	ET12332	E.tenella plastid gene 3.36e+02
c	446	15	10.2	2015	28	MMU39066	Murine MAP kinase kina 3.36e+02
	447	15	10.2	2029	20	GGVIPGE	G.gallus PHI encoding 3.36e+02
	448	15	10.2	2044	26	AF007152	Homo sapiens clone 236 3.36e+02
c	449	15	10.2	2086	23	PIU17009	Phytophthora infestans 3.36e+02
c	450	15	10.2	2089	18	SMSERRICH	S.mansoni mRNA for unk 3.36e+02

	451	15	10.2	2135	18	DROAMA	D.melanogaster amalgam 3.36e+02
c	452	15	10.2	2225	23	AF005048	Arabidopsis thaliana t 3.36e+02
	453	15	10.2	2247	25	HSSTROL3	Human stromelysin-3 mR 3.36e+02
	454	15	10.2	2256	21	I17126	Sequence 1 from patent 3.36e+02
	455	15	10.2	2305	26	HSU81787	Human Wnt10B mRNA, com 3.36e+02
c	456	15	10.2	2316	21	I31888	Sequence 10 from paten 3.36e+02
c	457	15	10.2	2316	21	I31887	Sequence 7 from patent 3.36e+02
	458	15	10.2	2333	25	HSALK4A	Homo sapiens ALK-4 mRN 3.36e+02
	459	15	10.2	2333	21	A38811	Sequence 7 from Patent 3.36e+02
c	460	15	10.2	2335	28	AF016545	Mus musculus Src-relat 3.36e+02
	461	15	10.2	2355	28	MMSIX1	M.musculus Six1 mRNA. 3.36e+02
	462	15	10.2	2401	23	SCGAS1	Yeast GAS1 gene for gl 3.36e+02
c	463	15	10.2	2408	33	CBVNPHI	Choristoneura biennis 3.36e+02
	464	15	10.2	2416	26	HUMRSC338	Human mRNA for KIAA000 3.36e+02
	465	15	10.2	2485	20	XLFOSP1A	X.laevis POSP-1A mRNA 3.36e+02
	466	15	10.2	2494	16	CLONITROGE	Clostridium pasteurian 3.36e+02
	467	15	10.2	2617	28	MMU19939	Mus musculus corticotr 3.36e+02
c	468	15	10.2	2649	19	BOVPNMTB	Bovine phenylethanolam 3.36e+02
c	469	15	10.2	2662	21	A08267	Nucleotide sequence fo 3.36e+02
c	470	15	10.2	2666	18	PFAAMA1A	Plasmodium falciparum 3.36e+02
c	471	15	10.2	2739	26	HSCDO2	Homo sapiens cysteine 3.36e+02
	472	15	10.2	2754	28	MUSNTF12	Mouse mRNA for zinc fi 3.36e+02
c	473	15	10.2	2803	18	HTDNALZF2	H.triserialis Lzf2 gen 3.36e+02
	474	15	10.2	2815	23	SCGGP1G	S. cerevisiae GGP1 gen 3.36e+02
	475	15	10.2	2827	28	RATCONPC5A	Rat convertase PC5 mRN 3.36e+02
	476	15	10.2	2848	28	MUSCONPC5A	Mouse convertase PC5 m 3.36e+02
c	477	15	10.2	2924	27	HSU39657	Human MAP kinase kinas 3.36e+02
	478	15	10.2	2932	23	SPPYP2	S.pombe pyp2-cDNA for 3.36e+02
c	479	15	10.2	2938	18	TTU12971	Tetrahymena thermophil 3.36e+02
	480	15	10.2	2941	23	D88260	Pisum sativum PsCHS4 g 3.36e+02
	481	15	10.2	2991	25	HUMVNTRA	Homo sapiens microsate 3.36e+02
	482	15	10.2	2991	27	S60754	(VNTR locus DXZ4, hype 3.36e+02
	483	15	10.2	2994	28	MUSPC6	Mouse mRNA for serine 3.36e+02
	484	15	10.2	3000	23	S51320	pyp2=protein tyrosine 3.36e+02
	485	15	10.2	3000	23	YSPPTPASE	Schirosaccharomyces po 3.36e+02
c	486	15	10.2	3072	24	PTR223620	Populus trichocarpa CC 3.36e+02
	487	15	10.2	3141	26	HSLAMA5	H.sapiens mRNA for lam 3.36e+02
c	488	15	10.2	3142	18	PFSC03110	Plasmodium falciparum 3.36e+02
	489	15	10.2	3210	23	GMABAB	Glycine max glycinin A 3.36e+02
c	490	15	10.2	3225	4	HSAC1453	Homo sapiens (subclone 3.36e+02
c	491	15	10.2	3279	23	YSCSIP	Saccharomyces cerevisi 3.36e+02
c	492	15	10.2	3315	23	YSCMTTGO4	Saccharomyces cerevisi 3.36e+02
c	493	15	10.2	3360	21	A17958	sau3AI R and sau3AI M 3.36e+02
c	494	15	10.2	3360	15	STASAU3AIM	S.aureus Sau3AI-restri 3.36e+02
c	495	15	10.2	3364	18	TTSNRU51	T.thermophila genes fo 3.36e+02
	496	15	10.2	3386	24	ACA59670	Amphidinium carterae l 3.36e+02
c	497	15	10.2	3387	23	ATJ001148	Arabidopsis thaliana C 3.36e+02
c	498	15	10.2	3514	33	RCSRNA2	Rice stripe virus geno 3.36e+02
c	499	15	10.2	3514	33	RCSRNA	Rice stripe virus geno 3.36e+02
	500	15	10.2	3527	23	GMGY1	Soybean Gy1 gene for g 3.36e+02
c	501	15	10.2	3530	23	AF005047	Arabidopsis thaliana t 3.36e+02

c	502	15	10.2	3542	23	YSCMTCG16	Saccharomyces cerevisi 3.36e+02
	503	15	10.2	3567	25	HSNADPTHY	H.sapiens mRNA for NAD 3.36e+02
	504	15	10.2	3572	23	PHTCYTOX	Phytophthora megasperm 3.36e+02
c	505	15	10.2	3630	23	MISC14	Yeast mitochondrial tR 3.36e+02
	506	15	10.2	3699	18	EHEXRDNA	E.histolytica extrachr 3.36e+02
c	507	15	10.2	3699	18	EHEXRDNA	E.histolytica extrachr 3.36e+02
	508	15	10.2	3709	23	MIWMTRNAV	W.mraxii mitochondriel 3.36e+02
c	509	15	10.2	3750	33	AF027657	Choristoneura fumifera 3.36e+02
c	510	15	10.2	3761	23	DCDC8	D. carota DC8 gene for 3.36e+02
c	511	15	10.2	3833	15	ECP1PARA	E. coli P1 par region 3.36e+02
c	512	15	10.2	3836	21	A39944	Sequence 1 from Patent 3.36e+02
c	513	15	10.2	3939	23	TOMPHEAMLY	Tomato phenylalanine a 3.36e+02
c	514	15	10.2	4084	18	PFSC03113	Plasmodium falciparum 3.36e+02
c	515	15	10.2	4098	28	RATACCYB	Rat cytoplasmic beta-a 3.36e+02
c	516	15	10.2	4100	28	RNAC01	Rat gene encoding cyto 3.36e+02
c	517	15	10.2	4163	23	DDIBP74	D.discoideum cyclic AM 3.36e+02
	518	15	10.2	4278	18	PFSC04090	Plasmodium falciparum 3.36e+02
	519	15	10.2	4288	15	MVFRCABDG	Methanococcus voltae f 3.36e+02
c	520	15	10.2	4321	28	MUSCADNA	Mouse neural cadherin 3.36e+02
c	521	15	10.2	4321	28	AB008811	Mus musculus mRNA for 3.36e+02
	522	15	10.2	4338	23	SCYJL069C	S.cerevisiae chromosom 3.36e+02
	523	15	10.2	4415	23	SPIRPL21AB	Spinacia oleracea nucl 3.36e+02
c	524	15	10.2	4619	25	HSU20982	Human insulin-like gro 3.36e+02
c	525	15	10.2	4800	23	MIAN08	Aspergillus nidlans mi 3.36e+02
c	526	15	10.2	4931	23	VXRNCRMBX	Cowpox virus secreted 3.36e+02
	527	15	10.2	4876	23	PSU28925	Pisum sativum cytosoli 3.36e+02
	528	15	10.2	4877	18	BMP25G	Bombyx mori gene for a 3.36e+02
	529	15	10.2	4962	20	FRY15172	Fugu rubripes surf5 an 3.36e+02
c	530	15	10.2	4992	25	HSINSU	Human gene for preproi 3.36e+02
	531	15	10.2	5117	26	AB011105	Homo sapiens mRNA for 3.36e+02
	532	15	10.2	5150	16	BFU61539	Bacillus firmus OrfA g 3.36e+02
	533	15	10.2	5198	21	I33131	Sequence 1 from patent 3.36e+02
	534	15	10.2	5208	28	MUSPC6B	Mouse mRNA for PC6B, p 3.36e+02
c	535	15	10.2	5245	26	D83778	Human mRNA for KIAA019 3.36e+02
c	536	15	10.2	5371	23	YSCMTGSA	Yeast (S.cerevisiae) m 3.36e+02
	537	15	10.2	5461	23	CAGTHIOGN	Capsicum annuum defens 3.36e+02
c	538	15	10.2	5484	26	HUMRETYKIN	Homo sapiens receptor 3.36e+02
	539	15	10.2	5656	23	PBU93196	Populus balsamifera su 3.36e+02
	540	15	10.2	5769	23	PSU28924	Pisum sativum cytosoli 3.36e+02
c	541	15	10.2	5786	19	BOVCYTB03	Bovine cytochrome b-5 3.36e+02
	542	15	10.2	5983	18	U00175	Dictyostelium mucoroid 3.36e+02
	543	15	10.2	6000	28	MMY17038	Mus musculus baasoon g 3.36e+02
c	544	15	10.2	6023	22	PP1REP	Bacteriophage P1 repli 3.36e+02
	545	15	10.2	6154	21	A37005	Sequence 1 from Patent 3.36e+02
	546	15	10.2	6208	33	S72415	putative protein kinas 3.36e+02
c	547	15	10.2	6340	23	YSCMTCG03	Saccharomyces cerevisi 3.36e+02
	548	15	10.2	6450	18	CEU36199	Caenorhabditis elegans 3.36e+02
	549	15	10.2	6534	15	D76414	Staphylococcus aureus 3.36e+02
	550	15	10.2	6540	15	BACJOJC	Bacillus subtilis dihy 3.36e+02
c	551	15	10.2	6618	27	AB007913	Homo sapiens mRNA for 3.36e+02
	552	15	10.2	6632	27	AB007917	Homo sapiens mRNA for 3.36e+02

	553	15	10.2	6807	28	AF004874	Mus musculus latent TG 3.36e+02
c	554	15	10.2	7057	28	MUSA1CC	Mouse mRNA for mouse a 3.36e+02
	555	15	10.2	7104	20	BRMHCIUCA	B.rerio UCA*01 class I 3.36e+02
	556	15	10.2	7198	15	PFU41818	Pseudomonas fluorescen 3.36e+02
c	557	15	10.2	7371	19	OAKII9	O.aries KII-9 gene for 3.36e+02
	558	15	10.2	7534	18	CELF58D7	Caenorhabditis elegans 3.36e+02
c	559	15	10.2	7589	18	DROROUGHA	Drosophila virilis rou 3.36e+02
c	560	15	10.2	7698	18	CCU60529	Ceratitis capitata yoy 3.36e+02
	561	15	10.2	7984	26	AB006651	Homo sapiens EXIM1 mRN 3.36e+02
	562	15	10.2	8033	18	PPINMP	Paramecium primaurelia 3.36e+02
	563	15	10.2	8076	28	MUSGPOAD	Mouse androgen-depende 3.36e+02
	564	15	10.2	8155	28	MMU20735	Mus musculus transcrip 3.36e+02
	565	15	10.2	8276	33	D42113	Feline herpesvirus typ 3.36e+02
	566	15	10.2	8628	25	HSU20734	Human transcription fa 3.36e+02
	567	15	10.2	8699	18	PFSC03018	Plasmodium falciparum 3.36e+02
c	568	15	10.2	8846	23	FPPPCA	F.pringlei ppcA1 gene 3.36e+02
	569	15	10.2	9123	23	D84656	Yeast DNA for bfr2+ pr 3.36e+02
	570	15	10.2	10028	15	U67503	Methanococcus jannasch 3.36e+02
c	571	15	10.2	10145	15	AE000839	Methanobacterium therm 3.36e+02
	572	15	10.2	10782	15	HPAE000611	Helicobacter pylori se 3.36e+02
	573	15	10.2	10914	18	PFSC04058	Plasmodium falciparum 3.36e+02
c	574	15	10.2	11136	15	AE000862	Methanobacterium therm 3.36e+02
	575	15	10.2	11485	23	AF052570	Populus balsamifera su 3.36e+02
	576	15	10.2	11553	15	HPAE000610	Helicobacter pylori se 3.36e+02
c	577	15	10.2	11680	15	HPAE000609	Helicobacter pylori se 3.36e+02
	578	15	10.2	11772	18	PFSC03022	Plasmodium falciparum 3.36e+02
	579	15	10.2	12318	18	CET23H4	Caenorhabditis elegans 3.36e+02
c	580	15	10.2	12403	15	SEHSD	S.enterica hsdM, hsdS 3.36e+02
	581	15	10.2	12507	28	RNY16563	Rattus norvegicus mRNA 3.36e+02
c	582	15	10.2	12936	18	CELR10H1	Caenorhabditis elegans 3.36e+02
	583	15	10.2	13242	15	AE000858	Methanobacterium therm 3.36e+02
c	584	15	10.2	13912	15	AE000929	Methanobacterium therm 3.36e+02
	585	15	10.2	14130	18	MIACDNA	A.coerulea complete mi 3.36e+02
c	586	15	10.2	14252	18	CTU07703	Chironomus thummi 7B g 3.36e+02
c	587	15	10.2	14440	23	EMEMTDNA	Emericella nidulans mt 3.36e+02
	588	15	10.2	14561	25	HSRING3GE	H.sapiens RING3 gene. 3.36e+02
c	589	15	10.2	15432	15	AE000878	Methanobacterium therm 3.36e+02
	590	15	10.2	15509	18	CEC03D6	Caenorhabditis elegans 3.36e+02
c	591	15	10.2	16011	31	XXU09128	pSacBII P1 cloning vec 3.36e+02
c	592	15	10.2	16202	23	ATZ83321	A.thaliana 16 kb chrom 3.36e+02
	593	15	10.2	16288	26	HSNCAMX1	H.sapiens gene for neu 3.36e+02
	594	15	10.2	16390	28	MUSSEQAA	Mus musculus (clone R2 3.36e+02
	595	15	10.2	16630	24	SPCC594	S.pombe chromosome III 3.36e+02
	596	15	10.2	16934	33	LCCVORFS	Little cherry clostero 3.36e+02
	597	15	10.2	17137	23	SCXCDNA	S.cerevisiae (S288C) X 3.36e+02
	598	15	10.2	17517	28	RATGCA	Rat guanylyl cyclase A 3.36e+02
	599	15	10.2	18630	22	BB103G	Bacteriophage B103 gen 3.36e+02
c	600	15	10.2	19500	31	CVU75992	Cloning vector pPAC4, 3.36e+02
	601	15	10.2	20369	18	CE2K525	Caenorhabditis elegans 3.36e+02
c	602	15	10.2	20992	18	LEIKPMAX	Leishmania tarentolae 3.36e+02
c	603	15	10.2	21552	18	CEM04D8	Caenorhabditis elegans 3.36e+02

	604	15	10.2	21733	18	CEW10D5	Caenorhabditis elegans 3.36e+02
c	605	15	10.2	22254	18	CET22G5	Caenorhabditis elegans 3.36e+02
	606	15	10.2	23775	15	BACYPIA	Bacillus subtilis (clo 3.36e+02
c	607	15	10.2	24554	18	CET05E11	Caenorhabditis elegans 3.36e+02
	608	15	10.2	25030	18	CEF30F8	Caenorhabditis elegans 3.36e+02
c	609	15	10.2	26122	18	CELW02F12	Caenorhabditis elegans 3.36e+02
c	610	15	10.2	26821	18	CEY7A9C	Caenorhabditis elegans 3.36e+02
c	611	15	10.2	27019	18	CEF47C8	Caenorhabditis elegans 3.36e+02
	612	15	10.2	27111	18	CELC45E1	Caenorhabditis elegans 3.36e+02
c	613	15	10.2	27567	18	CELT27C4	Caenorhabditis elegans 3.36e+02
c	614	15	10.2	28409	18	CELC06A6	Caenorhabditis elegans 3.36e+02
c	615	15	10.2	28585	18	CELC16B8	Caenorhabditis elegans 3.36e+02
	616	15	10.2	29286	23	SC9952X	S.cerevisiae chromosom 3.36e+02
c	617	15	10.2	29583	18	CEF22G12	Caenorhabditis elegans 3.36e+02
	618	15	10.2	29793	18	CEF20C5	Caenorhabditis elegans 3.36e+02
	619	15	10.2	29807	28	MMU73107	Mus musculus adenosine 3.36e+02
	620	15	10.2	30389	23	SPAC19D5	S.pombe chromosome I c 3.36e+02
c	621	15	10.2	31252	25	HSJ256K2B	Human DNA sequence fro 3.36e+02
c	622	15	10.2	31975	24	SPCC553	S.pombe chromosome III 3.36e+02
c	623	15	10.2	32263	18	CEF26H9	Caenorhabditis elegans 3.36e+02
c	624	15	10.2	33246	18	DMC25E8	Drosophila melanogaste 3.36e+02
	625	15	10.2	33550	26	AC003004	Human Chromosome 15 co 3.36e+02
	626	15	10.2	33614	18	CELC25B8	Caenorhabditis elegans 3.36e+02
c	627	15	10.2	33651	18	CELF44E2	C. elegans cosmid F44E 3.36e+02
	628	15	10.2	34263	18	CET02E9	Caenorhabditis elegans 3.36e+02
	629	15	10.2	34497	18	CELT22E5	Caenorhabditis elegans 3.36e+02
	630	15	10.2	34511	18	CELEC404	Caenorhabditis elegans 3.36e+02
c	631	15	10.2	34931	25	HUM85D2	Human PAC 85D2, comple 3.36e+02
	632	15	10.2	35000	15	AB000617	Bacillus subtilis geno 3.36e+02
	633	15	10.2	35365	18	CELF27D9	Caenorhabditis elegans 3.36e+02
c	634	15	10.2	35383	18	CELZC449	Caenorhabditis elegans 3.36e+02
c	635	15	10.2	35700	26	HS27C3	Homo sapiens DNA seque 3.36e+02
	636	15	10.2	36148	17	AC002321	*** SEQUENCING IN PROG 3.36e+02
	637	15	10.2	36388	26	AC002309	Homo sapiens Chromosom 3.36e+02
	638	15	10.2	36393	26	D87013	Human (lambda) DNA for 3.36e+02
	639	15	10.2	36741	25	HUMADAG	Human adenosine deamin 3.36e+02
	640	15	10.2	36822	18	CELC25E10	Caenorhabditis elegans 3.36e+02
c	641	15	10.2	36829	18	CELZC239	Caenorhabditis elegans 3.36e+02
c	642	15	10.2	36933	23	SPAC4G8	S.pombe chromosome I c 3.36e+02
c	643	15	10.2	37714	18	CET11F9	Caenorhabditis elegans 3.36e+02
c	644	15	10.2	38066	18	CELF15A8	Caenorhabditis elegans 3.36e+02
c	645	15	10.2	38110	16	MTCY427	Mycobacterium tubercul 3.36e+02
c	646	15	10.2	38168	26	HSU96H11	Human DNA sequence fro 3.36e+02
c	647	15	10.2	38383	18	CEF35H8	Caenorhabditis elegans 3.36e+02
c	648	15	10.2	38682	26	HSAC002077	Human cosmid clone LDC 3.36e+02
c	649	15	10.2	38826	26	AC005356	Homo sapiens chromosom 3.36e+02
c	650	15	10.2	38902	26	HSO27	Human DNA sequence fro 3.36e+02
c	651	15	10.2	38939	26	AC004678	Homo sapiens chromosom 3.36e+02
c	652	15	10.2	39526	26	AC003961	Human Cosmid g5129z180 3.36e+02
c	653	15	10.2	39594	23	SCESTGENE	S.cerevisiae DNA for h 3.36e+02
	654	15	10.2	39729	24	SPAC31G5	S.pombe chromosome I c 3.36e+02

c	655	15	10.2	40014	18	CELF10C1	Caenorhabditis elegans 3.36e+02
c	656	15	10.2	40662	27	HSN44A4	Human DNA sequence fro 3.36e+02
c	657	15	10.2	40740	26	HSO14	Human DNA sequence fro 3.36e+02
	658	15	10.2	41071	18	CELB0303	C.elegans cosmid B0303 3.36e+02
	659	15	10.2	41084	26	AC002500	Human Cosmid g5129z101 3.36e+02
	660	15	10.2	41088	18	CELB0280	Caenorhabditis elegans 3.36e+02
c	661	15	10.2	41150	18	CEF47H4	Caenorhabditis elegans 3.36e+02
	662	15	10.2	41230	18	CELC14A11	Caenorhabditis elegans 3.36e+02
c	663	15	10.2	41299	18	CELF55A4	Caenorhabditis elegans 3.36e+02
c	664	15	10.2	41306	25	HSAC000127	Human Cosmid g1572c198 3.36e+02
	665	15	10.2	42163	18	CEF49C5	Caenorhabditis elegans 3.36e+02
c	666	15	10.2	42393	27	AC005780	Homo sapiens chromosom 3.36e+02
c	667	15	10.2	42398	26	D84401	Homo sapiens genomic D 3.36e+02
c	668	15	10.2	42799	18	CELF47E1	Caenorhabditis elegans 3.36e+02
	669	15	10.2	43180	26	AC005360	Homo sapiens chromosom 3.36e+02
c	670	15	10.2	43455	18	CELT07A9	Caenorhabditis elegans 3.36e+02
c	671	15	10.2	43593	26	AC003960	Human Cosmid g5129g129 3.36e+02
c	672	15	10.2	43600	27	AC004209	Homo sapiens clone UWG 3.36e+02
	673	15	10.2	43661	26	AC004493	Homo sapiens chromosom 3.36e+02
c	674	15	10.2	45034	26	AF020803	Homo sapiens cosmid Q3 3.36e+02
c	675	15	10.2	45189	18	CELC34F11	Caenorhabditis elegans 3.36e+02
c	676	15	10.2	45701	17	CETY4	Caenorhabditis elegans 3.36e+02
	677	15	10.2	53517	17	AC002405	*** SEQUENCING IN PROG 3.36e+02
c	678	15	10.2	54336	26	HUMXPDG1	Homo sapiens ERCC2 gen 3.36e+02
	679	15	10.2	56959	27	HS726P20	Human DNA sequence *** 3.36e+02
	680	15	10.2	57328	18	PFSC03065	Plasmodium falciparum 3.36e+02
c	681	15	10.2	58407	16	MII2CG	Methanococcus jannasch 3.36e+02
c	682	15	10.2	59383	18	PFSC03066	Plasmodium falciparum 3.36e+02
c	683	15	10.2	59725	24	ENU34740	Emericella nidulans st 3.36e+02
c	684	15	10.2	59782	18	PFSC03093	Plasmodium falciparum 3.36e+02
	685	15	10.2	65958	23	AB009056	Arabidopsis thaliana g 3.36e+02
c	686	15	10.2	67833	18	AC004564	Drosophila melanogaste 3.36e+02
c	687	15	10.2	69034	18	AF007261	Reclinomonas americana 3.36e+02
c	688	15	10.2	70319	26	AC002453	Human BAC clone GS440B 3.36e+02
c	689	15	10.2	70881	27	HS293L6B	Human DNA sequence fro 3.36e+02
	690	15	10.2	71504	18	AC002473	Drosophila melanogaste 3.36e+02
	691	15	10.2	73350	18	AC002493	Drosophila melanogaste 3.36e+02
c	692	15	10.2	73350	18	AC002493	Drosophila melanogaste 3.36e+02
	693	15	10.2	74432	26	HS465G10B	Human DNA sequence fro 3.36e+02
c	694	15	10.2	78174	18	AC004280	Drosophila melanogaste 3.36e+02
c	695	15	10.2	78193	17	HS356B8	Human DNA sequence *** 3.36e+02
	696	15	10.2	78500	23	SCD9461	Saccharomyces cerevisi 3.36e+02
	697	15	10.2	78520	26	AC004455	Homo sapiens PAC clone 3.36e+02
c	698	15	10.2	79837	23	AB005233	Arabidopsis thaliana g 3.36e+02
	699	15	10.2	80315	23	AB005232	Arabidopsis thaliana g 3.36e+02
c	700	15	10.2	80970	23	AB006697	Arabidopsis thaliana g 3.36e+02
	701	15	10.2	82001	23	AB005242	Arabidopsis thaliana g 3.36e+02
c	702	15	10.2	82393	18	AC003923	Drosophila melanogaste 3.36e+02
	703	15	10.2	82952	24	AB015476	Arabidopsis thaliana g 3.36e+02
	704	15	10.2	83605	26	AC004220	Homo sapiens chromosom 3.36e+02
	705	15	10.2	83865	23	AB006700	Arabidopsis thaliana g 3.36e+02

	706	15	10.2	84246	18	AC002512	Drosophila melanogaste 3.36e+02
	707	15	10.2	84544	23	AB005239	Arabidopsis thaliana g 3.36e+02
c	708	15	10.2	84704	18	CEY48E1B	Caenorhabditis elegans 3.36e+02
	709	15	10.2	84736	17	AC004793	*** SEQUENCING IN PROG 3.36e+02
c	710	15	10.2	84736	17	AC004793	*** SEQUENCING IN PROG 3.36e+02
c	711	15	10.2	85214	24	ATAC005396	Arabidopsis thaliana c 3.36e+02
c	712	15	10.2	85463	24	ATAC004483	Arabidopsis thaliana c 3.36e+02
c	713	15	10.2	85833	26	HS344F17	Human DNA sequence fro 3.36e+02
	714	15	10.2	86419	18	AC004345	Drosophila melanogaste 3.36e+02
	715	15	10.2	87192	24	T8F5	Arabidopsis thaliana c 3.36e+02
c	716	15	10.2	87459	24	AB015477	Arabidopsis thaliana g 3.36e+02
c	717	15	10.2	87487	24	AB015474	Arabidopsis thaliana g 3.36e+02
	718	15	10.2	88057	18	AC004120	Drosophila melanogaste 3.36e+02
	719	15	10.2	88128	24	AB017069	Arabidopsis thaliana g 3.36e+02
c	720	15	10.2	89743	28	AC003062	Mouse Chromosome 16 Re 3.36e+02
	721	15	10.2	89779	24	ATAC004669	Arabidopsis thaliana c 3.36e+02
c	722	15	10.2	89779	23	AB005234	Arabidopsis thaliana g 3.36e+02
c	723	15	10.2	91690	25	HS96A9	Human DNA sequence fro 3.36e+02
	724	15	10.2	91927	27	AC004771	Homo sapiens chromosom 3.36e+02
	725	15	10.2	92802	17	AC004710	*** SEQUENCING IN PROG 3.36e+02
c	726	15	10.2	92863	20	GGBLOCUS	Gallus gallus (chicken 3.36e+02
	727	15	10.2	92879	17	AC004450	*** SEQUENCING IN PROG 3.36e+02
c	728	15	10.2	92879	17	AC004450	*** SEQUENCING IN PROG 3.36e+02
c	729	15	10.2	95272	18	AC005298	Drosophila melanogaste 3.36e+02
	730	15	10.2	96801	23	ATAC003033	Arabidopsis thaliana c 3.36e+02
	731	15	10.2	97979	26	AC004739	Homo sapiens PAC clone 3.36e+02
c	732	15	10.2	98274	27	HS714B7	Human DNA sequence fro 3.36e+02
c	733	15	10.2	98695	18	AC005556	Drosophila melanogaste 3.36e+02
c	734	15	10.2	100427	26	AC004048	Homo sapiens chromosom 3.36e+02
	735	15	10.2	101072	27	AC005752	Homo sapiens chromosom 3.36e+02
c	736	15	10.2	102703	17	HS968D22	Human DNA sequence *** 3.36e+02
c	737	15	10.2	102798	17	AC000107	*** SEQUENCING IN PROG 3.36e+02
	738	15	10.2	103391	17	HS892F13	Human DNA sequence *** 3.36e+02
	739	15	10.2	103632	24	AF074021	Arabidopsis thaliana B 3.36e+02
c	740	15	10.2	104597	26	HS393P12	Homo sapiens DNA seque 3.36e+02
c	741	15	10.2	104609	17	AC005683	*** SEQUENCING IN PROG 3.36e+02
	742	15	10.2	105119	17	AC004856	*** SEQUENCING IN PROG 3.36e+02
c	743	15	10.2	105654	23	ATAF001308	Arabidopsis thaliana c 3.36e+02
	744	15	10.2	105716	17	AC005724	*** SEQUENCING IN PROG 3.36e+02
c	745	15	10.2	105754	17	AC005000	*** SEQUENCING IN PROG 3.36e+02
	746	15	10.2	105830	24	AF069441	Arabidopsis thaliana B 3.36e+02
c	747	15	10.2	106274	17	AC004390	*** SEQUENCING IN PROG 3.36e+02
c	748	15	10.2	106420	26	AC002546	Homo sapiens chromosom 3.36e+02
c	749	15	10.2	106921	26	AC004263	Homo sapiens PAC clone 3.36e+02
c	750	15	10.2	107352	26	HS474A14	Human DNA sequence fro 3.36e+02
c	751	15	10.2	109866	26	AF043945	Homo sapiens chromosom 3.36e+02
	752	15	10.2	110000	17	CEY106G6	Caenorhabditis elegans 3.36e+02
	753	15	10.2	111368	26	AP000025	Homo sapiens genomic D 3.36e+02
c	754	15	10.2	112242	26	AC004741	Homo sapiens PAC clone 3.36e+02
c	755	15	10.2	112314	26	AC004002	Human BAC clone RG356F 3.36e+02
c	756	15	10.2	112400	17	HS30A23	Human DNA sequence *** 3.36e+02

	757	15	10.2	114929	26	AC002401	Homo sapiens chromosom 3.36e+02
c	758	15	10.2	115325	17	AC003659	*** SEQUENCING IN PROG 3.36e+02
	759	15	10.2	115601	17	HS981L23	Human DNA sequence *** 3.36e+02
	760	15	10.2	116513	26	HS487J7	Human DNA sequence fro 3.36e+02
	761	15	10.2	118357	27	HS1142C11	Human DNA sequence fro 3.36e+02
c	762	15	10.2	118758	26	HS450C20	Human DNA sequence fro 3.36e+02
	763	15	10.2	118791	17	HS1177I5	Human DNA sequence *** 3.36e+02
	764	15	10.2	119301	17	AC004920	*** SEQUENCING IN PROG 3.36e+02
c	765	15	10.2	119301	17	AC004920	*** SEQUENCING IN PROG 3.36e+02
c	766	15	10.2	119678	26	AF064866	Homo sapiens chromosom 3.36e+02
	767	15	10.2	119707	23	PINCPTRPG	Pinus thunbergii chlor 3.36e+02
	768	15	10.2	120321	17	HS341D10	Human DNA sequence *** 3.36e+02
	769	15	10.2	120359	27	AC005389	Homo sapiens chromosom 3.36e+02
	770	15	10.2	120423	26	HS326L12	Homo sapiens DNA seque 3.36e+02
c	771	15	10.2	121434	17	AC005116	*** SEQUENCING IN PROG 3.36e+02
	772	15	10.2	122000	26	AC003093	Human BAC clone RG385F 3.36e+02
	773	15	10.2	122928	23	ATAC002387	Arabidopsis thaliana c 3.36e+02
c	774	15	10.2	123233	17	AC005294	*** SEQUENCING IN PROG 3.36e+02
	775	15	10.2	123435	17	CEY61B8	Caenorhabditis elegans 3.36e+02
	776	15	10.2	123814	26	AC005393	Homo sapiens chromosom 3.36e+02
c	777	15	10.2	123891	26	AC004549	Homo sapiens BAC clone 3.36e+02
c	778	15	10.2	124497	26	HS232G24	Homo sapiens DNA seque 3.36e+02
	779	15	10.2	124519	28	AC003019	Mus musculus Bac B6, c 3.36e+02
c	780	15	10.2	124781	17	HS324C6	Human DNA sequence *** 3.36e+02
	781	15	10.2	125108	26	AF030453	Homo sapiens chromosom 3.36e+02
	782	15	10.2	125345	17	AC004818	*** SEQUENCING IN PROG 3.36e+02
c	783	15	10.2	125590	26	AC002382	Human BAC clone RG022J 3.36e+02
c	784	15	10.2	125927	27	AC005754	Homo sapiens chromosom 3.36e+02
c	785	15	10.2	125971	26	HS363L9	Human DNA sequence fro 3.36e+02
c	786	15	10.2	126270	26	HS211D12	Human DNA sequence fro 3.36e+02
c	787	15	10.2	126983	17	AC005283	*** SEQUENCING IN PROG 3.36e+02
c	788	15	10.2	127178	26	AC005160	Homo sapiens BAC clone 3.36e+02
c	789	15	10.2	127270	26	AC004776	Homo sapiens PAC clone 3.36e+02
	790	15	10.2	127506	26	AC005244	Homo sapiens chromosom 3.36e+02
	791	15	10.2	127757	26	AC004468	Homo sapiens Xp22 PAC 3.36e+02
	792	15	10.2	127766	27	AC005723	Homo sapiens chromosom 3.36e+02
c	793	15	10.2	128117	26	AC004000	Human PAC clone DJ404F 3.36e+02
c	794	15	10.2	128136	17	HS889N15	Human DNA sequence *** 3.36e+02
	795	15	10.2	128361	27	AC004877	Homo sapiens PAC clone 3.36e+02
	796	15	10.2	129169	17	AC005506	*** SEQUENCING IN PROG 3.36e+02
	797	15	10.2	129429	17	AC004951	*** SEQUENCING IN PROG 3.36e+02
	798	15	10.2	129861	17	AC004756	*** SEQUENCING IN PROG 3.36e+02
	799	15	10.2	129957	17	HS415G2	Human DNA sequence *** 3.36e+02
c	800	15	10.2	131447	26	HS380E11	Homo sapiens DNA seque 3.36e+02
	801	15	10.2	131648	17	AC004817	*** SEQUENCING IN PROG 3.36e+02
c	802	15	10.2	132826	17	AC005627	*** SEQUENCING IN PROG 3.36e+02
c	803	15	10.2	133687	17	AC002093	*** SEQUENCING IN PROG 3.36e+02
c	804	15	10.2	134478	17	HS154A10	Human DNA sequence *** 3.36e+02
	805	15	10.2	134534	17	HS1163J1	Human DNA sequence *** 3.36e+02
c	806	15	10.2	135090	26	HS164L12	Homo sapiens DNA seque 3.36e+02
c	807	15	10.2	135648	26	AC004069	Homo sapiens chromosom 3.36e+02

	808	15	10.2	135705	17	HS97P20	Human DNA sequence *** 3.36e+02
	809	15	10.2	136037	26	AC004104	Homo sapiens 12q24 PAC 3.36e+02
	810	15	10.2	136292	17	HS8B22	Human DNA sequence *** 3.36e+02
	811	15	10.2	137260	17	PFMAL1P4	Plasmodium falciparum 3.36e+02
	812	15	10.2	137968	17	AC004952	*** SEQUENCING IN PROG 3.36e+02
	813	15	10.2	138181	17	AC005171	*** SEQUENCING IN PROG 3.36e+02
	814	15	10.2	138445	17	AC002979	*** SEQUENCING IN PROG 3.36e+02
c	815	15	10.2	138445	17	AC002979	*** SEQUENCING IN PROG 3.36e+02
c	816	15	10.2	139158	17	HS316D5	Human DNA sequence *** 3.36e+02
c	817	15	10.2	139389	26	HS68O2	Homo sapiens DNA seque 3.36e+02
	818	15	10.2	139480	26	HUAC002331	Homo sapiens Chromosom 3.36e+02
c	819	15	10.2	140530	17	HS153G14	Human DNA sequence *** 3.36e+02
c	820	15	10.2	140704	17	AC004916	*** SEQUENCING IN PROG 3.36e+02
c	821	15	10.2	140769	17	HS377F16	Human DNA sequence *** 3.36e+02
c	822	15	10.2	140849	17	HS20B11	Human DNA sequence *** 3.36e+02
c	823	15	10.2	140901	17	HS888M10	Human DNA sequence *** 3.36e+02
c	824	15	10.2	140942	26	AC002525	Human PAC clone 257C22 3.36e+02
c	825	15	10.2	141387	27	AC005686	Homo sapiens clone UWG 3.36e+02
	826	15	10.2	141752	17	HS971N18	Human DNA sequence *** 3.36e+02
	827	15	10.2	141878	26	AC000004	Genomic sequence from 3.36e+02
	828	15	10.2	141983	17	HS169I5	Human DNA sequence *** 3.36e+02
c	829	15	10.2	142396	25	HSAC000120	Human BAC clone RG161K 3.36e+02
c	830	15	10.2	143308	15	D90913	Synechocystis sp. PCC6 3.36e+02
c	831	15	10.2	143577	17	AC004846	*** SEQUENCING IN PROG 3.36e+02
c	832	15	10.2	144227	26	HS119E23	Homo sapiens DNA seque 3.36e+02
c	833	15	10.2	144239	26	AC004015	Human BAC clone RG333F 3.36e+02
c	834	15	10.2	144474	26	AF051934	Homo sapiens chromosom 3.36e+02
c	835	15	10.2	144759	26	HS352A20	Homo sapiens DNA seque 3.36e+02
	836	15	10.2	145618	26	AC005138	Homo sapiens chromosom 3.36e+02
c	837	15	10.2	145961	17	AC005805	*** SEQUENCING IN PROG 3.36e+02
c	838	15	10.2	147358	17	AC005630	*** SEQUENCING IN PROG 3.36e+02
c	839	15	10.2	147542	17	HS754E20	Human DNA sequence *** 3.36e+02
c	840	15	10.2	147615	17	HS453C12	Human DNA sequence *** 3.36e+02
	841	15	10.2	147686	27	AC005702	Homo sapiens chromosom 3.36e+02
	842	15	10.2	147927	17	HS324M8	Human DNA sequence *** 3.36e+02
c	843	15	10.2	148416	26	HSAC002089	Human BAC clone RG308B 3.36e+02
c	844	15	10.2	148540	17	HS28H20	Human DNA sequence *** 3.36e+02
	845	15	10.2	149409	26	AC004081	Homo sapiens PAC clone 3.36e+02
c	846	15	10.2	149420	17	HS5O6	Human DNA sequence *** 3.36e+02
c	847	15	10.2	149764	26	HSAC002076	Human BAC clone GS345D 3.36e+02
	848	15	10.2	149765	26	AC004984	Homo sapiens PAC clone 3.36e+02
c	849	15	10.2	149963	17	HS466P17	Human DNA sequence *** 3.36e+02
	850	15	10.2	150158	17	HS1147O16	Human DNA sequence *** 3.36e+02
c	851	15	10.2	150158	17	HS1147O16	Human DNA sequence *** 3.36e+02
c	852	15	10.2	151018	17	PFMAL1P6	Plasmodium falciparum 3.36e+02
	853	15	10.2	151390	17	HS92N15	Human DNA sequence *** 3.36e+02
	854	15	10.2	151630	26	AC005154	Homo sapiens PAC clone 3.36e+02
c	855	15	10.2	151964	26	AF047825	Homo sapiens PAC 50H2 3.36e+02
c	856	15	10.2	152077	26	HS130G2	Human DNA sequence fro 3.36e+02
c	857	15	10.2	152184	17	HS593P22	Human DNA sequence *** 3.36e+02
c	858	15	10.2	152392	17	HS1018D12	Human DNA sequence *** 3.36e+02

c	859	15	10.2	153356	17	AC004995	*** SEQUENCING IN PROG 3.36e+02
c	860	15	10.2	153672	17	AC004917	*** SEQUENCING IN PROG 3.36e+02
c	861	15	10.2	153692	17	CEY26D4	Caenorhabditis elegans 3.36e+02
	862	15	10.2	154746	33	HSV2HG52	Herpes simplex virus t 3.36e+02
c	863	15	10.2	154909	17	AC005799	*** SEQUENCING IN PROG 3.36e+02
	864	15	10.2	154942	17	AC005227	*** SEQUENCING IN PROG 3.36e+02
	865	15	10.2	155398	17	AC003020	*** SEQUENCING IN PROG 3.36e+02
	866	15	10.2	155536	26	AC004112	Homo sapiens BAC clone 3.36e+02
	867	15	10.2	155939	24	CHNTXX	Tobacco chloroplast ge 3.36e+02
c	868	15	10.2	156089	17	AC004586	*** SEQUENCING IN PROG 3.36e+02
c	869	15	10.2	156197	25	HS97K10	Human DNA sequence fro 3.36e+02
	870	15	10.2	156490	26	AC004052	Homo sapiens chromosom 3.36e+02
	871	15	10.2	158172	17	AC004884	*** SEQUENCING IN PROG 3.36e+02
c	872	15	10.2	158293	26	AC004746	Homo sapiens BAC clone 3.36e+02
	873	15	10.2	158886	26	AC004674	Homo sapiens Xp22 BAC 3.36e+02
c	874	15	10.2	159948	17	AC005103	*** SEQUENCING IN PROG 3.36e+02
	875	15	10.2	160986	17	AC005505	*** SEQUENCING IN PROG 3.36e+02
c	876	15	10.2	161146	17	HS323A16	Human DNA sequence *** 3.36e+02
c	877	15	10.2	161242	17	HS366B10	Human DNA sequence *** 3.36e+02
	878	15	10.2	161824	17	AC005061	*** SEQUENCING IN PROG 3.36e+02
c	879	15	10.2	162472	17	AC005730	*** SEQUENCING IN PROG 3.36e+02
c	880	15	10.2	162485	17	AC004840	*** SEQUENCING IN PROG 3.36e+02
	881	15	10.2	162577	17	AC004086	*** SEQUENCING IN PROG 3.36e+02
	882	15	10.2	162750	17	AC003061	*** SEQUENCING IN PROG 3.36e+02
	883	15	10.2	163027	26	AC002549	Homo sapiens Xp22 BAC 3.36e+02
c	884	15	10.2	163908	27	AC005411	Homo sapiens chromosom 3.36e+02
c	885	15	10.2	164017	26	AC005304	Homo sapiens chromosom 3.36e+02
c	886	15	10.2	164049	17	AC005232	*** SEQUENCING IN PROG 3.36e+02
c	887	15	10.2	164237	17	AC004838	*** SEQUENCING IN PROG 3.36e+02
	888	15	10.2	164511	27	AC004894	Homo sapiens PAC clone 3.36e+02
	889	15	10.2	164536	17	HS134N8	Human DNA sequence *** 3.36e+02
	890	15	10.2	164675	26	AC004384	Human Chromosome X clo 3.36e+02
c	891	15	10.2	165659	26	AF003529	Homo sapiens glypican 3.36e+02
c	892	15	10.2	166088	17	AC005088	*** SEQUENCING IN PROG 3.36e+02
	893	15	10.2	166126	26	AC004142	Homo sapiens BAC clone 3.36e+02
c	894	15	10.2	167217	26	AC002350	Homo sapiens 12q24 PAC 3.36e+02
	895	15	10.2	169027	17	AC002370	*** SEQUENCING IN PROG 3.36e+02
	896	15	10.2	169283	17	AC004935	*** SEQUENCING IN PROG 3.36e+02
	897	15	10.2	169998	26	HS279N11	Human DNA sequence fro 3.36e+02
	898	15	10.2	170000	17	AC004524	*** SEQUENCING IN PROG 3.36e+02
c	899	15	10.2	170342	17	AC003668	*** SEQUENCING IN PROG 3.36e+02
c	900	15	10.2	170486	17	HS1014D13	Human DNA sequence *** 3.36e+02
	901	15	10.2	171368	26	HSD95738	Human chromosome 16p13 3.36e+02
c	902	15	10.2	172780	17	AC005086	*** SEQUENCING IN PROG 3.36e+02
c	903	15	10.2	172983	26	AC002454	Human BAC clone GS552A 3.36e+02
	904	15	10.2	174045	25	HS107N3	Human DNA sequence fro 3.36e+02
	905	15	10.2	174285	17	HS524E15	Human DNA sequence *** 3.36e+02
c	906	15	10.2	174424	27	HSU52112	Homo sapiens Xq28 geno 3.36e+02
	907	15	10.2	174952	26	AC002460	Human BAC clone RG191D 3.36e+02
c	908	15	10.2	175162	17	AC005182	*** SEQUENCING IN PROG 3.36e+02
c	909	15	10.2	175737	27	HS1033B10	Human DNA sequence *** 3.36e+02

c	910	15	10.2	176109	26	AC003992	Human BAC clone RG250D 3.36e+02
	911	15	10.2	176781	17	AC002988	*** SEQUENCING IN PROG 3.36e+02
c	912	15	10.2	176781	17	AC002989	*** SEQUENCING IN PROG 3.36e+02
	913	15	10.2	177142	26	AC004582	Human Chromosome 11p14 3.36e+02
c	914	15	10.2	177189	17	HS428A13	*** SEQUENCING IN PROG 3.36e+02
c	915	15	10.2	178615	17	HS742C19	Human DNA sequence *** 3.36e+02
	916	15	10.2	179074	17	HS191D12	Human DNA sequence *** 3.36e+02
	917	15	10.2	179531	26	AC003080	Human BAC clone GS368F 3.36e+02
c	918	15	10.2	179544	26	AF015720	Homo sapiens chromosom 3.36e+02
c	919	15	10.2	179861	17	CEY65A5	Caenorhabditis elegans 3.36e+02
c	920	15	10.2	181331	17	AC005230	*** SEQUENCING IN PROG 3.36e+02
c	921	15	10.2	181588	26	AC002487	Human BAC clone RG369K 3.36e+02
	922	15	10.2	181978	17	AC005058	*** SEQUENCING IN PROG 3.36e+02
	923	15	10.2	182433	26	AC003078	Human BAC clone GS117O 3.36e+02
	924	15	10.2	184159	17	AC005487	*** SEQUENCING IN PROG 3.36e+02
	925	15	10.2	184289	17	AC004828	*** SEQUENCING IN PROG 3.36e+02
	926	15	10.2	184982	17	HS327J16	Human DNA sequence *** 3.36e+02
c	927	15	10.2	184982	17	HS327J16	Human DNA sequence *** 3.36e+02
c	928	15	10.2	186314	28	AC003694	Mus musculus chromosom 3.36e+02
	929	15	10.2	187051	17	AC004688	*** SEQUENCING IN PROG 3.36e+02
	930	15	10.2	188629	17	AC004816	*** SEQUENCING IN PROG 3.36e+02
	931	15	10.2	188743	26	AC002542	Human BAC clone RG114A 3.36e+02
c	932	15	10.2	190000	17	AC004479	*** SEQUENCING IN PROG 3.36e+02
c	933	15	10.2	190248	17	AC004988	*** SEQUENCING IN PROG 3.36e+02
	934	15	10.2	191439	17	AC005040	*** SEQUENCING IN PROG 3.36e+02
c	935	15	10.2	192037	17	AC004486	*** SEQUENCING IN PROG 3.36e+02
	936	15	10.2	192853	17	AC005096	*** SEQUENCING IN PROG 3.36e+02
c	937	15	10.2	192853	17	AC005096	*** SEQUENCING IN PROG 3.36e+02
	938	15	10.2	193374	17	AC005139	*** SEQUENCING IN PROG 3.36e+02
c	939	15	10.2	195016	17	HS215F16	Human DNA sequence *** 3.36e+02
	940	15	10.2	195033	17	HS608E8	Human DNA sequence *** 3.36e+02
	941	15	10.2	196559	17	HS399M14	Human DNA sequence *** 3.36e+02
c	942	15	10.2	197391	17	AC005140	*** SEQUENCING IN PROG 3.36e+02
	943	15	10.2	199496	17	HS357K22	Human DNA sequence *** 3.36e+02
c	944	15	10.2	200349	27	HSU85195	Homo sapiens BAC129, c 3.36e+02
c	945	15	10.2	201964	28	MMXHC29N7	Mus musculus major his 3.36e+02
	946	15	10.2	202233	26	AC003664	Homo sapiens chromosom 3.36e+02
	947	15	10.2	204877	17	AC005375	*** SEQUENCING IN PROG 3.36e+02
c	948	15	10.2	205150	26	AC005274	Homo sapiens chromosom 3.36e+02
	949	15	10.2	205251	26	AC005392	Homo sapiens chromosom 3.36e+02
c	950	15	10.2	207931	17	HS101G11	Human DNA sequence *** 3.36e+02
c	951	15	10.2	207957	26	AC004470	Homo sapiens Xp22 BAC 3.36e+02
c	952	15	10.2	210000	17	AC004555	*** SEQUENCING IN PROG 3.36e+02
c	953	15	10.2	210276	17	HS1170K4	Human DNA sequence *** 3.36e+02
c	954	15	10.2	210364	17	AC005537	*** SEQUENCING IN PROG 3.36e+02
c	955	15	10.2	210964	17	AC005497	*** SEQUENCING IN PROG 3.36e+02
c	956	15	10.2	213680	15	BSUB0012	Bacillus subtilis comp 3.36e+02
	957	15	10.2	215286	26	HSU91322	Human chromosome 16p13 3.36e+02
c	958	15	10.2	215441	27	AC005736	Homo sapiens chromosom 3.36e+02
c	959	15	10.2	215640	15	BSUB0016	Bacillus subtilis comp 3.36e+02
c	960	15	10.2	216424	17	AC005414	*** SEQUENCING IN PROG 3.36e+02

c	961	15	10.2	218410	15	BSUB0015	Bacillus subtilis comp 3.36e+02
	962	15	10.2	219062	17	HS407F17	Human DNA sequence *** 3.36e+02
c	963	15	10.2	220046	17	AC005343	*** SEQUENCING IN PROG 3.36e+02
	964	15	10.2	220060	15	AF008220	Bacillus subtilis rrnB 3.36e+02
	965	15	10.2	220731	17	AC004671	*** SEQUENCING IN PROG 3.36e+02
	966	15	10.2	221160	15	BSUB0002	Bacillus subtilis comp 3.36e+02
c	967	15	10.2	222543	17	HS477J10	Human DNA sequence *** 3.36e+02
c	968	15	10.2	224129	17	CEY7A9	Caenorhabditis elegans 3.36e+02
	969	15	10.2	224681	17	HS44N10	Human DNA sequence *** 3.36e+02
c	970	15	10.2	224681	17	HS44N10	Human DNA sequence *** 3.36e+02
	971	15	10.2	225676	17	HS199L16	Human DNA sequence *** 3.36e+02
c	972	15	10.2	226916	27	AC005066	Homo sapiens BAC clone 3.36e+02
c	973	15	10.2	227215	17	HS245G19	Human DNA sequence *** 3.36e+02
	974	15	10.2	231329	17	AC002353	*** SEQUENCING IN PROG 3.36e+02
c	975	15	10.2	231329	17	AC002353	*** SEQUENCING IN PROG 3.36e+02
c	976	15	10.2	234053	26	AC002429	Human BAC clone GS200X 3.36e+02
c	977	15	10.2	236383	17	HS94M16	Human DNA sequence *** 3.36e+02
	978	15	10.2	240114	17	HS90L6	Human DNA sequence *** 3.36e+02
c	979	15	10.2	240114	17	HS90L6	Human DNA sequence *** 3.36e+02
	980	15	10.2	244649	17	AC005308	*** SEQUENCING IN PROG 3.36e+02
c	981	15	10.2	244649	17	AC005308	*** SEQUENCING IN PROG 3.36e+02
c	982	15	10.2	250529	26	HUAE000658	Homo sapiens T-cell re 3.36e+02
	983	15	10.2	250599	26	AF015262	Homo sapiens Down Synd 3.36e+02
	984	15	10.2	257506	17	HS1042K10	Human DNA sequence *** 3.36e+02
c	985	15	10.2	263833	17	AC005342	*** SEQUENCING IN PROG 3.36e+02
	986	15	10.2	281873	17	AC005533	*** SEQUENCING IN PROG 3.36e+02
	987	15	10.2	288002	26	HS229043	Homo sapiens 959 kb co 3.36e+02
	988	15	10.2	291015	17	CEY59A8	Caenorhabditis elegans 3.36e+02
	989	15	10.2	291821	17	CEY48E1	Caenorhabditis elegans 3.36e+02
	990	15	10.2	292828	17	AC004801	*** SEQUENCING IN PROG 3.36e+02
c	991	15	10.2	323000	26	HS229041	Homo sapiens 959 kb co 3.36e+02
c	992	15	10.2	348050	26	HS229042	Homo sapiens 959 kb co 3.36e+02
	993	14	9.5	392	30	HUM4S1340	Human Chromosome 4 (cl 1.33e+03
	994	14	9.5	463	30	HSU91783	Human chromosome 8 STS 1.33e+03
c	995	14	9.5	3673	33	VRANICAM	Puumala virus M segmen 1.33e+03
c	996	14	9.5	31786	18	CEC32C4	Caenorhabditis elegans 1.33e+03
	997	14	9.5	40978	18	CELC09B7	Caenorhabditis elegans 1.33e+03
c	998	14	9.5	103699	17	AC004556	*** SEQUENCING IN PROG 1.33e+03
c	999	14	9.5	133661	33	U93872	Kaposi's sarcoma-assoc 1.33e+03
c	1000	14	9.5	139958	27	HUAC002519	Human Chromosome 16 BA 1.33e+03

ALIGNMENTS

```
RESULT    1
LOCUS       AMU32452      310 bp     DNA              INV       18-APR-1996
DEFINITION  Apodemia mormo NADH dehydrogenase subunit 1 (ND1) gene,
            mitochondrial gene encoding mitochondrial protein, partial cds.
ACCESSION   U32452
NID         g1272199
KEYWORDS    .
```



```
SOURCE      Mormon metalmark.
  ORGANISM  Mitochondrion Apodemia mormo
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Arthropoda;
            Tracheata; Insecta; Pterygota; Lepidoptera; Papilionoidea;
            Lycaenidae; Riodininae; Apodemia.
REFERENCE   1  (bases 1 to 310)
  AUTHORS   Weller,S.J., Pashley,D.P. and Martin,J.A.
  TITLE     Reassessment of Butterfly Family Relationships using independent
            genes and morphological evidence
  JOURNAL   Ann. Entomol. Soc. Amer. (1995) In press
REFERENCE   2  (bases 1 to 310)
  AUTHORS   Weller,S.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUL-1995) Susan J. Weller, Department of Entomology,
            University of Minnesota, 219 Hodson Hall, 1980 Folwell Avenue,
            Saint Paul, MN 55108, USA
FEATURES              Location/Qualifiers
     source           1. .310
                      /organism="Apodemia mormo"
                      /mitochondrion
                      /db_xref="taxon:42267"
                      /map="mtDNA Drosophila yakuba 12,677-12,366 bp"
                      /tissue_type="muscle"
                      /dev_stage="adult"
     gene             1. .258
                      /gene="ND1"
     CDS              <1. .258
                      /gene="ND1"
                      /codon_start=1
                      /transl_table=5
                      /product="NADH dehydrogenase subunit 1"
                      /db_xref="PID:g1272200"
                      /translation="DIIVXIVGVLILLVGVLXXVAFLTLLXRKVLGYIQIRKGPNKLG
                      FMGMLQPFSDGIKLFTKEMXFPNYSNYLXYYISPILVLLYH"
BASE COUNT       93 a      22 c      36 g     147 t      12 others
ORIGIN

  Query Match           12.2%;  Score 18;  DB 18;  Length 310;
  Best Local Similarity 100.0%;  Pred. No. 3.50e+00;
  Matches    18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    278 TTATTATTTTAATATAAT 295
          ||||||||||||||||||
Cp    139 TTATTATTTTAATATAAT 122


RESULT    2
LOCUS       BPU32454      311 bp     DNA              INV       18-APR-1996
DEFINITION  Battus philenor NADH dehydrogenase subunit 1 (ND1) gene,
            mitochondrial gene encoding mitochondrial protein, partial cds.
ACCESSION   U32454
```



```
NID         g1272205
KEYWORDS    .
SOURCE      pipe-vine swallowtail.
  ORGANISM  Mitochondrion Battus philenor
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Arthropoda;
            Tracheata; Insecta; Pterygota; Lepidoptera; Papilionoidea;
            Papilionidae; Papilioninae; Battus.
REFERENCE   1  (bases 1 to 311)
  AUTHORS   Weller,S.J., Pashley,D.P. and Martin,J.A.
  TITLE     Reassessment of Butterfly Family Relationships using independent
            genes and morphological evidence
  JOURNAL   Ann. Entomol. Soc. Amer. (1995) In press
REFERENCE   2  (bases 1 to 311)
  AUTHORS   Weller,S.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUL-1995) Susan J. Weller, Department of Entomology,
            University of Minnesota, 219 Hodson Hall, 1980 Folwell Avenue,
            Saint Paul, MN 55108, USA
FEATURES              Location/Qualifiers
     source           1. .311
                      /organism="Battus philenor"
                      /mitochondrion
                      /db_xref="taxon:42288"
                      /map="mtDNA Drosophila yakuba 12,677-12,366 bp"
                      /tissue_type="muscle"
                      /dev_stage="adult"
     gene             1. .311
                      /gene="ND1"
     CDS              <1. .>311
                      /gene="ND1"
                      /codon_start=1
                      /transl_table=5
                      /product="NADH dehydrogenase subunit 1"
                      /db_xref="PID:g1272206"
                      /translation="DMMLFIIGMLIMIIGVLVGVAFLTLLERKVLGYIQIRKGPNKVG
                      FMGILQPFSDAIKLFTKEHTFPNYSNYMSYFSPVISFIISLMLWMLIPYYFNMISFN
                      LG"
BASE COUNT       99 a      25 c      36 g     151 t
ORIGIN

  Query Match           12.2%;  Score 18;  DB 18;  Length 311;
  Best Local Similarity 100.0%;  Pred. No. 3.50e+00;
  Matches    18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    279 TTATTATTTTAATATAAT 296
          ||||||||||||||||||
Cp    139 TTATTATTTTAATATAAT 122


RESULT    3
LOCUS       HMU25877      311 bp     DNA              INV       23-DEC-1996
```



```
DEFINITION  Hemileuca maia NADH dehydrogenase subunit 1 (ND1) gene,
            mitochondrial gene encoding mitochondrial protein, partial cds.
ACCESSION   U25877
NID         g1049033
KEYWORDS    .
SOURCE      buck moth.
  ORGANISM  Mitochondrion Hemileuca maia
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Arthropoda;
            Tracheata; Hexapoda; Insecta; Pterygota; Lepidoptera; Bombycoidea;
            Saturniidae; Hemileucinae; Hemileuca.
REFERENCE   1  (bases 1 to 311)
  AUTHORS   Weller,S.J. and Pashley,D.P.
  TITLE     In search of butterfly origins
  JOURNAL   Mol. Phylogenet. Evol. 4 (3), 235-246 (1995)
  MEDLINE   96111500
REFERENCE   2  (bases 1 to 311)
  AUTHORS   Weller,S.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-APR-1995) Department of Entomology, University of
            Minnesota, 219 Hodson Hall, 1980 Folwell Avenue, Saint Paul, MN
            55108, USA
FEATURES              Location/Qualifiers
     source           1. .311
                      /organism="Hemileuca maia"
                      /mitochondrion
                      /db_xref="taxon:40090"
                      /tissue_type="muscle"
                      /dev_stage="adult"
     gene             1. .311
                      /gene="ND1"
     CDS              <1. .>311
                      /gene="ND1"
                      /codon_start=1
                      /transl_table=5
                      /product="NADH dehydrogenase subunit 1"
                      /db_xref="PID:g1049034"
                      /translation="DYTMIFMGVIFLILGVLIGVAFLTLLERKVLGYIQLRKGPNKLG
                      ILGVXQPFSDAIKLFTKEQTYPNYSNYMCYYFSPIVSFVMSLFIWMLIPYYFNMISFN
                      LG"
BASE COUNT       95 a      29 c      45 g     141 t       1 others
ORIGIN

  Query Match           12.2%;  Score 18;  DB 18;  Length 311;
  Best Local Similarity 100.0%;  Pred. No. 3.50e+00;
  Matches    18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    279 TTATTATTTTAATATAAT 296
          ||||||||||||||||||
Cp    139 TTATTATTTTAATATAAT 122
```

```
RESULT    4
LOCUS       ISU32461      311 bp    DNA             INV       18-APR-1996
DEFINITION  Incisalia sp. NADH dehydrogenase subunit 1 (ND1) gene,
            mitochondrial gene encoding mitochondrial protein, partial cds.
ACCESSION   U32461
NID         g1272219
KEYWORDS    .
SOURCE      Incisalia sp.
  ORGANISM  Mitochondrion Incisalia sp.
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Arthropoda;
            Tracheata; Insecta; Pterygota; Lepidoptera; Papilionoidea;
            Lycaenidae; Theclinae; Incisalia.
REFERENCE   1  (bases 1 to 311)
  AUTHORS   Weller,S.J., Pashley,D.P. and Martin,J.A.
  TITLE     Reassessment of Butterfly Family Relationships using independent
            genes and morphological evidence
  JOURNAL   Ann. Entomol. Soc. Amer. (1995) In press
REFERENCE   2  (bases 1 to 311)
  AUTHORS   Weller,S.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUL-1995) Susan J. Weller, Department of Entomology,
            University of Minnesota, 219 Hodson Hall, 1980 Folwell Avenue,
            Saint Paul, MN 55108, USA
FEATURES             Location/Qualifiers
     source          1. .311
                     /organism="Incisalia sp."
                     /mitochondrion
                     /db_xref="taxon:42319"
                     /map="mtDNA Drosophila yakuba 12,677-12,366 bp"
                     /tissue_type="muscle"
                     /dev_stage="adult"
     gene            1. .311
                     /gene="ND1"
     CDS             <1. .>311
                     /gene="ND1"
                     /codon_start=1
                     /transl_table=5
                     /product="NADH dehydrogenase subunit 1"
                     /db_xref="PID:g1272220"
                     /translation="DLMNIIIGIXILIIGVMISLAFLTLLERKVLSYMQIRKGPNKLG
                     PLGMLQPFSDGIKLFTKEMIYLNSSNYLFYLSPIIGPILSLMVWLLIPYYFNMISFD
                     LG"
BASE COUNT      106 a     23 c     32 g    149 t      1 others
ORIGIN

  Query Match           12.2%;  Score 18;  DB 18;  Length 311;
  Best Local Similarity 100.0%;  Pred. No. 3.50e+00;
  Matches    18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    279 TTATTATTTTAATATAAT 296
          ||||||||||||||||||
```



```
Cp    139 TTATTATTTTAATATAAT 122


RESULT    5
LOCUS       PCU32464      311 bp    DNA             INV       18-APR-1996
DEFINITION  Parnassius clodius NADH dehydrogenase subunit 1 (ND1) gene,
            mitochondrial gene encoding mitochondrial protein, partial cds.
ACCESSION   U32464
NID         g1272225
KEYWORDS    .
SOURCE      Clodius parnassian.
  ORGANISM  Mitochondrion Parnassius clodius
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Arthropoda;
            Tracheata; Insecta; Pterygota; Lepidoptera; Papilionoidea;
            Papilionidae; Parnassinae; Parnassius.
REFERENCE   1  (bases 1 to 311)
  AUTHORS   Weller,S.J., Pashley,D.P. and Martin,J.A.
  TITLE     Reassessment of Butterfly Family Relationships using independent
            genes and morphological evidence
  JOURNAL   Ann. Entomol. Soc. Amer. (1995) In press
REFERENCE   2  (bases 1 to 311)
  AUTHORS   Weller,S.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUL-1995) Susan J. Weller, Department of Entomology,
            University of Minnesota, 219 Hodson Hall, 1980 Folwell Avenue,
            Saint Paul, MN 55108, USA
FEATURES             Location/Qualifiers
     source          1. .311
                     /organism="Parnassius clodius"
                     /mitochondrion
                     /db_xref="taxon:42292"
                     /map="mtDNA Drosophila yakuba 12,677-12,366 bp"
                     /tissue_type="muscle"
                     /dev_stage="adult"
     gene            1. .311
                     /gene="ND1"
     CDS             <1. .>311
                     /gene="ND1"
                     /codon_start=1
                     /transl_table=5
                     /product="NADH dehydrogenase subunit 1"
                     /db_xref="PID:g1272226"
                     /translation="DMMMMILGMLILILGVLIGVAFLTLLERKVLGYIQIRKGPNKVG
                     YMGILQPFSDAIKLFTKEQVFPMYSNYISYYFSPIISFILSLMVWMLIPYYFNMISFN
                     LG"
BASE COUNT      102 a     28 c     37 g    144 t
ORIGIN

  Query Match           12.2%;  Score 18;  DB 18;  Length 311;
  Best Local Similarity 100.0%;  Pred. No. 3.50e+00;
  Matches    18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```



```
Db    279 TTATTATTTTAATATAAT 296
          ||||||||||||||||||
Cp    139 TTATTATTTTAATATAAT 122


RESULT    6
LOCUS       ISU32462      311 bp    DNA             INV       18-APR-1996
DEFINITION  Ithomia sp. NADH dehydrogenase subunit 1 (ND1) gene, mitochondrial
            gene encoding mitochondrial protein, partial cds.
ACCESSION   U32462
NID         g1272221
KEYWORDS    .
SOURCE      Ithomia sp.
  ORGANISM  Mitochondrion Ithomia sp.
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Arthropoda;
            Tracheata; Insecta; Pterygota; Lepidoptera; Papilionoidea;
            Nymphalidae; Ithomiinae; Ithomia.
REFERENCE   1  (bases 1 to 311)
  AUTHORS   Weller,S.J., Pashley,D.P. and Martin,J.A.
  TITLE     Reassessment of Butterfly Family Relationships using independent
            genes and morphological evidence
  JOURNAL   Ann. Entomol. Soc. Amer. (1995) In press
REFERENCE   2  (bases 1 to 311)
  AUTHORS   Weller,S.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUL-1995) Susan J. Weller, Department of Entomology,
            University of Minnesota, 219 Hodson Hall, 1980 Folwell Avenue,
            Saint Paul, MN 55108, USA
FEATURES             Location/Qualifiers
     source          1. .311
                     /organism="Ithomia sp."
                     /mitochondrion
                     /db_xref="taxon:42317"
                     /map="mtDNA Drosophila yakuba 12,677-12,366 bp"
                     /tissue_type="muscle"
                     /dev_stage="adult"
     gene            1. .311
                     /gene="ND1"
     CDS             <1. .>311
                     /gene="ND1"
                     /codon_start=1
                     /transl_table=5
                     /product="NADH dehydrogenase subunit 1"
                     /db_xref="PID:g1272222"
                     /translation="DLFNMIIGGLILVIGVLVGVAPLTLLERKVLGYIQIRKGPNKMG
                     ILGKLQPFSDAIKLFSKEQVTLNTSNYISYYISPIMSFTMTLIIWMVXPYYFNMISFN
                     LG"
BASE COUNT      101 a     25 c     40 g    142 t      3 others
ORIGIN
```



```
  Query Match           12.2%;  Score 18;  DB 18;  Length 311;
  Best Local Similarity 100.0%;  Pred. No. 3.50e+00;
  Matches    18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    279 TTATTATTTTAATATAAT 296
          ||||||||||||||||||
Cp    139 TTATTATTTTAATATAAT 122


RESULT    7
LOCUS       EPU32458      311 bp    DNA             INV       18-APR-1996
DEFINITION  Enodia portlandia NADH dehydrogenase subunit 1 (ND1) gene,
            mitochondrial gene encoding mitochondrial protein, partial cds.
ACCESSION   U32458
NID         g1272213
KEYWORDS    .
SOURCE      pearly eye.
  ORGANISM  Mitochondrion Enodia portlandia
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Arthropoda;
            Tracheata; Insecta; Pterygota; Lepidoptera; Papilionoidea;
            Nymphalidae; Satyrinae; Enodia.
REFERENCE   1  (bases 1 to 311)
  AUTHORS   Weller,S.J., Pashley,D.P. and Martin,J.A.
  TITLE     Reassessment of Butterfly Family Relationships using independent
            genes and morphological evidence
  JOURNAL   Ann. Entomol. Soc. Amer. (1995) In press
REFERENCE   2  (bases 1 to 311)
  AUTHORS   Weller,S.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUL-1995) Susan J. Weller, Department of Entomology,
            University of Minnesota, 219 Hodson Hall, 1980 Folwell Avenue,
            Saint Paul, MN 55108, USA
FEATURES             Location/Qualifiers
     source          1. .311
                     /organism="Enodia portlandia"
                     /mitochondrion
                     /db_xref="taxon:42284"
                     /map="mtDNA Drosophila yakuba 12,677-12,366 bp"
                     /tissue_type="muscle"
                     /dev_stage="adult"
     gene            1. .311
                     /gene="ND1"
     CDS             <1. .>311
                     /gene="ND1"
                     /codon_start=1
                     /transl_table=5
                     /product="NADH dehydrogenase subunit 1"
                     /db_xref="PID:g1272214"
                     /translation="DLVNMIIGFLILIIGVLIGVAFLLLLERKVLGYIQIRKGPNKLG
                     FMGLLQPFSDAIKLFTKEQVYLNYSNYMFYYFSPVLSFMLSLMIWMVIPYYFNMIMFN
                     LG"
```

```
BASE COUNT      101 a     22 c     39 g    149 t
ORIGIN

  Query Match           12.2%;  Score 18;  DB 18;  Length 311;
  Best Local Similarity 100.0%;  Pred. No. 3.50e+00;
  Matches    18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    279 TTATTATTTTAATATAAT 296
          ||||||||||||||||||
Cp    139 TTATTATTTTAATATAAT 122


RESULT    8
LOCUS       AAU32451      311 bp    DNA              INV       18-APR-1996
DEFINITION  Anaea andria NADH dehydrogenase subunit 1 (ND1) gene, mitochondrial
            gene encoding mitochondrial protein, partial cds.
ACCESSION   U32451
NID         g1272197
KEYWORDS    .
SOURCE      goatweed butterfly.
  ORGANISM  Mitochondrion Anaea andria
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Arthropoda;
            Tracheata; Insecta; Pterygota; Lepidoptera; Papilionoidea;
            Nymphalidae; Charaxinae; Anaea.
REFERENCE   1  (bases 1 to 311)
  AUTHORS   Weller,S.J., Pashley,D.P. and Martin,J.A.
  TITLE     Reassessment of Butterfly Family Relationships using independent
            genes and morphological evidence
  JOURNAL   Ann. Entomol. Soc. Amer. (1995) In press
REFERENCE   2  (bases 1 to 311)
  AUTHORS   Weller,S.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUL-1995) Susan J. Weller, Department of Entomology,
            University of Minnesota, 219 Hodson Hall, 1980 Folwell Avenue,
            Saint Paul, MN 55108, USA
FEATURES             Location/Qualifiers
     source          1. .311
                     /organism="Anaea andria"
                     /mitochondrion
                     /db_xref="taxon:42263"
                     /map="mtDNA Drosophila yakuba 12,677-12,366 bp"
                     /tissue_type="muscle"
                     /dev_stage="adult"
     gene            1. .311
                     /gene="ND1"
     CDS             <1. .>311
                     /gene="ND1"
                     /codon_start=1
                     /transl_table=5
                     /product="NADH dehydrogenase subunit 1"
                     /db_xref="PID:g1272198"
```



```
                     /translation="DLLNIIIGMLILIIGVLIGVAFLVLLERKVLGYIQIRKGPNKVG
                     MMGLLQPFSDAIKLFTKEQIYLNYSNYLFYYFSPVLSFMXGXMSWMMIPYYFNMIMFX
                     LG"
BASE COUNT      98 a     21 c     38 g    149 t      5 others
ORIGIN

  Query Match           12.2%;  Score 18;  DB 18;  Length 311;
  Best Local Similarity 100.0%;  Pred. No. 3.50e+00;
  Matches    18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    279 TTATTATTTTAATATAAT 296
          ||||||||||||||||||
Cp    139 TTATTATTTTAATATAAT 122


RESULT    9
LOCUS       DSU25876      311 bp    DNA              INV       23-DEC-1996
DEFINITION  Diatraea saccharalis NADH dehydrogenase subunit 1 (ND1) gene,
            mitochondrial gene encoding mitochondrial protein, partial cds.
ACCESSION   U25876
NID         g1049029
KEYWORDS    .
SOURCE      sugarcane borer.
  ORGANISM  Mitochondrion Diatraea saccharalis
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Arthropoda;
            Tracheata; Hexapoda; Insecta; Pterygota; Lepidoptera; Pyraloidea;
            Pyralidae; Crambinae; Diatraea.
REFERENCE   1  (bases 1 to 311)
  AUTHORS   Weller,S.J. and Pashley,D.P.
  TITLE     In search of butterfly origins
  JOURNAL   Mol. Phylogenet. Evol. 4 (3), 235-246 (1995)
  MEDLINE   96111500
REFERENCE   2  (bases 1 to 311)
  AUTHORS   Weller,S.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-APR-1995) Department of Entomology, University of
            Minnesota, 219 Hodson Hall, 1980 Folwell Avenue, Saint Paul, MN
            55108, USA
FEATURES             Location/Qualifiers
     source          1. .311
                     /organism="Diatraea saccharalis"
                     /mitochondrion
                     /db_xref="taxon:40085"
                     /tissue_type="muscle"
                     /dev_stage="adult"
     gene            1. .311
                     /gene="ND1"
     CDS             <1. .>311
                     /gene="ND1"
                     /codon_start=1
                     /transl_table=5
```



```
                     /product="NADH dehydrogenase subunit 1"
                     /db_xref="PID:g1049030"
                     /translation="DLLMVLMGLIILILGVLIGVAFLTLLERKVLGYIQIRKGPNKVG
                     IIGILQPFSDGIKLFTKEQTYPMFSNYYSYYFSPVISFILSLMVWMLIPYYFNMIXFN
                     L"
BASE COUNT      101 a     26 c     38 g    143 t      3 others
ORIGIN

  Query Match           12.2%;  Score 18;  DB 18;  Length 311;
  Best Local Similarity 100.0%;  Pred. No. 3.50e+00;
  Matches    18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    279 TTATTATTTTAATATAAT 296
          ||||||||||||||||||
Cp    139 TTATTATTTTAATATAAT 122


RESULT   10
LOCUS       ACU25881      311 bp    DNA              INV       23-DEC-1996
DEFINITION  Agrius cingulata NADH dehydrogenase subunit 1 (ND1) gene,
            mitochondrial gene encoding mitochondrial protein, partial cds.
ACCESSION   U25881
NID         g1049027
KEYWORDS    .
SOURCE      sweetpotato hornworm.
  ORGANISM  Mitochondrion Agrius cingulata
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Arthropoda;
            Tracheata; Hexapoda; Insecta; Pterygota; Lepidoptera; Sphingiodea;
            Sphingidae; Agrius.
REFERENCE   1  (bases 1 to 311)
  AUTHORS   Weller,S.J. and Pashley,D.P.
  TITLE     In search of butterfly origins
  JOURNAL   Mol. Phylogenet. Evol. 4 (3), 235-246 (1995)
  MEDLINE   96111500
REFERENCE   2  (bases 1 to 311)
  AUTHORS   Weller,S.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-APR-1995) Department of Entomology, University of
            Minnesota, 219 Hodson Hall, 1980 Folwell Avenue, Saint Paul, MN
            55108, USA
FEATURES             Location/Qualifiers
     source          1. .311
                     /organism="Agrius cingulata"
                     /mitochondrion
                     /db_xref="taxon:40104"
                     /tissue_type="muscle"
                     /dev_stage="adult"
     gene            1. .311
                     /gene="ND1"
     CDS             <1. .>311
                     /gene="ND1"
```



```
                     /codon_start=1
                     /transl_table=5
                     /product="NADH dehydrogenase subunit 1"
                     /db_xref="PID:g1049028"
                     /translation="DLFMVMIGLLILILGVLIGVAFLVLLERKVLGYIQLRKGPNKVG
                     FLGIMQPFSDAIKLFSKETIYPNFSNIYCIYFSPVVGFILSLLIWMLVPYYFNMIXFN
                     L"
BASE COUNT      87 a     24 c     42 g    154 t      4 others
ORIGIN

  Query Match           12.2%;  Score 18;  DB 18;  Length 311;
  Best Local Similarity 100.0%;  Pred. No. 3.50e+00;
  Matches    18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    279 TTATTATTTTAATATAAT 296
          ||||||||||||||||||
Cp    139 TTATTATTTTAATATAAT 122


RESULT   11
LOCUS       HCU25873      311 bp    DNA              INV       23-DEC-1996
DEFINITION  Hyalophora cecropia NADH dehydrogenase subunit 1 (ND1) gene,
            mitochondrial gene encoding mitochondrial protein, partial cds.
ACCESSION   U25873
NID         g1049031
KEYWORDS    .
SOURCE      cecropia moth.
  ORGANISM  Mitochondrion Hyalophora cecropia
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Arthropoda;
            Tracheata; Hexapoda; Insecta; Pterygota; Lepidoptera; Bombycoidea;
            Saturniidae; Hyalophora.
REFERENCE   1  (bases 1 to 311)
  AUTHORS   Weller,S.J. and Pashley,D.P.
  TITLE     In search of butterfly origins
  JOURNAL   Mol. Phylogenet. Evol. 4 (3), 235-246 (1995)
  MEDLINE   96111500
REFERENCE   2  (bases 1 to 311)
  AUTHORS   Weller,S.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-APR-1995) Department of Entomology, University of
            Minnesota, 219 Hodson Hall, 1980 Folwell Avenue, Saint Paul, MN
            55108, USA
FEATURES             Location/Qualifiers
     source          1. .311
                     /organism="Hyalophora cecropia"
                     /mitochondrion
                     /db_xref="taxon:7123"
                     /tissue_type="muscle"
                     /dev_stage="adult"
     gene            1. .311
                     /gene="ND1"
```

```
     CDS             <1. .>311
                     /gene="ND1"
                     /codon_start=1
                     /transl_table=5
                     /product="NADH dehydrogenase subunit 1"
                     /db_xref="PID:g1049032"
                     /translation="DLYLVFIGLLILILGVMIGVAPLTLLERKVLGYIQIRKGPNKLG
                     MMGIMQPFSDAIKLFTKEQTTPNYSNYPCYYFSPVISFLMSLLIWMLIPYYFNMISFN
                     LG"
BASE COUNT       95 a     25 c     45 g    145 t      1 others
ORIGIN

  Query Match           12.2%;  Score 18;  DB 18;  Length 311;
  Best Local Similarity 100.0%;  Pred. No. 3.50e+00;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    279 TTATTATTTTAATATAAT 296
          ||||||||||||||||||
Cp    139 TTATTATTTTAATATAAT 122


RESULT   12
LOCUS       PCU25880      311 bp    DNA             INV       23-DEC-1996
DEFINITION  Pyrgus communis NADH dehydrogenase subunit 1 (ND1) gene,
            mitochondrial gene encoding mitochondrial protein, partial cds.
ACCESSION   U25880
NID         g1049043
KEYWORDS    .
SOURCE      checkered skipper.
  ORGANISM  Mitochondrion Pyrgus communis
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Arthropoda;
            Tracheata; Hexapoda; Insecta; Pterygota; Lepidoptera; Hesperioidea;
            Hesperiidae; Pyrginae; Pyrgus.
REFERENCE   1  (bases 1 to 311)
  AUTHORS   Weller,S.J. and Pashley,D.P.
  TITLE     In search of butterfly origins
  JOURNAL   Mol. Phylogenet. Evol. 4 (3), 235-246 (1995)
  MEDLINE   96111500
REFERENCE   2  (bases 1 to 311)
  AUTHORS   Weller,S.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-APR-1995) Department of Entomology, University of
            Minnesota, 219 Hodson Hall, 1980 Folwell Avenue, Saint Paul, MN
            55108, USA
FEATURES             Location/Qualifiers
     source          1. .311
                     /organism="Pyrgus communis"
                     /mitochondrion
                     /db_xref="taxon:40101"
                     /tissue_type="muscle"
                     /dev_stage="adult"
```



```
     gene            1. .309
                     /gene="ND1"
     CDS             <1. .309
                     /gene="ND1"
                     /codon_start=1
                     /transl_table=5
                     /product="NADH dehydrogenase subunit 1"
                     /db_xref="PID:g1049044"
                     /translation="DIFNIMIGILILIIGVMIGVAPLTLLERKVLGYIQIRKGPNKVG
                     FMGLLQPFSDAIKLFTKEQMTLNYSNYLIYYFSPIMSFILSLMIWMLIPYYFNMISFN
                     "
BASE COUNT      100 a     21 c     37 g    152 t      1 others
ORIGIN

  Query Match           12.2%;  Score 18;  DB 18;  Length 311;
  Best Local Similarity 100.0%;  Pred. No. 3.50e+00;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    279 TTATTATTTTAATATAAT 296
          ||||||||||||||||||
Cp    139 TTATTATTTTAATATAAT 122


RESULT   13
LOCUS       A30910        731 bp    DNA             PAT       04-AUG-1995
DEFINITION  D.discoideum DNA for ras promoter region from patent WO9220806.
ACCESSION   A30910
NID         g1246930
KEYWORDS    .
SOURCE      Dictyostelium discoideum.
  ORGANISM  Dictyostelium discoideum
            Eukaryotae; mitochondrial eukaryotes; Dictyosteliida;
            Dictyostelium.
REFERENCE   1  (bases 1 to 731)
  AUTHORS   .
  TITLE     DICTYOSTELID EXPRESSION VECTOR AND METHOD FOR EXPRESSING A DESIRED
            PROTEIN
  JOURNAL   Patent: WO 9220806-A 3 26-NOV-1992;
FEATURES             Location/Qualifiers
     source          1. .731
                     /organism="Dictyostelium discoideum"
                     /db_xref="taxon:44689"
BASE COUNT      317 a     52 c     42 g    320 t
ORIGIN

  Query Match           12.2%;  Score 18;  DB 21;  Length 731;
  Best Local Similarity 100.0%;  Pred. No. 3.50e+00;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    626 TTTGTACATTATTATTT 643
          |||||||||||||||||
```



```
Cp    147 TTTGTACATTATTATTT 130


RESULT   14
LOCUS       DDDDRAS       731 bp    DNA             INV       14-NOV-1991
DEFINITION  D.discoideum Dd ras gene for ras protein (promoter region).
ACCESSION   X54630
NID         g7254
KEYWORDS    Dd ras gene; promoter region; ras gene; ras-protein.
SOURCE      Dictyostelium discoideum.
  ORGANISM  Dictyostelium discoideum
            Eukaryotae; mitochondrial eukaryotes; Dictyosteliida;
            Dictyostelium.
REFERENCE   1  (bases 1 to 731)
  AUTHORS   Reymond,C.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-AUG-1990) Reymond C.D., ISREC, CH-1066 Epalinges,
            Switzerland
REFERENCE   2  (bases 1 to 731)
  AUTHORS   Louvion,J.F., Scholder,J.C., Pinaud,S. and Reymond,C.D.
  TITLE     Two independent promoters as well as 5' untranslated regions
            regulate Dd ras expression in Dictyostelium
  JOURNAL   Nucleic Acids Res. 19 (22), 6133-6138 (1991)
  MEDLINE   92066459
COMMENT     For related sequence see K02114.
FEATURES             Location/Qualifiers
     source          1. .731
                     /organism="Dictyostelium discoideum"
                     /strain="Ax3 K"
                     /db_xref="taxon:44689"
                     /germline
                     /cell_line="Ax3"
     prim_transcript 369. .>731
                     /gene="Dd ras"
                     /note="transcript 1"
     gene            369. .731
                     /gene="Dd ras"
     prim_transcript 490. .>731
                     /gene="Dd ras"
                     /note="transcript 2"
     prim_transcript 547. .>731
                     /gene="Dd ras"
                     /note="transcript 3"
     CDS             729. .>731
                     /gene="Dd ras"
                     /codon_start=1
                     /product="ras"
BASE COUNT      317 a     52 c     42 g    320 t
ORIGIN

  Query Match           12.2%;  Score 18;  DB 18;  Length 731;
```



```
  Best Local Similarity 100.0%;  Pred. No. 3.50e+00;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    626 TTTGTACATTATTATTT 643
          |||||||||||||||||
Cp    147 TTTGTACATTATTATTT 130


RESULT   15
LOCUS       S46989       1119 bp    mRNA            PLN       05-JAN-1993
DEFINITION  chalcone synthase [soybeans, mRNA, 1119 nt].
ACCESSION   S46989
NID         g257815
KEYWORDS    .
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryotae; mitochondrial eukaryotes; Viridiplantae;
            Charophyta/Embryophyta group; Embryophyta; Magnoliophyta;
            Magnoliopsida; Rutanae; Sapindales; Fabaceae; Papilionoideae;
            Glycine.
REFERENCE   1  (bases 1 to 1119)
  AUTHORS   Estabrook,E.M. and Sengupta-Gopalan,C.
  TITLE     Differential expression of phenylalanine ammonia-lyase and chalcone
            synthase during soybean nodule development
  JOURNAL   Plant Cell 3 (3), 299-308 (1991)
  MEDLINE   93005700
  REMARK    GenBank staff at the National Library of Medicine created this
            entry [NCBI gibbsq 116406] from the original journal article.
            This sequence comes from Fig. 1B.
FEATURES             Location/Qualifiers
     source          1. .1119
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
     gene            1. .1119
                     /gene="chalcone synthase"
BASE COUNT      313 a     216 c     277 g     313 t
ORIGIN

  Query Match           12.2%;  Score 18;  DB 23;  Length 1119;
  Best Local Similarity 100.0%;  Pred. No. 3.50e+00;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1041 TTATATTAAAATAATAAT 1058
          ||||||||||||||||||
Qy    123 TTATATTAAAATAATAAT 140


RESULT   16
LOCUS       DDRASX       1676 bp    DNA             INV       17-FEB-1997
DEFINITION  D.discoideum ras gene.
ACCESSION   Z11804 K02114 X58190
```

```
NID             g7346
KEYWORDS        ras cellular oncogene.
SOURCE          Dictyostelium discoideum.
  ORGANISM      Dictyostelium discoideum
                Eukaryotae; mitochondrial eukaryotes; Dictyosteliida;
                Dictyostelium.
REFERENCE       1  (bases 1 to 735)
  AUTHORS       Esch,R.K. and Firtel,R.A.
  TITLE         cAMP and cell sorting control the spatial expression of a
                developmentally essential cell-type-specific ras gene in
                Dictyostelium
  JOURNAL       Genes Dev. 5 (1), 9-21 (1991)
  MEDLINE       91115102
REFERENCE       2  (bases 643 to 1676)
  AUTHORS       Reymond,C.D., Gomer,R.H., Mehdy,M.C. and Firtel,R.A.
  TITLE         Developmental regulation of a Dictyostelium gene encoding a protein
                homologous to mammalian ras protein
  JOURNAL       Cell 39 (1), 141-148 (1984)
  MEDLINE       85024887
REFERENCE       3  (bases 1 to 1676)
  AUTHORS       Firtel,R.A.
  TITLE         Direct Submission
  JOURNAL       Submitted (27-DEC-1991) Richard A Firtel, Biology, University of
                California - San Diego, Center For Molecular Genetics, 9500 Gilman
                Drive, La Jolla, CA, 92093-0634, USA
  REMARK        Updates [2]
FEATURES                Location/Qualifiers
     source             1. .1676
                        /organism="Dictyostelium discoideum"
                        /strain="KAx-3"
                        /db_xref="taxon:44689"
                        /dev_stage="aggregate"
                        /cell_type="vegetative and prestalk"
                        /clone_lib="cDNA and genomic"
                        /clone="DdrasD"
     CDS                join(701. .775,871. .883,984. .1034,1129. .1553)
                        /codon_start=1
                        /product="ras protein"
                        /db_xref="PID:g7347"
                        /db_xref="SWISS-PROT:P03967"
                        /translation="MTEYKLVIVGGGGVGKSALTIQLIQNHFIDEYDPTIEDSYRKQV
                        SIDDETCLLDILDTAGQEEYSAMRDQYMRTGQGFLCVYSITSRSSYDEIASFREQILR
                        VKDKDRVPLILVGNKADLDHERQVSVNEGQSLAKGFNCPFMESSAKSRINVEEAFYSL
                        VREIRKELKGDQSSGKAQKKKKQCLIL"
     exon               701. .775
                        /number=1
     intron             776. .870
                        /number=1
     old_sequence       791. .792
                        /citation=[2]
                        /replace="ttttat"
```



```
     old_sequence       850. .851
                        /citation=[2]
                        /replace="atattt"
     exon               871. .883
                        /number=2
     intron             884. .983
                        /number=2
     exon               984. .1034
                        /number=3
     intron             1035. .1128
                        /number=3
     old_sequence       1040
                        /citation=[2]
                        /replace=""
     exon               1129. .1553
                        /number=4
     old_sequence       1394
                        /citation=[2]
                        /replace=""
     old_sequence       1404
                        /citation=[2]
                        /replace=""
     old_sequence       1410
                        /citation=[2]
                        /replace=""
     old_sequence       1579. .1580
                        /citation=[2]
                        /replace="caa"
BASE COUNT      724 a    118 c    160 g    674 t
ORIGIN

  Query Match           12.2%;  Score 18;  DB 18;  Length 1676;
  Best Local Similarity 100.0%;  Pred. No. 3.50e+00;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    598 TTTGTACATTATTATTTT 615
          ||||||||||||||||||
Cp    147 TTTGTACATTATTATTTT 130


RESULT   17
LOCUS       SPFMTTGRG     2256 bp    DNA             INV       25-JAN-1993
DEFINITION  Spodoptera frugiperda 16S rRNA gene, Val-tRNA, and Leu-tRNA genes,
            and ND-1 protein gene, 5' end.
ACCESSION   M76713
NID         g343352
KEYWORDS    16S ribosomal RNA; ND-1 protein; transfer RNA-Leu; transfer
            RNA-Val.
SOURCE      Mitochondrion Spodoptera frugiperda (strain RICE, organelle
            Mitochondrion Spodoptera frugiperda) DNA.
  ORGANISM  Mitochondrion Spodoptera frugiperda
```



```
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Arthropoda;
            Tracheata; Insecta; Pterygota; Lepidoptera; Noctuoidea; Noctuidae;
            Spodoptera.
REFERENCE   1  (bases 1 to 2256)
  AUTHORS   Pashley,D.P. and Ke,L.D.
  TITLE     Sequence evolution in the large ribosomal and ND-1 genes in a
            lepidopteran mitochondrial genome: Implications for phylogenetic
            analyses
  JOURNAL   Mol. Biol. Evol. 9, 1061-1075 (1992)
  MEDLINE   93061985
FEATURES            Location/Qualifiers
     source         1. .2256
                    /organism="Spodoptera frugiperda"
                    /mitochondrion
                    /strain="RICE"
                    /db_xref="taxon:7108"
     tRNA           11. .75
                    /gene="tRNA-Val"
                    /product="transfer RNA-Val"
     gene           11. .75
                    /gene="tRNA-Val"
     rRNA           76. .1412
                    /gene="16S rRNA"
                    /note="5' end approximate"
                    /product="16S ribosomal RNA"
     gene           76. .1412
                    /gene="16S rRNA"
     tRNA           1477. .1546
                    /gene="tRNA-Leu"
                    /product="transfer RNA-Leu"
     gene           1477. .1546
                    /gene="tRNA-Leu"
     gene           1553. .2256
                    /gene="ND-1"
     CDS            1553. .2256
                    /partial
                    /gene="ND-1"
                    /codon_start=1
                    /transl_table=5
                    /product="ND-1 protein"
                    /db_xref="PID:g552886"
                    /translation="MNDIYMIPLGLLILIIGILVGVAYLTLLERKVLGYIQIRKGPNK
                    VGFMGILQPFSDAIKLFTKEQTYPNFSNYLSYYFSPILSFILSLMIWLLIPYYFNMIS
                    FNLGILFFLCCTSLGVYTVMVAGWSSNSNYALLGGLRAVAQTISYEVSLALILMSSIV
                    MIMDFNLMMFFVYQNYVWFIILMFPLSLCWMSSSLAETNRTPFDFAEGESELVSGFNI
                    EYSSGGPALIFLAEYSS"
BASE COUNT      883 a    126 c    264 g    983 t
ORIGIN

  Query Match           12.2%;  Score 18;  DB 18;  Length 2256;
  Best Local Similarity 100.0%;  Pred. No. 3.50e+00;
```



```
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1837 TTATTATTTTAATATAAT 1854
          ||||||||||||||||||
Cp    139 TTATTATTTTAATATAAT 122


RESULT   18
LOCUS       SOYCHS7A     2688 bp    DNA             PLN       21-JUN-1994
DEFINITION  Glycine max chalcone synthase (chs7) gene, complete cds.
ACCESSION   M98871
NID         g169936
KEYWORDS    chalcone synthase.
SOURCE      Glycine max (strain Williams) DNA.
  ORGANISM  Glycine max
            Eukaryotae; mitochondrial eukaryotes; Viridiplantae;
            Charophyta/Embryophyta group; Embryophyta; Magnoliophyta;
            Magnoliopsida; Rutanae; Sapindales; Fabaceae; Papilionoideae;
            Glycine.
REFERENCE   1  (bases 1 to 2688)
  AUTHORS   Akada,S., Kung,S.D. and Dube,S.K.
  TITLE     Nucleotide sequence and putative regulatory elements of a
            nodule-development-specific member of the soybean (Glycine max)
            chalcone synthase multigene family, Gmchs 7
  JOURNAL   Plant Physiol. 102, 321-323 (1993)
  MEDLINE   94151429
FEATURES            Location/Qualifiers
     source         1. .2688
                    /organism="Glycine max"
                    /strain="Williams"
                    /db_xref="taxon:3847"
     TATA_signal    678. .684
                    /gene="chs7"
     gene           join(798. .975,1421. .2412)
                    /gene="chs7"
     CDS            join(798. .975,1421. .2412)
                    /gene="chs7"
                    /codon_start=1
                    /product="chalcone synthase"
                    /db_xref="PID:g169937"
                    /translation="MVSVAEIRQAQRAEGPATILAIGTANPPNRVDQSTYPDYYFRIT
                    NSDHMTELKEKFQRMCDKSMIRTRYMYLNEELLKENPNMCAYMAPSLDARQDMVVVEV
                    PKLGKEAAVKAIKEWGQPKSKITHLIFCTTSGVDMPGADYQLTKQLGLRPFVKRYMMY
                    QQGCFAGGTVLRLAKDLAENNKGARVLVVCSEITAVTFRGPSDTHLDSLVGQALFGDG
                    AAAVIVGSDPIPQVEKPLYELVWTAQTIAPDSEGAIDGHLREVGLTFHLLKDVPGIVS
                    KNIDKALFEAFNPLNISDYNSIFWIAHPGGPAILDQVEQKLGLKPEKMKATRDVLSEY
                    GNMSSACVLFILDEMRRKSAENGHKTTGEGLEWGVLFGFGPGLTIETVVLHSVAI"
     intron         976. .1420
                    /gene="chs7"
BASE COUNT      853 a    476 c    510 g    849 t
ORIGIN
```

```
  Query Match          12.2%;  Score 18;  DB 23;  Length 2688;
  Best Local Similarity 100.0%;  Pred. No. 3.50e+00;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2526 TTATATTAAAATAATAAT 2543
          ||||||||||||||||||
Qy    123 TTATATTAAAATAATAAT 140


RESULT   19
LOCUS       CEC49G9       6393 bp    DNA              INV       21-SEP-1998
DEFINITION  Caenorhabditis elegans cosmid C49G9, complete sequence.
ACCESSION   AL022593
NID         g3642157
KEYWORDS    HTG.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 6393)
  AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
  TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
  JOURNAL   Nature 368 (6466), 32-38 (1994)
  MEDLINE   94150718
REFERENCE   2  (bases 1 to 6393)
  AUTHORS   Lennard,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-NOV-1996) Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     On Sep 22, 1998 this sequence version replaced gi:3080429.
            Coding sequences below are predicted from computer analysis, using
            predictions from Genefinder (P. Green, U. Washington), and other
            available information.
            For a graphical representation of this sequence and its analysis
            see:-
            http://weface.sanger.ac.uk/cgi-
            bin/display?db=wormace&class=Sequence &object=C49G9
            Current sequence finishing criteria for the C. elegans genome
            sequencing consortium are that all bases are either sequenced
```



```
            unambiguously on both strands, or on a single strand with both a
            dye primer and dye terminator reaction, from distinct subclones.
            Exceptions are indicated by an explicit note.
            IMPORTANT:  This sequence is NOT necessarily the entire insert of
            the specified clone.  It may be shorter because we only sequence
            overlapping sections once, or longer because we arrange for a small
            overlap between neighbouring submissions.
            IMPORTANT: This sequence is not the entire insert of clone C49G9.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we arrange for a small overlap between
            neighbouring submissions.
            The true left end of clone F33E2 is at 6282 in this sequence. The
            true right end of clone W02D9 is at 109 in this sequence. The end
            of this sequence (6282. .6393) overlaps with the start of sequence
            Z84574.
FEATURES             Location/Qualifiers
     source          1. .6393
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /chromosome="I"
                     /clone="C49G9"
     gene            2256. .3539
                     /gene="C49G9.1"
     CDS             join(2256. .2378,3190. .3425,3509. .3539)
                     /gene="C49G9.1"
                     /note="predicted using Genefinder"
                     /codon_start=1
                     /db_xref="PID:e1321971"
                     /db_xref="PID:g3642158"
                     /translation="MGGGSGLLINSNRRKGSSSGPGCEGWPCDLMDGAMVMLVVMVIA
                     TALIGCLKACLCKKCKQLLREPKPTSEAAQEVPNGPELPQISEAPPPMLPVVIVLENP
                     PEYAENDQYMPPAYCTLRFNNDEEAQI"
     CDS             join(complement(Z84574:2170. .2317),
                     complement(Z84574:1634. .2131),
                     complement(Z84574:1388. .1537),
                     complement(Z84574:971. .1280),complement(Z84574:461. .842),
                     complement(6191. .6351),complement(5940. .6099),
                     complement(4455. .5110),complement(3863. .3952))
                     /gene="F33E2.2"
                     /note="similar to Eukaryotic protein kinase domain; cDNA
                     EST EMBL:Z14451 comes from this gene; cDNA EST EMBL:D34773
                     comes from this gene; cDNA EST EMBL:D36020 comes from this
                     gene; cDNA EST EMBL:D68216 comes from this gene; cDNA EST
                     EMBL:D68574 comes from this gene; cDNA EST yk219e4.5 comes
                     from this gene; cDNA EST yk371f1.5 comes from this gene;
                     cDNA EST yk413g9.5 comes from this gene; cDNA EST
                     yk421c10.5 comes from this gene"
                     /codon_start=1
                     /db_xref="PID:e1321972"
                     /db_xref="PID:g3642159"
                     /translation="MVAATGNASKPSLNSFYADGLGQLRNGLFSCFQPVPGYFGTKTT
```



```
                     VEIEKSEDELWEIPFDAISELEWLGSGSQGAVFRGQLENRTVAVKKVNQLKETEIKHL
                     RHLRHQNIIEFLGVCSKSPCYCIVMEYCSKGQLCTVLKSRNTITRELFAQWVKEIADG
                     MHYLHQNKVIHRDLKSPNILISAEDSIKICDFGTSHMQKKMDSTMMSPCGTVSWMAPE
                     MIKKQPCNEKVDVYSFGVVLWEMLTRETPYANIAQMAIIPGVGTNILSLPMPEEAPKG
                     LVLLIKQCLSQKGRNRPSFSHIRQHWEIPKPELFEMTEEEWQLAWDSTREFAKCIQTP
                     STVTRDHGGPKSAFAMZEEIQKREEQLNHIKDIRNMYDKLLKETNERMYDKLQGCFTE
                     LKLKESELAEWEKDLTEREQWHNQNSPKAVAAPRAQLRGYPHEGYDDMSDEDVQPCR
                     GSPYRCSNTSSSSGVQSSPFSRQSSSRSSSAGQQTRRSEGANPPKILRNDAIRHSGSYW
                     ETLGGARGSPARDSGFSQDSGMWSAGAGSCTAINGGGQQVCYSQTLYRNGDGRWSDGR
                     IASRRRVSTSVNKSTAVPGQPVPFTRDSPSRVPHARSSSKLNRSSYPSRNAPHQLEDG
                     CCCAHARAPRAKSIAVPMTSSSRARSPTPYDNDPENAESFVDPESPKNLKNLEKIVNL
                     PESTSYDEALCNSDVTMNPIYTSPITTYSNPCHVELVDEENANDVDLTSSMDSRRSRS
                     DDADVESSEEDEGNGNNILNTSMESEDLRYRIDTSQSTMMSSLERSLEIGATRSDGLS
                     DNEMRVQAVKMSIKTHRRTGSNPQALIHQCIDEYTTSATDDSDDAGAVRI"
BASE COUNT     2001 a    1205 c    1155 g    2032 t
ORIGIN

  Query Match          12.2%;  Score 18;  DB 18;  Length 6393;
  Best Local Similarity 100.0%;  Pred. No. 3.50e+00;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   3588 TTATATTAAAATAATAAT 3605
          ||||||||||||||||||
Qy    123 TTATATTAAAATAATAAT 140


RESULT   20
LOCUS       HSU159B9     40267 bp    DNA              PRI       06-FEB-1996
DEFINITION  Human DNA sequence from cosmid U159B9, between markers DXS366 and
            DXS87 on chromosome X.
ACCESSION   Z69367
NID         g1182004
KEYWORDS    X.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 40267)
  AUTHORS   Dodsworth,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-FEB-1995) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1RQ, UK. E-mail enquires: humquery@sanger.ac.uk
COMMENT     IMPORTANT:  This sequence is the entire insert of clone U159B9. The
            true left end of clone U159B9 is at 1 in this sequence. The true
            right end of clone U159B9 is at 40267.
            U159B9 is from the human chromosome X-specific cosmid library.
FEATURES             Location/Qualifiers
     source          1. .40267
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
```



```
                     /map="X"
                     /clone="U159B9"
     repeat_region   1436. .1720
                     /note="L1 element fragment"
     repeat_region   2333. .2385
                     /note="MIR2 element fragment"
     repeat_region   5078. .5191
                     /note="L1 element fragment"
     repeat_region   6131. .6253
                     /note="L1 element fragment"
     repeat_region   6691. .6981
                     /partial
                     /note="Alu repeat: matches 308. .1 of consensus"
     repeat_region   7634. .7804
                     /note="L1 element fragment"
     repeat_region   8909. .10043
                     /note="L1 element fragment"
     repeat_region   11305. .11590
                     /note="Alu repeat: matches 1. .308 of consensus"
     repeat_region   11711. .11857
                     /note="L1 element fragment"
     repeat_region   11937. .13696
                     /note="L1 element fragment"
     repeat_region   13755. .14291
                     /note="L1 element fragment"
     repeat_region   14360. .14431
                     /note="36 copies of 2 mer 96 % conserved"
     repeat_region   14432. .14769
                     /note="L1 element fragment"
     repeat_region   14811. .16706
                     /note="L1 element fragment"
     repeat_region   18185. .18559
                     /note="L1 element fragment"
     repeat_region   19169. .19494
                     /note="L1 element fragment"
     repeat_region   19899. .20231
                     /note="L1 element fragment"
     repeat_region   20326. .20891
                     /note="L1 element fragment"
     repeat_region   20913. .21325
                     /note="MSTA element fragment"
     repeat_region   21331. .21782
                     /note="L1 element fragment"
     repeat_region   21858. .21959
                     /note="L1 element fragment"
     repeat_region   22016. .22124
                     /note="L1 element fragment"
     repeat_region   24568. .24666
                     /note="L1 element fragment"
     repeat_region   26186. .26352
                     /note="L1 element fragment"
```

```
     repeat_region   26408. .26542
                     /note="L1 element fragment"
     repeat_region   26696. .26764
                     /note="L1 element fragment"
     repeat_region   26832. .27460
                     /note="L1 element fragment"
     repeat_region   27771. .29494
                     /note="L1 element fragment"
     repeat_region   29629. .30674
                     /note="L1 element fragment"
     repeat_region   31179. .31218
                     /note="20 copies of 2 mer 88 % conserved"
     repeat_region   32818. .32968
                     /note="L1 element fragment"
     repeat_region   34237. .34652
                     /note="MER29 element fragment"
     repeat_region   34861. .35294
                     /note="MER17 element fragment"
     repeat_region   35080. .35294
                     /note="MER7 element fragment"
     repeat_region   35499. .35600
                     /note="L1 element fragment"
     repeat_region   35663. .35953
                     /partial
                     /note="Alu repeat: matches 308. .1 of consensus"
     repeat_region   37509. .38257
                     /note="L1 element fragment"
     repeat_region   38299. .39528
                     /note="L1 element fragment"
     repeat_region   39624. .39865
                     /note="MER42C element fragment"
     repeat_region   39924. .40197
                     /note="L1 element fragment"
BASE COUNT    12780 a   6850 c   7065 g  13572 t
ORIGIN

  Query Match           12.2%;  Score 18;  DB 25;  Length 40267;
  Best Local Similarity 100.0%;  Pred. No. 3.50e+00;
  Matches    18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  17587 TATTTTAATATAATCTGT 17604
          ||||||||||||||||||
Cp    135 TATTTTAATATAATCTGT 118


RESULT   21
LOCUS       CELC28H8    46255 bp    DNA             INV       19-FEB-1995
DEFINITION  Caenorhabditis elegans cosmid C28H8.
ACCESSION   U20861
NID         g669006
KEYWORDS    .
```



```
SOURCE      Caenorhabditis elegans strain=Bristol N2.
  ORGANISM  Caenorhabditis elegans
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Nematoda;
            Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea;
            Rhabditidae; Caenorhabditis.
REFERENCE   1  (bases 1 to 46255)
  AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
  TITLE     The C. elegans genome project: Contiguous nucleotide sequence of
            over two megabases from chromosome III
  JOURNAL   Nature (1994) In press
REFERENCE   2  (bases 1 to 46255)
  AUTHORS   Miller,N.
  TITLE     The sequence of C. elegans cosmid C28H8
  JOURNAL   Unpublished (1994)
REFERENCE   3  (bases 1 to 46255)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-FEB-1995)
COMMENT               Submitted by:
                      Genome Sequencing Center
                      Department of Genetics, Washington University,
                      St. Louis, MO  63110, USA, and
                      Sanger Centre, Hinxton Hall
                      Cambridge CB10 IRQ, England
                      e-mail: rw@nematode.wustl.edu and jes@sanger.ac.uk


            NEIGHBORING COSMID INFORMATION:


            The 5' cosmid is F01F1 (500 bp overlap); gap on 3' end.  Actual
            start of this cosmid is bp 497 of this sequence; and ends at 46255
            of this sequence

            NOTES:

            Coding sequences below are predicted from computer analysis, using
            the program Genefinder (P. Green and L. Hillier, ms in
            preparation).
FEATURES              Location/Qualifiers
```



```
     source          1. .46255
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /map="III"
                     /strain="Bristol N2"
     gene            1609. .3077
                     /gene="C28H8.5"
     CDS             join(1609. .1878,1926. .2019,2175. .2358,2571. .2704,
                     2757. .2860,2955. .3077)
                     /gene="C28H8.5"
                     /codon_start=1
                     /evidence=not_experimental
                     /db_xref="PID:g669011"
                     /translation="MRLPRRLKIKRPHKPSCSSILPKYPGCPPSLMLSLSFLNPTHYD
                     VHLQMINRFFPNVYYSVVIPIILLSTSYSAKIAPKCRDTDMNCAVWVATNTSDCENVE
                     LVNSHCPRTCQTCGEPIDPKYDVKLLPAKLKSIAWMVGRWRSEPGGKAFFPTIPKFTY
                     GEQVDITIADNSQDTHTPLLNYTAPAWDINMPDGDPTEIHSENGYIAVEYDKEQEKEY
                     VSLNTAMSNGFMTIEEGESGPNQVKFRLQRIGRISFSHDSAVRIMFREWTLLDENRLE
                     ARLLMTTTITRRLMEHTAVIYKKIYP"
     gene            complement(3149. .4258)
                     /gene="C28H8.7"
     CDS             complement(join(3149. .3418,3494. .3724,3777. .3917,
                     4049. .4258))
                     /gene="C28H8.7"
                     /codon_start=1
                     /evidence=not_experimental
                     /db_xref="PID:g669013"
                     /translation="MSPHPRNIKLLFALILLLLVYIARQKRFEQQYFDRYLEHPRGGD
                     STKSKFQTEVDKWVKCMEKDLYVDRESWYYVRFATERCKNASIFLNVPIMKLETNVEY
                     QTGIRSYDMPIATKKEPPVPVTIGNYTDSRTRKNMTDELPTDLKTYGTHLNSSNIYWD
                     YGSLANYLNQSESAGIGHFSTYLKTQVNHKVIDLISIKEEVNNYKIPHEISRMGKLDE
                     NGQIVCQIEIRLAPPTSQTIKHLMAGLSLIFNDSRYVLLSFHNLELFLVNYEHQECRG
                     KYLAPYI"
     gene            complement(5150. .6077)
                     /gene="C28H8.8"
     CDS             complement(join(5150. .5362,6015. .6077))
                     /gene="C28H8.8"
                     /codon_start=1
                     /evidence=not_experimental
                     /db_xref="PID:g669014"
                     /translation="MFIILFLCFTVLFMLFYPDYKEPSVVVTLGIGLDVKAETTLKET
                     LPTNSRFYAADPIYTGNGELYEPVGSYFPPAVGKETDVSTALVLKNR"
     gene            complement(7701. .9666)
                     /gene="C28H8.6"
     CDS             complement(join(7701. .7979,8626. .8760,8809. .8904,
                     9147. .9320,9580. .9666))
                     /gene="C28H8.6"
                     /note="contains LIM domains; weakly similar to zyxin
                     (chicken) and C. elegans lin-11"
                     /codon_start=1
                     /evidence=not_experimental
```



```
                     /db_xref="PID:g669012"
                     /translation="MNGELSSKHGVNTIPKGDCAACGKPIIGQVVIALGKWWHPEHYT
                     CCECGAELGQRPFFERNGRAFCEEDYHNQFSPKCQGCHRAITDRCVSVMNKNFHIECF
                     TCAECNQPFGEDGFHEKNGQTYCKRDFFRLFAPKCNGCSQPITSNFITALGTHWHPDC
                     FVCQHCGVSFNGASFFEHNGAPLCERHYHESRGSICSQCRGAINGRCVAAMGRKFHPE
                     HFRCSYCNHQLTKGTFKEVDRRPFCHKCYNNTYALTPA"
     gene            complement(21184. .22352)
                     /gene="C28H8.9"
     CDS             complement(join(21184. .21333,21385. .21617,21665. .21747,
                     22210. .22352))
                     /gene="C28H8.9"
                     /note="weakly similar to Neuro-D4 protein (R. norvegicus)"
                     /codon_start=1
                     /evidence=not_experimental
                     /db_xref="PID:g669015"
                     /translation="MLGNCTRLFNFLHALVFPLIMHVSPEIQGKLNQKNGKFGLKNKF
                     WKIIAKYKSLAGLSYHQSYLHDQKSSQPLVKLLSPSIEISTSCDFCSGTAFMNKNTKL
                     PEDLVSCHDCGRSGHPSCLNFNQNVTKIIKRSGWQCLECKSCTICGTSENDDKLLPCD
                     DCDRGYHLYCLTPALEKAPDDEYSCRLCQVEPGDKASAPAKK"
     gene            complement(22407. .23417)
                     /gene="C28H8.10"
     CDS             complement(join(22407. .22649,22699. .22904,22955. .23208,
                     23335. .23417))
                     /gene="C28H8.10"
                     /codon_start=1
                     /evidence=not_experimental
                     /db_xref="PID:g669016"
                     /translation="MISENGYMDLMRNCIKWNSRQMEDRKKRLRFPYYEHQTATAQRE
                     SRFTTRNVDHMYPSNDPNTVVQFATERWKKSKSQPPSDAVEMSMFLRENPSIQVAIDH
                     LTPQVVGPTTESVSDSSNDSTTIRPSRQTQIKEEYRDDYVLDDELSPDEPGSDEDDWS
                     SRKRRKGNLGPVQKATSSRKKVPTTRSSVSRLTPSRSIVKETKYEEPEEKTYPCDSEW
                     LYGGLAKPSRPGEHTILLVLFSLICLPHRPQHVSYREHYHHDF"
     gene            24360. .25492
                     /gene="C28H8.4"
     CDS             join(24360. .24551,24596. .24838,25286. .25492)
                     /gene="C28H8.4"
                     /note="similar to endoplasmic reticulum lumen protein
                     retaining receptor"
                     /codon_start=1
                     /evidence=not_experimental
                     /db_xref="PID:g669010"
                     /translation="MNIFRISADMSHLLAIIILLLKIWKSRSCSGISARSQILFALVF
                     TARYLDLPSTYISLYNTTMKITPLAATYATVYLMFFKFRSTYMRESDTFRVELLIVPA
                     AILALLINHDFAPFELLWTFSIYLEAVAILPQLFLLQSTGSAEVITAHYLFALGSYRA
                     LYIFNWIYRYYTEDYFDPIVVVAGIVQTVLYADFFYLIYTRVVQTRKGMELPI"
     gene            26085. .31786
                     /gene="C28H8.3"
     CDS             join(26085. .26850,26897. .27339,27388. .28734,28779. .29002,
                     29052. .29453,29501. .30439,30524. .31166,31247. .31513,
                     31649. .31786)
                     /gene="C28H8.3"
```

/note="similar to yeast antiviral protein SKI2 and ATP-dependent DNA-helicases; coded for by C. elegans cDNA yk28d12.3; coded for by C. elegans cDNA yk31b5.5; coded for by C. elegans cDNA yk7a1.3; coded for by C. elegans cDNA yk13a9.3; coded for by C. elegans cDNA yk31b4.3; coded for by C. elegans cDNA yk28d12.5; coded for by C. elegans cDNA yk13a9.5; coded for by C. elegans cDNA yk31b4.5; coded for by C. elegans cDNA yk44g7.5; coded for by C. elegans cDNA yk7d6.5"
/codon_start=1
/db_xref="PID:g669009"
/translation="MIAVCFLTFQASTFKNLTDSENENVVVEEKPVILPRARQPSIAK SLEKVQILENSVAGSEKNASDDDSDASSVMSDDLESLGETTEVLNFSLEQLSMMKKNL AQFPCSYKNIISEFADVEIFLISIDSLIVECAAHSYHNWDVAGQSMVLNKQIDRFLQQ FVDIGGRFKLVVFSDLTTQPAKDTTLSFARSTALAHLAWGPHARDLLFFTNPTDPEWD KLLNDLTPSPLMISTDNVTQNVCASQEIDLTKQFETIAFDVLTRSMSVVLLHSIKVNF VSVEAYYIQPLLVVAPDWQAFLAAHWDNNGTLLRNHLSKEIKFNQFETPADLWVKVIT DAKATGKGSDTFDTLAAAVILSSLICSRRGAHRIYYPTRTDAKRGLNVIRDRRLILNA AVVLLDKADHANSKLDLSDLWDGRMVYSIFDELSANETVLPYRLQDDFAKYHQKAGLT VPLATDTNEKLFDPIPEITDPLVDLPILYSVTSPMIKRFVPEIEKMTSQNAVEEGTVQ DYADYLKDTSQWRLKPIEETTYAQKEEKIEDAWQLKKANRSKQFLMRWYELFANSLEGR GSNLLVDFSRVPKGIAVVEEEVTDEKKTGKGGWSGQKQQKPGAGGKKGATKESGKSKK DLILEANKKAKDQKVAESEKVKIKYGCQQGKESVIFLNNLYSSLDLPETRALCVYEIA VREGRTIFDQHQGTDKQEVRRNAAIDLVGHLKDCFVKHWDHLESKQKEQIVDLWVSLG FFAPAGSKPSSEAKQKKLNLGINWVTYQLQYGGELIDIQSDPKKDDRVSGFAPDQWQR KWLDSVDRGNSALIIAPTSAGKTFVSYICIEKVLRSSDNDVVVYVAPSKALINQVCGS VYARPRNKSMKRGISLPGTLTQETSQNAMQCQVLITVPECLQELMLSRTPAVQKFVSH IKYVVFDEVHSIGASEESHIWEQLLLLIQCPPLALSATIGNANKLHEWLNSSEQAKSA GKRKVELINYGERYSELELSILNINDPHGEDDGAVHKKAGERAVIPLMPYGVYMPEKL RMFSIPEDQQLTARQVLHLYNMMAEVDDATKKEFEPCKPPGQHGTRAVWISRSELRRL ENALKERFMEWLSSDEQKINSILKILKEPVNTQLSYRARPPNKEKIANDYIVTLVDDL KEKGELPAICFNDDRHVCEKLAVTLANELERRELEYMETDEFKNKYMIKDESKLVKLA KRKRDDAEKKKKGDKDEDAGPEKDDDEMDVLAMKKAKLARALERFKLRGRNGGDPDIY AKMVERMQKGAKTRESTQVLLKLFERGIGTHHAGLNTVERGAVEVLFRSGNLAVLFST STLSLGVNMPCKTVMFGVDTLQLTPLLYRQMSGRAGRGFDHSGNVIFMSIPTSKVRR LLTASLSNLQKNPPFTVPLRLFSVLRLFDNQDILNEEGQKVSTMKQRAFAAKSLLESHSF SIHTRREAQDGILQKQLRMFSAFSPQLLRHLQLLSPPGEGKNFAEMAIHSSSGANGTL LFIYLMQKKCFHQLIKSYDTAEQAQLGILEVLANLFTNLRMTPFHERSDNLENVQVTL RGLPSLLKPYVEEYNSTVTGLYKRFMAASSQDGNLFDPSFAVSGKLDSESVSLTEDFL VAPLFDQYSHDESFLPVIDFNKKDHRGRKIQRNAFAYDFTVHGSRNMLMDVNGLHVSV AWFLLHDFAAILERLAIGVHNMARPQDPLVLVLEELHRNYDEKFRKAPGMRTRD"
gene complement(35012. .37253)

...

Note: remainder of annotations omitted.

Query Match 12.2%; Score 18; DB 18; Length 46255;
Best Local Similarity 100.0%; Pred. No. 3.50e+00;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 6372 ATTATTATTTTAATATAA 6389
||||||||||||||||||

Cp 140 ATTATTATTTTAATATAA 123

RESULT 22
LOCUS CEY38G4 117864 bp DNA HTG 03-OCT-1997
DEFINITION Caenorhabditis elegans DNA *** SEQUENCING IN PROGRESS *** from clone Y38G4; HTGS phase 1.
ACCESSION Z99713
NID g2467102
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryotae; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 117864)
AUTHORS Sulston,J.
TITLE Direct Submission
JOURNAL Submitted (03-OCT-1997) Nematode Sequencing Project, Sanger Centre, Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics, Washington University, St. Louis, MO 63110, USA. E-mail: jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT On Oct 7, 1997 this sequence version replaced gi:2464982.
IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and
the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated
with foreign sequence from E.coli, yeast, vector, phage etc. Order of segments is not known; 800 n's separate segments.
*** ***
*** WARNING: Phase 1 High Throughput Genome Sequence ***
*** ***
* This sequence is unfinished. When sequencing is complete,
* the sequence data presented in this record will be replaced
*by a single finished sequence with the same accession number.
FEATURES Location/Qualifiers
source 1. .117864
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/clone="Y38G4"
/chromosome="II"
BASE COUNT 26604 a 15058 c 14789 g 26050 t 35363 others
ORIGIN

Query Match 12.2%; Score 18; DB 17; Length 117864;
Best Local Similarity 100.0%; Pred. No. 3.50e+00;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 37064 ACAGATTATATTAAAATA 37081
||||||||||||||||||

Qy 118 ACAGATTATATTAAAATA 135

RESULT 23
LOCUS CEY17G7B 143092 bp DNA INV 22-SEP-1998
DEFINITION Caenorhabditis elegans cosmid Y17G7B, complete sequence.
ACCESSION AL023828
NID g3217816
KEYWORDS HTG.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 143092)
AUTHORS Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M., Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J., Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A., Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M., Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N., Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B., O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A., Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E., Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M., Vaughan,K., Waterston,R., Watson,A., Weinstock,L., Wilkinson-Sproat,J. and Wohldman,P.
TITLE 2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans
JOURNAL Nature 368 (6466), 32-38 (1994)
MEDLINE 94150718
REFERENCE 2 (bases 1 to 143092)
AUTHORS Smye,R.
TITLE Direct Submission
JOURNAL Submitted (06-NOV-1996) Nematode Sequencing Project, Sanger Centre, Hinxton, Cambridge CB10 1RQ, England and Department of Genetics, Washington University, St. Louis, MO 63110, USA. E-mail: jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT Coding sequences below are predicted from computer analysis, using predictions from Genefinder (P. Green, U. Washington), and other available information.
For a graphical representation of this sequence and its analysis see:-
http://webace.sanger.ac.uk/cgi-bin/display?db=wormace&class=Sequence &object=Y17G7B
Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small

overlap between neighbouring submissions.
IMPORTANT: This sequence is not the entire insert of clone Y17G7B. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.
The true right end of clone W03C9 is at 100 in this sequence. The true right end of clone Y17G7 is at 143092 in this sequence. The start of this sequence (1. .100) overlaps with the end of sequence Z66516.
FEATURES Location/Qualifiers
source 1. .143092
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="II"
/clone="Y17G7B"
gene complement(3569. .4874)
/gene="Y17G7B.1"
CDS complement(join(3569. .3750,3804. .3945,4851. .4874))
/gene="Y17G7B.1"
/note="similar to Acyl CoA binding protein"
/codon_start=1
/db_xref="PID:e1323259"
/db_xref="PID:g3646792"
/translation="MPISELEKKWTGLTFEIAAEEMRRLKSEPTDRERLKLYALYKQA LHGDIPNEDVTPVPAGDEVGRKKYAAWKSQKGANSEKCRADYVAIAEEMIKKYGRNIV RCKWNSEVWSVDY"
gene 11683. .14666
/gene="Y17G7B.2c"
gene 11683. .19658
/gene="Y17G7B.2b"
gene 11683. .19658
/gene="Y17G7B.2a"
CDS join(11683. .11753,11836. .11928,11980. .12304,13242. .13462, 14218. .14331,15865. .16127,17187. .17299,18228. .18428, 18624. .18714,19438. .19658)
/gene="Y17G7B.2b"
/note="cDNA EST EMBL:D74713 comes from this gene; cDNA EST yk470e5.5 comes from this gene"
/codon_start=1
/db_xref="PID:e1323260"
/db_xref="PID:g3646793"
/translation="MRSSKGGRGRQAAPKTAPTTVCYCDGKRELGSVEVVCSTCLKWF HGRCLKEPHELNSNGVPFMICYTPTCKQCRPTAEDWKAKKADLVQMCVTVLATLGAER LKADCKLGAEHVPEDPTYLSLKDEIVPYNNENWYMLYAIKQKKEWHQNLAPTLLKEKN IPVQHNDDDDLFALAEKNLSLLGPLHEAVKLIGKRPIERENREPRHIELPPIEGPKTR GASKRRHAEAPVTGKKQKLAADYSSTAAPNGVQIDIPPSKDNYRYYLTEVDPNVPEDP AWNQNQSSATVIPSPHYRELLNPTVNVSSNDRAPQLSINGNSITGPEGYSMARASHGV SKGTWYFEVNFDDQPDDSHIRIGWSQSTASLQACVGYNKFSYGWRSKHGTKFHEAAGR KYHPGGFKQGDVLGCLIHLPVDKKLQIPANLPSEKYLPVSHKGFNLISFKANYPPEVQ EESADIAKTLVEMPGSYIEFPHNGKSCGKAYENIYAGAYYPSISIFKSATATHNLGPK FRNLPRGATGIHARADEQQHEQTLSDMLYLVSKEVNLDHPPRVKREDDDDVKDIKKEI

```
                KQEI"
     CDS        join(11683. .11753,11836. .11928,11980. .12304,13242. .13462,
                14218. .14331,14622. .14666)
                /gene="Y17G7B.2c"
                /note="cDNA EST yk470e5.3 comes from this gene"
                /codon_start=1
                /db_xref="PID:e1323262"
                /db_xref="PID:g3646795"
                /translation="MRSSKGGRGRGRQAAPKTAPTTVCYCDGKRELGSVEVVCSTCLKWF
                HGRCLKEPHELNSNGVPFMICYTFTCKQCRPTAEDWKAKKADLVQMCVTVLATLSAER
                LKADGKLSAEHVPEDPTYLSLKDEIVPYMNENWYMLTAIKQKKEWHQNLAPTLLKEKN
                IFVQHNDDDDLFALAEKNLSLLGPLHEAVKLIGKRPIERENREPRHIELPPIEGPKTR
                GASKRRHAEAPVTGKKQKLAADYSSTAAPNGVQIDIPFSKDNYRYYLTEVDPNVPEEP
                IFPIFLAPSLFYI"
     CDS        join(11683. .11753,11836. .11928,11974. .12304,13242. .13462,
                14218. .14331,15865. .16127,17187. .17299,18228. .18420,
                18624. .18714,19438. .19658)
                /gene="Y17G7B.2b"
                /note="cDNA EST EMBL:D35727 comes from this gene"
                /codon_start=1
                /db_xref="PID:e1323261"
                /db_xref="PID:g3646794"
                /translation="MRSSKGGRGRQAAPKTAPTTVCYCDGKRELGSVEVVCSTCLKWF
                HGRCLKEPHEFSRNSNGVPFMICYTFTCKQCRPTAEDWKAKKADLVQMCVTVLATLSA
                ERLKADGKLSAEHVPEDPTYLSLKDEIVPYMNENWYMLTAIKQKKEWHQNLAPTLLKE
                KNIFVQHNDDDDLFALAEKNLSLLGPLHEAVKLIGKRPIERENREPRHIELPPIEGPK
                TRGASKRRHAEAPVTGKKQKLAADYSSTAAPNGVQIDIPFSKDNYRYYLTEVDPNVPE
                DPAWNQNQSSATVIPSFHYRELLNPTVMVSSNDRAPQLSINGNSITGFEGYSMARASH
                GVSKGTWYFEVNFDDQPDDSHIRIGWSQSYASLQACVGYNKFSYGWRSKHGTKFHEAK
                GKKYHFGGFKQGDVLGCLIHLPVDKKLQIPANLPSEKYLPVSHKGFNLISFKANYPFE
                VQEESADIAKTLVEMPGSYIEFPHNGKSCGKAYENIYAGAYYPSISIFKSATATMNLG
                PKFRNLPRGATGIHARADEQQHEQTLSDMLYLVSKEVNLDHPPRVKREDDDDVKDIKK
                EIKQEI"
     gene       20244. .23044
                /gene="Y17G7B.3"
     CDS        join(20244. .20381,21383. .21553,22571. .23044)
                /gene="Y17G7B.3"
                /note="similar to Metallo-beta-lactamase superfamily"
                /codon_start=1
                /db_xref="PID:e1323263"
                /db_xref="PID:g3646796"
                /translation="MVYVKSLLRRADNFMYIVKKSSEARAALLVDLVNEEDYKELADK
                ENIDITAVLTTHHHYDHCGGNEGFRRQFPNVMILGGDSRIPAMDRHVKHGDMAEFAGL
                QIKCLSTPCHTSGHICYHITNPAADSTSPGVVFTGDTLFIAGCGRFFEGTAPQMDVAL
                NEILKNLPVETQIFPGHEYTVANLKFACHVEPGNEKAAQKLEWAQRQIEQGGPTVPST
                VAEEKATNPFMRVRESEEIQKSIGTCDAVVGMAKLREMKNKF"
     gene       24223. .29110
                /gene="Y17G7B.4"
     CDS        join(24223. .24435,25013. .25162,26304. .26685,28307. .28435,
                28869. .29110)
                /gene="Y17G7B.4"
```



```
                /note="cDNA EST EMBL:D73584 comes from this gene"
                /codon_start=1
                /db_xref="PID:e1323264"
                /db_xref="PID:g3646797"
                /translation="MSTNEAAASQEDIALAQGAVLVKSCQVPDGSIPIRGFDFSTASG
                PDFSLSAILSSYMSTGFQATHLAQAIQQVNQMLSLRDTPLTCDDDEKLPPYPEGRQKR
                SCTIPLGYTSNLVTSGLREVLRYCVQRNLVDCIVTSAGGIEEDLIKCLKPSYLGTFTM
                DGAKLRSNGMNRAGNVLIPNDNYCAFEDWLMPILDECLVEQEEKHLNWTPSKLIQRLG
                ERIGDESSILYWAAKHRIPVFCPALTDGSLGDMLYFHSVKSSPGLRVDIVEDVRHINT
                IAVKSFKTGSIILGGGVVKHHINNANLMRNGADHTVYINTGQEFDGSDSGAQPDEAVS
                WGKVKPSAGAVKVHAEATLVFPLLVAETFAKHEGHKD"
     gene       complement(32881. .37816)
                /gene="Y17G7B.5"
     CDS        complement(join(32881. .33253,34402. .35645,36629. .37493,
                37653. .37816))
                /gene="Y17G7B.5"
                /note="similar to MCM2/3/5 family; cDNA EST yk244e4.5
                comes from this gene"
                /codon_start=1
                /db_xref="PID:e1323265"
                /db_xref="PID:g3646798"
                /translation="MADEANNDDDVDRQPLPIADDADDDVDGIDEMFNNEDEDDPEDEE
                GENLPGDDMERDYREQPELDQYSESGMDDASDVGSLSVSARRAAERENAQRDQLLDDD
                ALMYEDGDSEEVDTRRRGRGRRGRGDAADDDSVPMEEEDIPVDILENIRGRTIRDHVS
                DEAVAKEIERRFKNFLRSFHEPGNKQTKYIQMIKSMAADNRESLEVSFTDLSDDNGEQ
                NISYFLPEAPNEMLAIMDRAATEVVMNMYPFYSRVCNEIKVRISQLPVEEDIRMLRQV
                HLNMLIRTAGVVTIASGILPQLAVVKYDCVACGYLLGPFVQQNDEEVRPTICPSCQGK
                GPFELNVENTVYHNYQRITMQESPNKVAAGRLPRSKDVILLGDLCDSCKPGDEIEVTG
                VTTNNFDGSLNYKQGFPVFNTLIHANHITNKDKMASDQLTDEDIKAIRELSQDPNISQ
                RVFSSIAPSITGHDDVKRAIALALFRGEARNPGAKHRLRGDINVLLCGDPGTAKSQFL
                RYAAHIAPRSVLTTGQGASAVGLTAYVQRHPVTREWTLEAGAMVLADKGVCLIDEFDK
                MSDQDRTSIHEAMEQQSISISKAGIVTSLHARCTVIAASNPIGGRYNPTRTFAENVDL
                TEPILSRFDVLCVIRDSVDSVEDERLAKFVVGNHRTHHPDAKKIVKEGDELEEDQMDE
                RTGVRLIPQDLLRKYIIYAREKCHPTLPEQHSEKFSNIFAQMRKESMATGSVAITVRH
                VESMIRLSEAHAKLHLRSYVNDEDCAAAIRVMLESFVNTQKASIMRMMKTFSRHLTE
                NRSANELLLFILKQLIRQQMHYATARAAAGTILQSVTIPESEFIEKAQQLRIENVKPF
                YTSEIFASNNFLYDPSKKTIVQEIP"
     gene       52521. .54875
                /gene="Y17G7B.6"
     CDS        join(52521. .52555,53285. .53390,53728. .53819,53916. .54102,
                54595. .54744,54795. .54875)
                /gene="Y17G7B.6"
                /note="similar to Globin"
                /codon_start=1
                /db_xref="PID:e1323272"
                /db_xref="PID:g3646805"
                /translation="MLTLLGKIAHFREPLSRQLSLSDTNVTPQISRCAKLNLSVKQKK
                LLRQSFNAMNSGGTFLKLMEKIFRRLETKCPDMRSIFLTTAFVNSLSRERQTPPLVKT
                EYDHCKCMVGIFERLIENLENINEQLTMIRHYGEKHAQMAESGFTGAMIEQPGEISVF
                VIGSQDVVKFNHETVKAWRLLLLACVTDEMKVGFDRMSRINGRRNSCNPPSTVPSGN"
     gene       55074. .57050
```



```
...
Note: remainder of annotations omitted.

  Query Match           12.2%;  Score 18;  DB 18;  Length 143092;
  Best Local Similarity 100.0%;  Pred. No. 3.50e+00;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  44179 TATTTTAATATAATCTGT 44196
          ||||||||||||||||||
Cp    135 TATTTTAATATAATCTGT 118


RESULT   24
LOCUS       AC005073   176011 bp    DNA             HTG       12-JUN-1998
DEFINITION  *** SEQUENCING IN PROGRESS *** Homo sapiens clone RG189J21; HTGS
            phase 1, 15 unordered pieces.
ACCESSION   AC005073
NID         g3212918
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 176011)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 176011)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-1998) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     ***                                                        ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                        ***
            * This sequence is unfinished. It consists of 15 contigs for
            * which the order is not known; their order in this record is
            * arbitrary. In some cases, the exact lengths of the gaps
            * between the contigs are also unknown; these gaps are  presented
            * as runs of N as a convenience only. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
            *        1     2258: contig of 2258 bp in length
            *     2259     2275: gap of unknown length
            *     2276     4628: contig of 2353 bp in length
            *     4629     4645: gap of unknown length
            *     4646     8205: contig of 3560 bp in length
            *     8206     8222: gap of unknown length
            *     8223    13646: contig of 5424 bp in length
            *    13647    13663: gap of unknown length
```



```
            *    13664    18633: contig of 4970 bp in length
            *    18634    18650: gap of unknown length
            *    18651    24946: contig of 6296 bp in length
            *    24947    24963: gap of unknown length
            *    24964    31163: contig of 6200 bp in length
            *    31164    31180: gap of unknown length
            *    31181    40495: contig of 9315 bp in length
            *    40496    40512: gap of unknown length
            *    40513    47925: contig of 7413 bp in length
            *    47926    47942: gap of unknown length
            *    47943    58647: contig of 10705 bp in length
            *    58648    58664: gap of unknown length
            *    58665    68826: contig of 10162 bp in length
            *    68827    68843: gap of unknown length
            *    68844    84172: contig of 15329 bp in length
            *    84173    84189: gap of unknown length
            *    84190    96721: contig of 12532 bp in length
            *    96722    96738: gap of unknown length
            *    96739   136674: contig of 39936 bp in length
            *   136675   136691: gap of unknown length
            *   136692   176011: contig of 39320 bp in length.
FEATURES             Location/Qualifiers
     source          1. .176011
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RG189J21"
BASE COUNT    44764 a  42635 c  44477 g  43852 t    283 others
ORIGIN

  Query Match           12.2%;  Score 18;  DB 17;  Length 176011;
  Best Local Similarity 100.0%;  Pred. No. 3.50e+00;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db 103947 CCAGCCCTGGCCCCCCCT 103964
          ||||||||||||||||||
Qy     80 CCAGCCCTGGCCCCCCCT 97


RESULT   25
LOCUS       AC002480   200167 bp    DNA             PRI       21-AUG-1997
DEFINITION  Human BAC clone GS055R18 from 7p15-p21, complete sequence.
ACCESSION   AC002480
NID         g2340098
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 200167)
  AUTHORS   Maggi,L.
```

```
  TITLE     The sequence of H. sapiens BAC clone GS055K18
  JOURNAL   Unpublished (1997)
REFERENCE   2  (bases 1 to 200167)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-AUG-1997) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     SUBMITTED BY:
            Genome Sequencing Center
            Department of Genetics
            Washington University
            St. Louis MO 63108, USA
            http://genome.wustl.edu/gsc
            mailto:sapiens@watson.wustl.edu

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded or sequenced with an alternate
            chemistry; an attempt was made to resolve all sequencing problems,
            such as compressions and repeats; all regions were covered by
            sequence from more than one subclone; and the assembly was
            confirmed by restriction digest.

            MAPPING INFORMATION:
            The sequence of this clone was established as part of a mapping and
            sequencing collaboration between the NHGRI Chromosome 7 Mapping
            Project (Eric D. Green, Director), John D. McPherson in the
            Department of Genetics (Washington University), and the Washington
            University Genome Sequencing Center.  For additional information
            about the map position of this sequence, see
            http://www.nhgri.nih.gov/DIR/GTB/CHR7 or send
            mailto:egreen@nhgri.nih.gov

            SOURCE INFORMATION:
            This clone is from the first BAC library from Genome Systems, Inc.
            (http://www.genomesystems.com).
            Cell line:  lymphoblastoid
            Haplotypes:  two
            VECTOR:  pBeloBAC
            Selection:  chloramphenicol

            NEIGHBORING SEQUENCE INFORMATION:
            The actual start of this clone is at base position 1 of GS055K18;
            actual end is at 200167 of GS055K18.  The orientation of this clone
            is unknown.

            This clone contains 2 runs of A's, with varying numbers of A's,
```



```
            represented on both strands.  The regions are bases 99673-99691 and
            193551-193565.  This sequence represents the largest number in each
            run.

            This clone contains STS sWSS182 (NID:g292551).
FEATURES              Location/Qualifiers
     source           1. .200167
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="7"
                      /clone="GS055K18"
                      /clone_lib="GSBAC1"
                      /map="7p15-p21"
     repeat_region    344. .643
                      /rpt_family="MER"
     repeat_region    complement(892. .991)
                      /rpt_family="MER"
     repeat_region    complement(993. .1032)
                      /rpt_family="L1"
     repeat_region    complement(1044. .1307)
                      /rpt_family="MER"
     repeat_region    complement(2163. .2255)
                      /rpt_family="L1"
     repeat_region    2352. .2702
                      /rpt_family="MER"
     repeat_region    complement(2748. .2851)
                      /rpt_family="L1"
     repeat_region    3541. .3567
                      /rpt_family="L1"
     repeat_region    complement(5165. .5454)
                      /rpt_family="ALU"
     repeat_region    complement(9297. .9320)
                      /rpt_family="L1"
     repeat_region    12340. .12379
                      /rpt_family="L1"
     repeat_region    complement(13841. .14131)
                      /rpt_family="ALU"
     repeat_region    complement(14276. .14595)
                      /rpt_family="L1"
     repeat_region    complement(14672. .14773)
                      /rpt_family="L1"
     misc_feature     14852. .15251
                      /note="similar to EST N66612 (NID:g1218737) yy70c01.s1"
     misc_feature     complement(14852. .15203)
                      /note="similar to EST W06912 (NID:g1280973) za91g10.r1"
     misc_feature     14852. .15500
                      /note="similar to EST AA173083 (NID:g1754353) zp21g09.s1"
     misc_feature     14852. .15490
                      /note="similar to EST AA173083 (NID:g1754353) zp21g09.s1"
     misc_feature     complement(14853. .14955)
                      /note="similar to EST N45529 (NID:g1186695) yy75a07.r1"
```



```
     misc_feature     complement(14869. .15015)
                      /note="similar to EST H88631 (NID:g1070891) yw22g11.r1"
     misc_feature     complement(14876. .15340)
                      /note="similar to EST N45529 (NID:g1186695) yy75a07.r1"
     misc_feature     complement(14876. .15371)
                      /note="similar to EST W06912 (NID:g1280973) za91g10.r1"
     misc_feature     complement(14878. .15465)
                      /note="similar to EST N45559 (NID:g1186725) yy75g03.r1"
     misc_feature     complement(14880. .15111)
                      /note="similar to EST N45559 (NID:g1186725) yy75g03.r1"
     misc_feature     complement(14882. .15543)
                      /note="similar to EST AA075482 (NID:g1615547) zm88g09.r1"
     misc_feature     complement(14894. .15481)
                      /note="similar to EST AA166976 (NID:g1745369) zo87f08.r1"
     misc_feature     complement(14903. .15481)
                      /note="similar to EST AA166976 (NID:g1745369) zo87f08.r1"
     misc_feature     complement(14904. .15543)
                      /note="similar to EST AA075482 (NID:g1615547) zm88g09.r1"
     misc_feature     14922. .15650
                      /note="similar to EST AA206359 (NID:g1801809) zq58e08.s1"
     misc_feature     14936. .15680
                      /note="similar to EST AA206359 (NID:g1801809) zq58e08.s1"
     misc_feature     complement(14953. .15133)
                      /note="similar to EST W31015 (NID:g1312005) zb85e06.r1"
     misc_feature     complement(14971. .15369)
                      /note="similar to EST H88631 (NID:g1070891) yw22g11.r1"
     misc_feature     complement(14983. .15221)
                      /note="similar to EST AA166953 (NID:g1745329) zo87h04.r1"
     misc_feature     complement(15022. .15579)
                      /note="similar to EST AA166953 (NID:g1745329) zo87h04.r1"
     misc_feature     complement(15025. .15525)
                      /note="similar to EST W31015 (NID:g1312005) zb85e06.r1"
     misc_feature     complement(15046. .15579)
                      /note="similar to EST AA166953 (NID:g1745329) zo87h04.r1"
     misc_feature     15143. .16452
                      /note="probable elongation factor 1-alpha pseudogene;
                      similar to X16869 (PID:31092)"
     misc_feature     complement(15170. .15701)
                      /note="similar to EST AA504677 (NID:g2240837) ab02c09.r1"
     misc_feature     complement(15171. .15701)
                      /note="similar to EST AA504677 (NID:g2240837) ab02c09.r1"
     misc_feature     complement(15188. .15701)
                      /note="similar to EST AA504677 (NID:g2240837) ab02c09.r1"
     misc_feature     15206. .15418
                      /note="similar to EST H66612 (NID:g1218737) yy70c01.s1"
     misc_feature     complement(15368. .15886)
                      /note="similar to EST W56442 (NID:g1358359) zc59e06.r1"
     misc_feature     complement(15447. .15613)
                      /note="similar to EST W07309 (NID:g1281507) za97d09.r1"
     misc_feature     complement(15493. .16019)
                      /note="similar to EST W07309 (NID:g1281507) za97d09.r1"
```



```
     misc_feature     complement(15506. .16019)
                      /note="similar to EST H39112 (NID:g908611) yp66e08.s1"
     misc_feature     complement(15512. .16023)
                      /note="similar to EST H39112 (NID:g908611) yp66e08.s1"
     misc_feature     complement(15530. .16023)
                      /note="similar to EST H39112 (NID:g908611) yp66e08.s1"
     misc_feature     complement(15564. .16019)
                      /note="similar to EST W07309 (NID:g1281507) za97d09.r1"
     misc_feature     complement(15763. .16246)
                      /note="similar to EST C17784 (NID:g1579387)"
     misc_feature     complement(15766. .16246)
                      /note="similar to EST C17784 (NID:g1579387)"
     misc_feature     complement(15838. .16373)
                      /note="similar to EST AA314028 (NID:g1966357)"
     misc_feature     complement(15895. .16179)
                      /note="similar to EST AA329651 (NID:g1981893)"
     misc_feature     complement(15984. .16155)
                      /note="similar to EST T66039 (NID:g675084) yc24d10.r1"
     misc_feature     complement(15986. .16298)
                      /note="similar to EST N58402 (NID:g1202292) yv70e08.s1"
     misc_feature     complement(16005. .16481)
                      /note="similar to EST T66039 (NID:g675084) yc24d10.r1"
     misc_feature     complement(16061. .16528)
                      /note="similar to EST N58402 (NID:g1202292) yv70e08.s1"
     misc_feature     complement(16065. .16503)
                      /note="similar to EST T66039 (NID:g675084) yc24d10.r1"
     misc_feature     complement(16086. .16528)
                      /note="similar to EST N58402 (NID:g1202292) yv70e08.s1"
     misc_feature     complement(16110. .16503)
                      /note="similar to EST T66039 (NID:g675084) yc24d10.r1"
     misc_feature     complement(16210. .16468)
                      /note="similar to EST AA329651 (NID:g1981893)"
     repeat_region    complement(16572. .16992)
                      /rpt_family="L1"
     repeat_region    complement(17098. .17240)
...
Note: remainder of annotations omitted.

  Query Match           12.2%;  Score 18;  DB 26;  Length 200167;
  Best Local Similarity 100.0%;  Pred. No. 3.50e+00;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db 172516 CTCAGGCCAGCCCTGGCC 172533
          ||||||||||||||||||
Qy     74 CTCAGGCCAGCCCTGGCC 91


RESULT   26
LOCUS       CEY17G7    207139 bp    DNA              HTG       22-SEP-1998
DEFINITION  Caenorhabditis elegans DNA *** SEQUENCING IN PROGRESS *** from
            clone Y17G7; HTGS phase 1.
```

```
ACCESSION   Z96049
NID         g3378012
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 207139)
  AUTHORS   Smye,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-SEP-1998) Nematode Sequencing Project, Sanger Centre,
            Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics,
            Washington University, St. Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     On Aug 4, 1998 this sequence version replaced gi:2546913.
            Order of segments is not known; 800 n's separate segments.
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and the release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc.
            ***                                                          ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                          ***
            * This sequence is unfinished. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
FEATURES             Location/Qualifiers
     source          1. .207139
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /chromosome="II"
                     /clone="Y17G7"
BASE COUNT    68582 a  34583 c  35567 g  67606 t    801 others
ORIGIN

  Query Match           12.2%;  Score 18;  DB 17;  Length 207139;
  Best Local Similarity 100.0%;  Pred. No. 3.50e+00;
  Matches    18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db 102704 TATTTTAATATAATCTGT 102721
          ||||||||||||||||||
Cp    135 TATTTTAATATAATCTGT 118


RESULT   27
LOCUS       AC005406   226063 bp    DNA             PRI       04-OCT-1998
DEFINITION  , complete sequence.
ACCESSION   AC005406
NID         g3694911
```



```
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 226063)
  AUTHORS   Fu,Y., Pan,H., McDermid,H. and Roe,B.A.
  TITLE     Homo sapiens 22q11 BAC Clone b128f11 In CES Region
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 226063)
  AUTHORS   Fu,Y., Pan,H., McDermid,H. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-AUG-1998) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   3  (bases 1 to 226063)
  AUTHORS   Fu,Y., Pan,H., McDermid,H. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-SEP-1998) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   4  (bases 1 to 226063)
  AUTHORS   Fu,Y., Pan,H., McDermid,H. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-SEP-1998) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   5  (bases 1 to 226063)
  AUTHORS   Fu,Y., Pan,H., McDermid,H. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-1998) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   6  (bases 1 to 226063)
  AUTHORS   Fu,Y., Pan,H., McDermid,H. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-OCT-1998) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On Oct 4, 1998 this sequence version replaced gi:3687261.
FEATURES             Location/Qualifiers
     source          1. .226063
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT    70672 a  47539 c  43413 g  64438 t      1 others
ORIGIN

  Query Match           12.2%;  Score 18;  DB 27;  Length 226063;
  Best Local Similarity 100.0%;  Pred. No. 3.50e+00;
  Matches    18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```



```
Db 122722 TTATATTAAAATAATAAT 122739
          ||||||||||||||||||
Qy    123 TTATATTAAAATAATAAT 140


RESULT   28
LOCUS       AC005071   256404 bp    DNA             HTG       12-JUN-1998
DEFINITION  *** SEQUENCING IN PROGRESS *** Homo sapiens clone RG161A02; HTGS
            phase 1, 35 unordered pieces.
ACCESSION   AC005071
NID         g3212920
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 256404)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 256404)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-1998) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     ***                                                          ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                          ***
            * This sequence is unfinished. It consists of 35 contigs for
            * which the order is not known; their order in this record is
            * arbitrary. In some cases, the exact lengths of the gaps
            * between the contigs are also unknown; these gaps are  presented
            * as runs of N as a convenience only. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
            *       1     1600: contig of 1600 bp in length
            *    1601     1618: gap of unknown length
            *    1619     3213: contig of 1595 bp in length
            *    3214     3231: gap of unknown length
            *    3232     4831: contig of 1600 bp in length
            *    4832     4849: gap of unknown length
            *    4850     6665: contig of 1816 bp in length
            *    6666     6683: gap of unknown length
            *    6684     8472: contig of 1789 bp in length
            *    8473     8490: gap of unknown length
            *    8491    10001: contig of 1511 bp in length
            *   10002    10019: gap of unknown length
            *   10020    11650: contig of 1631 bp in length
            *   11651    11668: gap of unknown length
            *   11669    13497: contig of 1829 bp in length
```



```
            *   13498    13515: gap of unknown length
            *   13516    15944: contig of 2429 bp in length
            *   15945    15962: gap of unknown length
            *   15963    17496: contig of 1534 bp in length
            *   17497    17514: gap of unknown length
            *   17515    19092: contig of 1578 bp in length
            *   19093    19110: gap of unknown length
            *   19111    21089: contig of 1979 bp in length
            *   21090    21107: gap of unknown length
            *   21108    22787: contig of 1680 bp in length
            *   22788    22805: gap of unknown length
            *   22806    24419: contig of 1614 bp in length
            *   24420    24437: gap of unknown length
            *   24438    25932: contig of 1495 bp in length
            *   25933    25950: gap of unknown length
            *   25951    27541: contig of 1591 bp in length
            *   27542    27559: gap of unknown length
            *   27560    29831: contig of 2272 bp in length
            *   29832    29849: gap of unknown length
            *   29850    31952: contig of 2103 bp in length
            *   31953    31970: gap of unknown length
            *   31971    33770: contig of 1800 bp in length
            *   33771    33788: gap of unknown length
            *   33789    35495: contig of 1707 bp in length
            *   35496    35513: gap of unknown length
            *   35514    37491: contig of 1978 bp in length
            *   37492    37509: gap of unknown length
            *   37510    40587: contig of 3078 bp in length
            *   40588    40605: gap of unknown length
            *   40606    42880: contig of 2275 bp in length
            *   42881    42898: gap of unknown length
            *   42899    44639: contig of 1741 bp in length
            *   44640    44657: gap of unknown length
            *   44658    65048: contig of 20391 bp in length
            *   65049    65066: gap of unknown length
            *   65067    92258: contig of 27192 bp in length
            *   92259    92276: gap of unknown length
            *   92277   141560: contig of 49284 bp in length
            *  141561   141578: gap of unknown length
            *  141579   191048: contig of 49470 bp in length
            *  191049   191066: gap of unknown length
            *  191067   246503: contig of 55437 bp in length
            *  246504   246520: gap of unknown length
            *  246521   248115: contig of 1595 bp in length
            *  248116   248132: gap of unknown length
            *  248133   249837: contig of 1705 bp in length
            *  249838   249854: gap of unknown length
            *  249855   251310: contig of 1456 bp in length
            *  251311   251327: gap of unknown length
            *  251328   252893: contig of 1566 bp in length
            *  252894   252910: gap of unknown length
```

```
            *   252911   254626: contig of 1716 bp in length
            *   254627   254643: gap of unknown length
            *   254644   256404: contig of 1761 bp in length.
FEATURES             Location/Qualifiers
     source          1. .256404
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RG161A02"
BASE COUNT    64346 a  62984 c  62157 g  66233 t    684 others
ORIGIN

  Query Match           12.2%;  Score 18;  DB 17;  Length 256404;
  Best Local Similarity 100.0%;  Pred. No. 3.50e+00;
  Matches    18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db 199387 CCAGCCCTGGCCCCCCCT 199404
          ||||||||||||||||||
Qy     80 CCAGCCCTGGCCCCCCCT 97


RESULT   29
LOCUS       CFDOPRX01     338 bp    RNA             MAM       17-MAR-1992
DEFINITION  C.familiaris mRNA DOPCRX01 for olfactory receptor.
ACCESSION   X64967
NID         g885
KEYWORDS    DOPCRX01 gene; G protein coupled receptor; olfactory receptor.
SOURCE      dog.
  ORGANISM  Canis familiaris
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE   1  (bases 1 to 338)
  AUTHORS   Parmentier,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAR-1992) M. Parmentier, Universite Libre de
            Bruxelles, Campus Hopital Erasme, Route de Lennik 808, 1070
            Bruxelles, BELGIUM
REFERENCE   2  (bases 1 to 338)
  AUTHORS   Parmentier,M., Libert,F., Schurmans,S., Schiffmann,S., Lefort,A.,
            Eggerickx,D., Ledent,C., Mollereau,C., Gerard,C., Perret,J.,
            Grootegoed,J.A. and Vassart,G.
  TITLE     Expression of members of the putative olfactory receptor gene
            family in mammalian germ cells
  JOURNAL   Nature 355 (6359), 453-455 (1992)
  MEDLINE   92131132
FEATURES             Location/Qualifiers
     source          1. .338
                     /organism="Canis familiaris"
                     /db_xref="taxon:9615"
                     /tissue_type="olfactory mucosa"
                     /clone="DOPCRX01"
     misc_feature    44. .118
```



```
                     /note="transmembrane segment IV"
     misc_feature    215. .283
                     /note="transmembrane segment V"
BASE COUNT       57 a      86 c      84 g     111 t
ORIGIN

  Query Match           11.6%;  Score 17;  DB 19;  Length 338;
  Best Local Similarity 100.0%;  Pred. No. 1.71e+01;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    160 CTGTGAGGCCCCCACTC 176
          |||||||||||||||||
Qy     60 CTGTGAGGCCCCCACTC 76


RESULT   30
LOCUS       PARMTDI7B     405 bp    DNA             INV       30-SEP-1988
DEFINITION  paramecium species 7,325 mt dna dimer: replication init. region.
ACCESSION   K00919
NID         g342966
KEYWORDS    origin of replication.
SOURCE      paramecium species 7,325 mitochondrial dna.
  ORGANISM  Mitochondrion Paramecium sp.
            Eukaryotae; mitochondrial eukaryotes; Alveolata; Ciliophora;
            Nassophorea; Peniculida; Paramecium.
REFERENCE   1  (bases 1 to 405)
  AUTHORS   Pritchard,A.E., Lapping,J.L., Seilhamer,J.J. and Cummings,D.J.
  TITLE     inter-species sequence diversity in the replication initiation
            region of paramecium mitochondrial dna
  JOURNAL   J. Mol. Biol. 164, 1-15 (1983)
  MEDLINE   83189096
REFERENCE   2  (bases 1 to 405)
  AUTHORS   Pritchard,A.E.
  JOURNAL   Unpublished (1984)
COMMENT     [2]  corrected sequence for [1].
            [2] provided revised sequences for [1] because there were several
            errors in the published sequences. the sequences provided by [2]
            are used in this entry. the revisions have not been noted.
FEATURES             Location/Qualifiers
     source          1. .405
                     /organism="Paramecium sp."
                     /mitochondrion
                     /db_xref="taxon:5887"
BASE COUNT      142 a      26 c      29 g     208 t
ORIGIN

  Query Match           11.6%;  Score 17;  DB 18;  Length 405;
  Best Local Similarity 100.0%;  Pred. No. 1.71e+01;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    348 TTATTATTTTAATATAA 364
```



```
          ||||||||||||||||||
Cp    139 TTATTATTTTAATATAA 123


RESULT   31
LOCUS       S52967S4      434 bp    DNA             PRI       08-MAR-1993
DEFINITION  alpha-(1.3/1,4) fucosyltransferase/FucT-III, alpha-1,3
            fucosyltransferase/FucT-VI [human, Genomic, 434 nt, segment 4 of
            4].
ACCESSION   S52969
NID         g263663
KEYWORDS    .
SEGMENT     4 of 4
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 434)
  AUTHORS   Nishihara,S., Nakazato,M., Kudo,T., Kimura,H., Ando,T. and
            Narimatsu,H.
  TITLE     Human alpha-1,3 fucosyltransferase (FucT-VI) gene is located at
            only 13 kb 3' to the Lewis type fucosyltransferase (FucT-III) gene
            on chromosome 19
  JOURNAL   Biochem. Biophys. Res. Commun. 190 (1), 42-46 (1993)
  MEDLINE   93135789
  REMARK    GenBank staff at the National Library of Medicine created this
            entry [NCBI gibbsq 122635] from the original journal article.
            This sequence comes from Fig. 2.
            Map location: 19.
FEATURES             Location/Qualifiers
     source          1. .434
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     gene            order(S52968:82. .87,1. .5)
                     /partial
                     /gene="alpha-1,3 fucosyltransferase/FucT-VI"
     CDS             join(S52968:82. .87,1. .5)
                     /partial
                     /gene="alpha-1,3 fucosyltransferase/FucT-VI"
                     /codon_start=1
                     /exception="Only initiation and stop codons shown"
BASE COUNT       60 a     138 c     130 g     106 t
ORIGIN

  Query Match           11.6%;  Score 17;  DB 25;  Length 434;
  Best Local Similarity 100.0%;  Pred. No. 1.71e+01;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     58 CCTGAGTGGGGGCCTCA 74
          ||||||||||||||||||
Cp     79 CCTGAGTGGGGGCCTCA 63
```



```
RESULT   32
LOCUS       HSMETTRAN     769 bp    RNA             PRI       17-JAN-1995
DEFINITION  H.sapiens mRNA for methyltransferase.
ACCESSION   X54220
NID         g34558
KEYWORDS    methyltransferase.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 769)
  AUTHORS   Hayakawa,H., Koike,G. and Sekiguchi,M.
  TITLE     Expression and cloning of complementary DNA for a human enzyme that
            repairs O6-methylguanine in DNA
  JOURNAL   J. Mol. Biol. 213 (4), 739-747 (1990)
  MEDLINE   90294292
FEATURES             Location/Qualifiers
     source          1. .769
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="Raji (Mer+)"
                     /clone="1H3"
     CDS             41. .664
                     /codon_start=1
                     /product="O-6-methylguanine-DNA methyltransferase"
                     /db_xref="PID:g34559"
                     /db_xref="SWISS-PROT:P16455"
                     /translation="MDKDCEMKRTTLDSPLGKLELSGCEQGLHEIKLLGKGTSAADAV
                     EVPAPAAVLGGPEPLMQCTAWLNAYFHQPEAIEEFPVPALHHPVFQQESFTRQVLWKL
                     LKVVKFGEVISYQQLAALAGNPKAARAVGGAMRGNPVPILIPCHRVVCSSGAVGNYSG
                     GLAVKEWLLAHEGHRLGKPGLGGSSGLAGAWLKGAGATSGSPPAGRN"
     polyA_site      769
BASE COUNT      162 a     216 c     249 g     142 t
ORIGIN

  Query Match           11.6%;  Score 17;  DB 25;  Length 769;
  Best Local Similarity 100.0%;  Pred. No. 1.71e+01;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     22 CCCCTTGGTACTTGGAA 38
          ||||||||||||||||||
Qy     93 CCCCTTGGTACTTGGAA 109


RESULT   33
LOCUS       E02953        791 bp    RNA             PAT       26-NOV-1996
DEFINITION  cDNA sequence encoding human
            O6-methylguanine-DNA-methyltrausferase.
ACCESSION   E02953
```

```
NID         g2171175
KEYWORDS    JP 1991147789-A/1.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 791)
  AUTHORS   Mutsuo,S. and Hiroshi,H. .
  TITLE     CDNA CLONE CAPABLE OF CODING HUMAN O6-METHYLGUANINE-DNA-
            METHYLTRANSFERASE
  JOURNAL   Patent: JP 1991147789-A 1 24-JUN-1991;
            RES DEV CORP OF JAPAN
COMMENT     OS   Homo sapiens (human)
            PN   JP 1991147789-A/1
            PD   24-JUN-1991
            PF   02-NOV-1989 JP 1989287223
            PI   SEKIGUCHI MUTSUO, HAYAKAWA HIROSHI
            PC   C12N15/54;
            CC   strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            CC   *source: cell_line=Raji;
            FH   Key             Location/Qualifiers
            FH
            FT   5'UTR           1. .40
            FT   CDS             41. .664
            FT                   /product='human
            FT                   O6-methylguanine-DNA-methyltransferase' FT
            3'UTR          665. .791.
FEATURES            Location/Qualifiers
     source         1. .791
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
BASE COUNT      184 a    215 c    247 g    145 t
ORIGIN

  Query Match           11.6%;  Score 17;  DB 21;  Length 791;
  Best Local Similarity 100.0%;  Pred. No. 1.71e+01;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      22 CCCCTTGGTACTTGGAA 38
           |||||||||||||||||
Qy      93 CCCCTTGGTACTTGGAA 109

RESULT   34
LOCUS       HSU27332      915 bp    mRNA            PRI       31-AUG-1995
DEFINITION  Human alpha (1,3) fucosyltransferase (FUT6) mRNA, isoform II,
            complete cds.
```



```
ACCESSION   U27332
NID         g967200
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 915)
  AUTHORS   Cameron,H.S., Szczepaniak,D. and Weston,B.W.
  TITLE     Expression of human chromosome 19p alpha(1,3)-fucosyltransferase
            genes in normal tissues. Alternative splicing, polyadenylation, and
            isoforms
  JOURNAL   J. Biol. Chem. 270 (34), 20112-20122 (1995)
  MEDLINE   95378269
REFERENCE   2  (bases 1 to 915)
  AUTHORS   Weston,B.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAY-1995) Brent W. Weston, Pediatrics, University of
            North Carolina at Chapel Hill, Division of Hematology/Oncology, 418
            MacNider Building, Chapel Hill, NC 27599-7220, USA
FEATURES            Location/Qualifiers
     source         1. .915
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /map="19p"
                    /chromosome="19"
                    /clone="FUT6 isoform II"
                    /tissue_type="kidney"
     gene           1. .915
                    /gene="FUT6"
     CDS            1. .915
                    /gene="FUT6"
                    /codon_start=1
                    /product="alpha (1,3) fucosyltransferase"
                    /db_xref="PID:g967201"
                    /translation="MDPLGPAKPQWSWRCCLTTLLFQLLMAVCFFSYLRVSQDDPTVY
                    PNGSRFPDSTGTPAHSIPLILLWTWPFNKPIALPRCSEMVPGTADCNITADRKVYPQA
                    DAVIVHHREVMYNPSAQLPRSPRRQGQRWIWFSMESPSHCRQLEAMDGYFNLTMSYRS
                    DSDIFTPYGWLEPWSGQPAHPPLNLSAKTELVAWAVSNWGPNSARVRYYQSLQAHLKV
                    DVYGRSHKPLPQGTMMETLSRYKFYLAFENSLHPDYITEKLWRNALEAWAVPVVLGPS
                    RRNLIFLGPHLSGGLIYLRTRLPEASPA"
BASE COUNT      170 a    318 c    251 g    176 t
ORIGIN

  Query Match           11.6%;  Score 17;  DB 27;  Length 915;
  Best Local Similarity 100.0%;  Pred. No. 1.71e+01;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     858 CCTGAGTGGGGGCCTCA 874
           |||||||||||||||||
Cp      79 CCTGAGTGGGGGCCTCA 63
```



```
RESULT   35
LOCUS       OCU04276      952 bp    mRNA            MAM       06-APR-1994
DEFINITION  Oryctolagus cuniculus New Zealand white clone 80A prostaglandin EP3
            receptor subtype mRNA, partial cds.
ACCESSION   U04276
NID         g467984
KEYWORDS    .
SOURCE      European rabbit.
  ORGANISM  Oryctolagus cuniculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
REFERENCE   1  (bases 1 to 952)
  AUTHORS   Breyer,R.M., Emeson,R.B., Tarng,J.L., Breyer,M.D., Davis,L.S.,
            Abromson,R.M. and Ferrenbach,S.M.
  TITLE     Alternative splicing generates multiple isoforms of a rabbit
            prostaglandin E2 receptor
  JOURNAL   J. Biol. Chem. 269, 6163-6169 (1994)
  MEDLINE   94164982
REFERENCE   2  (bases 1 to 952)
  AUTHORS   Breyer,R.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-1993) Richard M. Breyer, Division of Nephrology,
            Vanderbilt University, 1161 21st Ave. South, Nashville, TN
            37232-2372, USA
FEATURES            Location/Qualifiers
     source         1. .952
                    /organism="Oryctolagus cuniculus"
                    /strain="New Zealand white"
                    /db_xref="taxon:9986"
                    /clone="80A"
                    /clone_lib="Lambda ZAP II, Stratagene"
                    /tissue_type="kidney cortex"
     CDS            <1. .655
                    /codon_start=2
                    /product="prostaglandin EP3 receptor subtype"
                    /db_xref="PID:g467985"
                    /translation="VLLGVWLAVLAFALLPVLGVGQYTIQWPGTWCFISTGRGDNGTS
                    SSHNWGNLFFASTFAFLGLLALAITFACNLATIKALVSRCRAKAAASQSSAQWGRITT
                    ETAIQLMGIMCVLSVCWSPLLIMMLKMIFNQTSVEHCKTDTGKQKECNFFLIAVRLAS
                    LNQILDPWVYLLLRKILLRKFFCQHSPAIGDLQISTHISKTNKYFEGLMKTFHSLAYL"
BASE COUNT      245 a    224 c    205 g    278 t
ORIGIN

  Query Match           11.6%;  Score 17;  DB 19;  Length 952;
  Best Local Similarity 100.0%;  Pred. No. 1.71e+01;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     761 TATATTAAAATAATAAT 777
           |||||||||||||||||
```



```
Qy     124 TATATTAAAATAATAAT 140

RESULT   36
LOCUS       PTY14035      1212 bp    DNA             PRI       01-DEC-1997
DEFINITION  Pan troglodytes FUT6 gene.
ACCESSION   Y14035
NID         g2655931
KEYWORDS    alpha-3-L-fucosyltransferase; FUT6 gene.
SOURCE      chimpanzee.
  ORGANISM  Pan troglodytes
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 1212)
  AUTHORS   Costache,M., Apoil,P.A., Cailleau,A., Elmgren,A., Larson,G.,
            Henry,S., Blancher,A., Iordachescu,D., Oriol,R. and Mollicone,R.
  TITLE     Evolution of fucosyltransferase genes in vertebrates
  JOURNAL   J. Biol. Chem. 272 (47), 29721-29728 (1997)
  MEDLINE   98037800
REFERENCE   2  (bases 1 to 1212)
  AUTHORS   Costache,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUN-1997) M. Costache, U-178 INSERM, University of
            Paris XI, 16, Avenue Paul Vaillant-Couturier, 94807, Villejuif
            CEDEX, FRANCE
COMMENT     Related sequence: L01698.
FEATURES            Location/Qualifiers
     source         1. .1212
                    /organism="Pan troglodytes"
                    /db_xref="taxon:9598"
                    /cell_type="leukocytes"
     gene           34. .1113
                    /gene="FUT6"
     CDS            34. .1113
                    /gene="FUT6"
                    /EC_number="2.4.1.152"
                    /note="allele A"
                    /codon_start=1
                    /evidence=experimental
                    /product="alpha-3-L-fucosyltransferase"
                    /db_xref="PID:e1193490"
                    /db_xref="PID:g2655932"
                    /translation="MDPLGPARPQWSWRCCLTTLLFQLLVAVCFFSYLRVSRDDPTVY
                    PNGSHFPDSTGTPAHSIPLILLWTWPFNKPIALPRCSEMVPGTADCNITADRKVYPQA
                    DAVIVHHREVMYNPSAQLPRSPRRQGQRWIWFSMESPSNCRHLEALDGYFNLTMSYRS
                    DSDIFTPYGWLQPWSGQPVHPPLNLSAKTELVAWAVSNWGPNSARVRYYQSLQAHLKV
                    DVYGRSHKPLPQGTMMETLSRYKFYLAFENSLHPDYITEKLWRNALEAWAVPVVLGPS
                    RSNYERFLPPDAFIHVDDPQSPKDLARYLQELDKDHARYLSYFRWRETLRPRFFSWAL
                    AFCKACWKLQEESRYQTRSIAAWFT"
     allele         404
                    /gene="FUT6"
```

```
                     /note="allele B: amino acid 124 P -> L"
                     /replace="t"
     allele          547
                     /gene="FUT6"
                     /note="allele B: amino acid 172 Q -> E"
                     /replace="g"
     allele          608
                     /gene="FUT6"
                     /note="allele B: amino acid 192 A -> C"
                     /replace="c"
BASE COUNT      217 a    423 c    342 g    230 t
ORIGIN

  Query Match           11.6%;  Score 17;  DB 26;  Length 1212;
  Best Local Similarity 100.0%;  Pred. No. 1.71e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1166 CCTGAGTGGGGGCCTCA 1182
           ||||||||||||||||||
Cp     79 CCTGAGTGGGGGCCTCA 63


RESULT   37
LOCUS       HUMFUT41A     1287 bp    DNA             PRI       27-JUN-1996
DEFINITION  Homo sapiens galactoside 3-fucosyltransferase (FUT6) gene, complete
            cds.
ACCESSION   L01698
NID         g182792
KEYWORDS    galactoside 3-fucosyltransferase; sialyl Lewis-x.
SOURCE      Homo sapiens DNA.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1287)
  AUTHORS   Weston,B.W., Smith,P.L., Kelly,R.J. and Lowe,J.B.
  TITLE     Molecular cloning of a fourth member of a human alpha
            (1,3)fucosyltransferase gene family. Multiple homologous sequences
            that determine expression of the Lewis x, sialyl Lewis x, and
            difucosyl sialyl Lewis x epitopes
  JOURNAL   J. Biol. Chem. 267 (34), 24575-24584 (1992)
  MEDLINE   93077550
  REMARK    Erratum:[J Biol Chem 1993 Aug 25;268(24):18398]]
FEATURES             Location/Qualifiers
     source          1. .1287
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="19p"
                     /chromosome="19"
     gene            97. .1176
                     /gene="FUT6"
     CDS             97. .1176
```



```
                     /gene="FUT6"
                     /note="alias FUC-TVI"
                     /codon_start=1
                     /product="galactoside 3-fucosyltransferase"
                     /db_xref="PID:g1280210"
                     /db_xref="GDB:GDB:131563"
                     /translation="MDPLGPARPQWSWRCCLTTLLPQLLMAVCFFSYLRVSQDDPTVI
                     PNGSRFPDSTGTPAHSIPLILLWTWPFNKPIALPRCSEMVPGTADCNITADRKVTPQA
                     DAVIVHHREVMYNPSAQLPRSPRRQGQRWIWFSMESPSHCWQLKAMDGIFNLTMSYRS
                     DSDIFTPTGWLEPWSGQPAHPPLNLSAKTELVAWAVSNWGPNSARVRYYQSLQAHLKV
                     DVYGRSHKPLPQGTMMETLSRYKFYLAFENSLHPDYITEKLWRNALEAWAVPVVLGPS
                     RSNYERFLPPDAFIHVDDFQSPKDLARYLQELDKDHARYLSYFRWRETLRPRSFSWAL
                     AFCKACWKLQEESRYQTRGIAAWFT"
BASE COUNT      240 a    442 c    356 g    249 t
ORIGIN

  Query Match           11.6%;  Score 17;  DB 25;  Length 1287;
  Best Local Similarity 100.0%;  Pred. No. 1.71e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1229 CCTGAGTGGGGGCCTCA 1245
           ||||||||||||||||||
Cp     79 CCTGAGTGGGGGCCTCA 63


RESULT   38
LOCUS       XELXP42       1468 bp    mRNA            VRT       30-JUL-1991
DEFINITION  X.laevis myelin basic protein kinase-like protein (Xp42) mRNA,
            complete cds.
ACCESSION   M60977
NID         g214930
KEYWORDS    myelin basic protein kinase-like protein.
SOURCE      Xenopus laevis oocyte, cDNA to mRNA.
  ORGANISM  Xenopus laevis
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea;
            Pipidae; Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 1468)
  AUTHORS   Posada,J., Sanghera,J., Pelech,S., Aebersold,R.H. and Cooper,J.A.
  TITLE     Tyrosine phosphorylation and activation of homologous protein
            kinases during Oocyte maturation and mitogenic activation of
            fibroblasts
  JOURNAL   Mol. Cell. Biol. 11, 2517-2528 (1991)
  MEDLINE   91203872
FEATURES             Location/Qualifiers
     source          1. .1468
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /dev_stage="oocyte"
     gene            42. .1127
                     /gene="Xp42"
```



```
     CDS             42. .1127
                     /gene="Xp42"
                     /note="An alternative initiation codon is located at
                     position 87-89."
                     /codon_start=1
                     /product="myelin basic protein kinase-like protein"
                     /db_xref="PID:g214931"
                     /translation="MAAAAASSNPGGGPEMVRGQAFDVGPRYTNLSYIGEGAYGMVCS
                     AHCNINKVRVAIKKISPFEHQTYCQRTLREIKILLRFKHENIIGINDIIRAPTIEQMK
                     DVYIVQDLMETDLYKLLKTQHLSNDHICYFLYQILRGLKYIHSANVLHRDLKPSNLLL
                     NTTCDLKICDFGLARVADPDHDHTGFLTEYVATRWYRAPEIMLNSKGYTKSIDIWSVG
                     CILAEMLSNRPIFPGKHYLDQLNHILGILGSPSQEDLNCIINLKARNYLLSLPHKNKV
                     PWNRLFPNADPKALDLLDKMLTFNPHKRIEVEAALAHPYLEQYYDPSDEPVAEAPFKF
                     EMELDDLPKETLKELIFEETARFQPGY"
BASE COUNT      402 a    329 c    312 g    425 t
ORIGIN

  Query Match           11.6%;  Score 17;  DB 20;  Length 1468;
  Best Local Similarity 100.0%;  Pred. No. 1.71e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1390 TTGTACATTATTATTTT 1406
           |||||||||||||||||
Cp    146 TTGTACATTATTATTTT 130


RESULT   39
LOCUS       I34194        1654 bp    DNA             PAT       27-JAN-1997
DEFINITION  Sequence 13 from patent US 5595900.
ACCESSION   I34194
NID         g1824985
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1654)
  AUTHORS   Lowe,J.B.
  TITLE     Methods and products for the synthesis of oligosaccharide
            structures on glycoproteins, glycolipids, or as free molecules, and
            for the isolation of cloned genetic sequences that determine these
            structures
  JOURNAL   Patent: US 5595900-A 13 21-JAN-1997;
FEATURES             Location/Qualifiers
     source          1. .1654
                     /organism="unknown"
BASE COUNT      298 a    557 c    459 g    340 t
ORIGIN

  Query Match           11.6%;  Score 17;  DB 21;  Length 1654;
  Best Local Similarity 100.0%;  Pred. No. 1.71e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```



```
Db   1261 CCTGAGTGGGGGCCTCA 1277
           ||||||||||||||||||
Cp     79 CCTGAGTGGGGGCCTCA 63


RESULT   40
LOCUS       HSU27336      2145 bp    mRNA            PRI       31-AUG-1995
DEFINITION  Human alpha (1,3) fucosyltransferase (FUT6) mRNA, minor transcript
            II, complete cds.
ACCESSION   U27336
NID         g967208
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2145)
  AUTHORS   Cameron,H.S., Szczepaniak,D. and Weston,B.W.
  TITLE     Expression of human chromosome 19p alpha(1,3)-fucosyltransferase
            genes in normal tissues. Alternative splicing, polyadenylation, and
            isoforms
  JOURNAL   J. Biol. Chem. 270 (34), 20112-20122 (1995)
  MEDLINE   95378269
REFERENCE   2  (bases 1 to 2145)
  AUTHORS   Weston,B.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAY-1995) Brent W. Weston, Pediatrics, University of
            North Carolina at Chapel Hill, Division of Hematology/Oncology, 418
            MacNider Building, Chapel Hill, NC 27599-7220, USA
FEATURES             Location/Qualifiers
     source          1. .2145
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="19p"
                     /chromosome="19"
                     /clone="minor transcript II"
                     /tissue_type="colon, liver"
     gene            352. .1431
                     /gene="FUT6"
     CDS             352. .1431
                     /gene="FUT6"
                     /codon_start=1
                     /product="alpha (1,3) fucosyltransferase"
                     /db_xref="PID:g967209"
                     /translation="MDPLGPARPQWSWRCCLTTLLPQLLMAVCFFSYLRVSQDDPTVI
                     PNGSRFPDSTGTPAHSIPLILLWTWPFNKPIALPRCSEMVPGTADCNITADRKVTPQA
                     DAVIVHHREVMYNPSAQLPRSPRRQGQRWIWFSMESPSHCWQLKAMDGIFNLTMSYRS
                     DSDIFTPTGWLEPWSGQPAHPPLNLSAKTELVAWAVSNWGPNSARVRYYQSLQAHLKV
                     DVYGRSHKPLPQGTMMETLSRYKFYLAFENSLHPDYITEKLWRNALEAWAVPVVLGPS
                     RSNYERFLPPDAFIHVDDFQSPKDLARYLQELDKDHARYLSYFRWRETLRPRSFSWAL
```

```
                    APCKACWKLQEESRYQTRGIAAWFT"
BASE COUNT      400 a    689 c    604 g    452 t
ORIGIN

  Query Match           11.6%;  Score 17;  DB 27;  Length 2145;
  Best Local Similarity 100.0%;  Pred. No. 1.71e+01;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1484 CCTGAGTGGGGGCCTCA 1500
         ||||||||||||||||||
Cp     79 CCTGAGTGGGGGCCTCA 63


RESULT   41
LOCUS       HSU27334     2270 bp    mRNA            PRI       31-AUG-1995
DEFINITION  Human alpha (1,3) fucosyltransferase (FUT6) mRNA, major transcript
            II, complete cds.
ACCESSION   U27334
NID         g967204
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2270)
  AUTHORS   Cameron,H.S., Szczepaniak,D. and Weston,B.W.
  TITLE     Expression of human chromosome 19p alpha(1,3)-fucosyltransferase
            genes in normal tissues. Alternative splicing, polyadenylation, and
            isoforms
  JOURNAL   J. Biol. Chem. 270 (34), 20112-20122 (1995)
  MEDLINE   95378269
REFERENCE   2  (bases 1 to 2270)
  AUTHORS   Weston,B.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAY-1995) Brent W. Weston, Pediatrics, University of
            North Carolina at Chapel Hill, Division of Hematology/Oncology, 418
            MacNider Building, Chapel Hill, NC 27599-7220, USA
FEATURES             Location/Qualifiers
     source          1. .2270
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="19p"
                     /chromosome="19"
                     /clone="major transcript II"
                     /tissue_type="colon, liver"
     gene            477. .1556
                     /gene="FUT6"
     CDS             477. .1556
                     /gene="FUT6"
                     /codon_start=1
                     /product="alpha (1,3) fucosyltransferase"
```



```
                     /db_xref="PID:g967205"
                     /translation="MDPLGPAKPQWSWRCCLTTLLFQLLMAVCFFSYLRVSQDDPTVY
                     PNGSRFPDSTGTPAHSIPLILLWTWPFNKPIALPRCSEMVPGTADCNITADRKVTPQA
                     DAVIVHHREVMYNPSAQLPRSPRRQGQRWIWFSMESPSHCWQLKAMDGYFNLTMSYRS
                     DSDIFTPYGWLEPWSGQPAHPPLNLSAKTELVAWAVSNWGPNSARVRYYQSLQAHLKV
                     DVYGRSHKPLPQGTMMETLSRYKFYLAFENSLHPDYITEKLWRNALEAWAVPVVLGPS
                     RSNYERFLPPDAFIHVDDFQSPKDLARYLQELDKDHARYLSTFRWRETLRPRSFSWAL
                     APCKACWKLQEESRYQTRGIAAWFT"
BASE COUNT      428 a    727 c    639 g    476 t
ORIGIN

  Query Match           11.6%;  Score 17;  DB 27;  Length 2270;
  Best Local Similarity 100.0%;  Pred. No. 1.71e+01;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1609 CCTGAGTGGGGGCCTCA 1625
         ||||||||||||||||||
Cp     79 CCTGAGTGGGGGCCTCA 63


RESULT   42
LOCUS       HSU27337     2612 bp    mRNA            PRI       31-AUG-1995
DEFINITION  Human alpha (1,3) fucosyltransferase (FUT6) mRNA, minor transcript
            III, complete cds.
ACCESSION   U27337
NID         g967210
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2612)
  AUTHORS   Cameron,H.S., Szczepaniak,D. and Weston,B.W.
  TITLE     Expression of human chromosome 19p alpha(1,3)-fucosyltransferase
            genes in normal tissues. Alternative splicing, polyadenylation, and
            isoforms
  JOURNAL   J. Biol. Chem. 270 (34), 20112-20122 (1995)
  MEDLINE   95378269
REFERENCE   2  (bases 1 to 2612)
  AUTHORS   Weston,B.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAY-1995) Brent W. Weston, Pediatrics, University of
            North Carolina at Chapel Hill, Division of Hematology/Oncology, 418
            MacNider Building, Chapel Hill, NC 27599-7220, USA
FEATURES             Location/Qualifiers
     source          1. .2612
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="19p"
                     /chromosome="19"
                     /clone="minor transcript III"
```



```
                     /tissue_type="kidney"
     gene            639. .1718
                     /gene="FUT6"
     CDS             639. .1718
                     /gene="FUT6"
                     /codon_start=1
                     /product="alpha (1,3) fucosyltransferase"
                     /db_xref="PID:g967211"
                     /translation="MDPLGPAKPQWSWRCCLTTLLFQLLMAVCFFSYLRVSQDDPTVY
                     PNGSRFPDSTGTPAHSIPLILLWTWPFNKPIALPRCSEMVPGTADCNITADRKVTPQA
                     DAVIVHHREVMYNPSAQLPRSPRRQGQRWIWFSMESPSHCWQLKAMDGYFNLTMSYRS
                     DSDIFTPYGWLEPWSGQPAHPPLNLSAKTELVAWAVSNWGPNSARVRYYQSLQAHLKV
                     DVYGRSHKPLPQGTMMETLSRYKFYLAFENSLHPDYITEKLWRNALEAWAVPVVLGPS
                     RSNYERFLPPDAFIHVDDFQSPKDLARYLQELDKDHARYLSTFRWRETLRPRSFSWAL
                     APCKACWKLQEESRYQTRGIAAWFT"
BASE COUNT      510 a    814 c    735 g    553 t
ORIGIN

  Query Match           11.6%;  Score 17;  DB 27;  Length 2612;
  Best Local Similarity 100.0%;  Pred. No. 1.71e+01;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1771 CCTGAGTGGGGGCCTCA 1787
         ||||||||||||||||||
Cp     79 CCTGAGTGGGGGCCTCA 63


RESULT   43
LOCUS       ATHCRBAA     2682 bp    DNA             PLN       06-AUG-1990
DEFINITION  A.thaliana 12S storage protein CRA1 gene, exons 1-4.
ACCESSION   M37248
NID         g166677
KEYWORDS    12S storage protein.
SOURCE      A.thaliana, cDNA to mRNA.
  ORGANISM  Arabidopsis thaliana
            Eukaryotae; mitochondrial eukaryotes; Viridiplantae;
            Charophyta/Embryophyta group; Embryophyta; Magnoliophyta;
            Magnoliopsida; Capparales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 2682)
  AUTHORS   Pang,P.P., Pruitt,R.E. and Meyerowitz,E.M.
  TITLE     Molecular cloning, genomic organization, expression and evolution
            of 12S seed storage protein genes of Arabidopsis thaliana
  JOURNAL   Plant Mol. Biol. 11, 805-820 (1988)
FEATURES             Location/Qualifiers
     source          1. .2682
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
     CDS             join(378. .660,871. .1133,1445. .1873,2113. .2505)
                     /note="12S storage protein CRB"
                     /codon_start=1
                     /db_xref="PID:g166678"
```



```
                     /translation="MGRVSSIISFSLTLLILFNGYTAQQWPNECQLDQLNALEPSQII
                     KSEGGRIEVWDHHAPQLRCSGPAFERFVIEPQGLPLPTFLNAGKLTFVVHGRGLMGRV
                     IPGCAETFMESPVFGEGQGQQQSQGFRDMHQKVEHLRCGDTIATPSGVAQWFTNNGNE
                     PLILVAAADLASNQNQLDRNLRPFLIAGNNPQGQEWLQGRKQQKQNNIFNGFAPEILA
                     QAFKINVETAQQLQNQQDNRGNIVKVNGPFGVIRPPLRRGEGGQQQPHEIANGLEETLC
                     TMRCTENLDDPSDADVYKPSLGYISTLNSYNLPILRLLRLSALRGSIRKNAMVLPQWN
                     VNANALLTWGKAHIQMVNDWGERVFDQEISSGQLLVVPQGFSVMKHRIGEQFEWIE
                     FKTNENAQVNTLAGRTSVMRGLPLEVITNGYQISPEEAKRVKFSTIETTLTHSSPMSY
                     GRPRA"
     exon            <378. .660
                     /note="12S storage protein CRB"
                     /number=1
     intron          661. .870
                     /note="12S storage protein CRB intron A"
     exon            871. .1133
                     /number=2
     intron          1134. .1444
                     /note="12S storage protein CRB intron B"
     exon            1445. .1873
                     /number=3
     intron          1874. .2112
                     /note="12S storage protein CRB intron C"
     exon            2113. .>2505
                     /note="12S storage protein CRB"
                     /number=4
BASE COUNT      853 a    525 c    514 g    790 t
ORIGIN

  Query Match           11.6%;  Score 17;  DB 23;  Length 2682;
  Best Local Similarity 100.0%;  Pred. No. 1.71e+01;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2573 AAATAATAATGTACAAA 2589
         ||||||||||||||||||
Qy    131 AAATAATAATGTACAAA 147


RESULT   44
LOCUS       ATCRB        2682 bp    DNA             PLN       18-MAY-1995
DEFINITION  Arabidopsis CRB gene for 12S seed storage protein.
ACCESSION   X14313
NID         g16236
KEYWORDS    CRB gene; seed storage protein; storage protein.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryotae; mitochondrial eukaryotes; Viridiplantae;
            Charophyta/Embryophyta group; Embryophyta; Magnoliophyta;
            Magnoliopsida; Capparales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 2682)
  AUTHORS   Pang,P.P., Pruitt,R.E. and Meyerowitz,E.M.
  TITLE     Molecular cloning, genomic organization, expression and evolution
```

```
            of 12S seed storage protein genes of Arabidopsis thaliana
  JOURNAL   Plant Mol. Biol. 11, 805-820 (1988)
FEATURES             Location/Qualifiers
     source          1. .2682
                     /organism="Arabidopsis thaliana"
                     /cultivar="Landsberg erecta"
                     /db_xref="taxon:3702"
                     /clone_lib="lambda EMBL4"
     promoter        314. .323
                     /note="TATA-box"
     exon            378. .660
                     /number=1
     CDS             join(378. .660,871. .1133,1445. .1873,2113. .2505)
                     /codon_start=1
                     /product="12S seed storage protein"
                     /db_xref="PID:g808937"
                     /db_xref="SWISS-PROT:P15456"
                     /translation="MGRVSSIISFSLTLLILFNGYTAQQWPNECQLDQLNALEPSQII
                     KSEGGRIEVWDHHAPQLRCSGFAFERFVIEPQGLFLPTFLNAGKLTFVVHGRGLMGRV
                     IPGCAETFMESPVFGEGQGQGQSQGFRDMHQKVEHLRCGDTIIATPSGVAQWFYNNGNE
                     PLILVAAADLASNQNQLDRNLRPFLIAGNNPQGQEWLQGRKQQKQNNIFNGFAPEILA
                     QAFKINVETAQQLQNQQDNRGNIVKVNGPFGVIRPPLRRGEGGQQPHEIANGLEETLC
                     TMRCTENLDDPSDADVTKPSLGYISTLNSYNLPILRLLRLSALRGSIRKNAMVLPQWN
                     VNANAALYVTNGKAHIQMVNDNGERVFDQEISSGQLLVVPQGFSVMKRHIGEQFEWIE
                     FKTNENAQVNTLAGRTSVMRGLPLEVITNGYQISPEEAKRVKFSTIETTLTHSSPMSY
                     GRPRA"
     intron          661. .870
                     /number=1
     exon            871. .1133
                     /number=2
     intron          1134. .1444
                     /number=2
     exon            1445. .1873
                     /number=3
     intron          1874. .2112
                     /number=3
     exon            2113. .2505
                     /number=4
     misc_feature    2574. .2578
                     /note="pot. polyA signal"
     polyA_signal    2574. .2578
BASE COUNT      853 a    525 c    514 g    790 t
ORIGIN

  Query Match           11.6%;  Score 17;  DB 23;  Length 2682;
  Best Local Similarity 100.0%;  Pred. No. 1.71e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2573 AAATAATAATGTACAAA 2589
          |||||||||||||||||
Qy    131 AAATAATAATGTACAAA 147
```



```
RESULT   45
LOCUS       A36402       2973 bp    DNA             PAT       05-MAR-1997
DEFINITION  Sequence 1 from Patent WO9324639.
ACCESSION   A36402
NID         g2293775
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 2973)
  AUTHORS   Schaaf,J. and Walter,M.H.
  TITLE     EXPRESSION OF GENES IN TRANSGENIC PLANTS
  JOURNAL   Patent: WO 9324639-A 1 09-DEC-1993;
            ZENECA LTD (GB)
COMMENT     Other publication AU 4334693 931230.
FEATURES             Location/Qualifiers
     source          1. .2973
                     /organism="unidentified"
                     /db_xref="taxon:32644"
                     /clone="GPVME1"
BASE COUNT     1077 a    413 c    424 g    1059 t
ORIGIN

  Query Match           11.6%;  Score 17;  DB 21;  Length 2973;
  Best Local Similarity 100.0%;  Pred. No. 1.71e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1694 ATTATTATTTTAATATA 1710
          |||||||||||||||||
Cp    140 ATTATTATTTTAATATA 124


RESULT   46
LOCUS       XLMPK1G      2997 bp    RNA             VRT       21-AUG-1991
DEFINITION  X.laevis MPK1 mRNA for MAP kinase.
ACCESSION   X59813
NID         g64893
KEYWORDS    M phase activation; MAP kinase; MBP kinase; microtubule-associated
            protein 2; mitogen-activated protein kinase; MPK1 gene; protein
            kinase.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea;
            Pipidae; Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 2997)
  AUTHORS   Moriyama,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-MAY-1991) K. Moriyama, Dept of Biophysics &
```



```
            Biochemistry, Faculty of Science, University of Tokyo, Hongo 7-3-1,
            Bunkyoku, Tokyo, Japan
REFERENCE   2  (bases 1 to 1478)
  AUTHORS   Gotoh,Y., Moriyama,K., Matsuda,S., Okumura,E., Kishimoto,T.,
            Kawasaki,H., Suzuki,K., Yahara,I., Sakai,H. and Nishida,E.
  TITLE     Xenopus M phase MAP kinase: isolation of its cDNA and activation by
            MPF
  JOURNAL   EMBO J. 10 (9), 2661-2668 (1991)
  MEDLINE   91330892
FEATURES             Location/Qualifiers
     source          1. .2997
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /germline
                     /dev_stage="oocyte"
                     /clone="xmk4"
     mRNA            1. .2997
                     /gene="MPK1"
     gene            1. .2997
                     /gene="MPK1"
     CDS             231. .1316
                     /gene="MPK1"
                     /codon_start=1
                     /product="MAP kinase"
                     /db_xref="PID:g64894"
                     /db_xref="SWISS-PROT:P26696"
                     /translation="MAAAGAASNPGGGPEMVRGQAFDVGPRYINLAYIGEGAYGMVCS
                     AHDNVNKVRVAIKKISPFEHQTYCQRTLREIKILLRFKHENIIGINDIIRAPTIEQMK
                     DVYIVQDLMETDLYKLLKTQHLSNDHICYFLYQILRGLKYIHSANVLHRDLKPSNLLL
                     NTTCDLKICDFGLARVADPDHDHTGFLTEYVATRWYRAPEIMLNSKGYTKSIDIWSVG
                     CILAEMLSNRPIFPGKHYLDQLNHILGILGSPSQEDLNCIINLKARNYLLSLPHKNKV
                     PWNRLFPNADPKALDLLDKMLTFNPHKRIEVEAALAHPYLEQYYDPSDEPVAEAPFKF
                     EMELDDLPKETLKELIFEETARFQPGY"
     polyA_signal    2018. .2023
                     /gene="MPK1"
     polyA_signal    2929. .2935
                     /gene="MPK1"
     polyA_site      2951
                     /gene="MPK1"
BASE COUNT      833 a    652 c    598 g    914 t
ORIGIN

  Query Match           11.6%;  Score 17;  DB 20;  Length 2997;
  Best Local Similarity 100.0%;  Pred. No. 1.71e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1584 TTGTACATTATTATTTT 1600
          |||||||||||||||||
Cp    146 TTGTACATTATTATTTT 130
```



```
RESULT   47
LOCUS       HSU27335     3002 bp    mRNA            PRI       31-AUG-1995
DEFINITION  Human alpha (1,3) fucosyltransferase (FUT6) mRNA, minor transcript
            I, complete cds.
ACCESSION   U27335
NID         g967206
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3002)
  AUTHORS   Cameron,H.S., Szczepaniak,D. and Weston,B.W.
  TITLE     Expression of human chromosome 19p alpha(1,3)-fucosyltransferase
            genes in normal tissues. Alternative splicing, polyadenylation, and
            isoforms
  JOURNAL   J. Biol. Chem. 270 (34), 20112-20122 (1995)
  MEDLINE   95378269
REFERENCE   2  (bases 1 to 3002)
  AUTHORS   Weston,B.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAY-1995) Brent W. Weston, Pediatrics, University of
            North Carolina at Chapel Hill, Division of Hematology/Oncology, 418
            MacNider Building, Chapel Hill, NC 27599-7220, USA
FEATURES             Location/Qualifiers
     source          1. .3002
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="19p"
                     /chromosome="19"
                     /clone="minor transcript I"
                     /tissue_type="kidney, liver"
     gene            1029. .2108
                     /gene="FUT6"
     CDS             1029. .2108
                     /gene="FUT6"
                     /codon_start=1
                     /product="alpha (1,3) fucosyltransferase"
                     /db_xref="PID:g967207"
                     /translation="MDPLGPAKPQWSWRCCLTTLLFQLLMAVCFFSYLRVSQDDPTVY
                     PNGSRFPPDSTGTPAHSIPLILLWTWPFNKPIALPRCSEMVPGTADCNITADRKVYPQA
                     DAVIVHHREVMYNPSAQLPRSPRRQGQRWIWFSMESPSHCWQLKAMDGYFNLTMSYRS
                     DSDIFTPYGWLEPWSGQPAHPPLNLSAKTELVAWAVSNWGPNSARVRYYQSLQAHLKV
                     DVYGRSHKPLPQGTMMETLSRYKFYLAFENSLHPDYITEKLWRNALEAWAVPVVLGPS
                     RSNYERFLPPDAFIHVDDFQSPKDLARYLQELDKDHARYLSYFRWRETLRPRSFSWAL
                     APCKACWKLQEESRYQTRGIAAWFT"
BASE COUNT      591 a    916 c    835 g    660 t
ORIGIN

  Query Match           11.6%;  Score 17;  DB 27;  Length 3002;
  Best Local Similarity 100.0%;  Pred. No. 1.71e+01;
```

```
    Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2161 CCTGAGTGGGGGCCTCA 2177
          |||||||||||||||||
Cp     79 CCTGAGTGGGGGCCTCA 63


RESULT   48
LOCUS       DDU60169     3085 bp    DNA             INV       07-JUL-1996
DEFINITION  Dictyostelium discoideum putative protein kinase MkcA (mkcA) gene,
            complete cds.
ACCESSION   U60169
NID         g1405365
KEYWORDS    .
SOURCE      Dictyostelium discoideum.
  ORGANISM  Dictyostelium discoideum
            Eukaryotae; mitochondrial eukaryotes; Dictyosteliida;
            Dictyostelium.
REFERENCE   1  (bases 1 to 3085)
  AUTHORS   Shaulsky,G., Escalante,R. and Loomis,W.F.
  TITLE     Developmental signal transduction pathways uncovered by genetic
            suppressors
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3085)
  AUTHORS   Shaulsky,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUN-1996) Dept. of Biology 0322, University of
            California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-032,
            USA
COMMENT     mkcA was cloned as a morphological suppresser of tagB. mkcA null
            strains are essentially indistinguishable from the wild type.  mkcA
            mRNA is expressed at equal levels in vegetative cells and in both
            prespore and prestalk cellls.  The predicted protein is a putative
            protein kinase, homologous to STE20 and PAK65.
FEATURES             Location/Qualifiers
     source          1. .3085
                     /organism="Dictyostelium discoideum"
                     /db_xref="taxon:44689"
                     /cell_line="AX4"
                     /chromosome="4"
                     /map="4.4.8"
     mRNA            join(<364. .786,896. .>3055)
                     /gene="mkcA"
     exon            <364. .786
                     /gene="mkcA"
                     /number=1
     gene            364. .3055
                     /gene="mkcA"
     CDS             join(364. .786,896. .3055)
                     /gene="mkcA"
                     /note="putative protein kinase"
                     .
```



```
                     /codon_start=1
                     /product="MkcA"
                     /db_xref="PID:g1405366"
                     /translation="MDNNSTLNNITITVASSHNNTTTNSNNNNNNNNGSNSNNNNEQL
                     NTLSNSITSLKSLNISGSGVITNNSEVNTPSEGNITPLPLLSLSDALSPLRELYSCLY
                     VGDIRALMSNQYLKFEENEKIRILGFPEPHLEIDEKTQKYFSYQIIEFESDEDILSIA
                     KSFEQYSQLIIGRTPTQPVVIQECCPGLVLASIYFIKYELFSLPQAHDLIGYNDFFGT
                     SKKHSHFEKVDSCTLINDITSIYYNNMSSKRKTTIFSNNISFTNLQTTTPPTTTTTTT
                     TTTTTSSSNNSEQIESTSKNNSSGITSPTSIKPKNKVGLLKKLFNKMDKESSNNNNNK
                     EKVEKIEKKESSKKERSANDAKEEKERERERKEHQLAKNIPPAELYKHHMEQIQKIQ
                     KQQIAELNATSPDGGNDLSNGSNGSKNGNNNNNNNSNNNNNNNSNEPNSQHQNNNP
                     ITNLMDQELIENYDDGSDRMKNSKTIISPNFESLPDEYKTMVKTWDIKQESVDEHLEI
                     ILSLLHFHTKQRFYSSDPTNLPDKPKYYNVGNVCGNCTPPMILDPAGSGQMVVAKSFY
                     KRPFDQICKKPTNEMKKLYTFREKVGKGGFGTVYLVRNNMDKSKSRIAIKKLPHVRKK
                     EKKFNVKEIRVLEPTNHPNIITYYNSHMLHDEVWIAMEFMEGGTLAEASNQYPFQESN
                     IAYVAREVLQGLAYLHGIQLVHRDLKSQNIMMTTSGEIKLIDFGLCANLSKGERIRMC
                     GSPIWMPPEMIQQKQHGYTCDIWSTAICLLELANRNNKLRKNPIKTMFMVGSEGIKDP
                     FEDPHKWSDEFHDIINKCLQMDPHKRPSAQELLKHPFIQLADNKKKMGKILSSIFLKS
                     IVGI"
     intron          787. .895
                     /gene="mkcA"
                     /number=1
     exon            896. .>3055
                     /gene="mkcA"
                     /number=2
BASE COUNT     1347 a    354 c    392 g    992 t
ORIGIN

  Query Match           11.6%;  Score 17;  DB 18;  Length 3085;
  Best Local Similarity 100.0%;  Pred. No. 1.71e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    114 ATTATTATTTAATATA 130
          ||||||||||||||||
Cp    140 ATTATTATTTAATATA 124


RESULT   49
LOCUS       HUMDIHM      3110 bp    DNA             PRI       04-JUL-1993
DEFINITION  Human gene for myeloperoxidase, upstream region containing
            undermethylated site.
ACCESSION   D14998
NID         g285913
KEYWORDS    Alu sequence; DNase I hypersensitivity; myeloperoxidase;
            undermethylation.
SOURCE      Homo sapiens (library: Charon 4 library of T. maniatis) myeloid
            leukemia HL-60 cells DNA, clone lambdaMPO-18.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
```



```
REFERENCE   1  (bases 1 to 3110)
  AUTHORS   Hashinaka,K. and Yamada,M.
  TITLE     Undermethylation and DNase I hypersensitivity of myeloperoxidase
            gene in HL-60 cells before and after differentiation
  JOURNAL   Arch. Biochem. Biophys. 293 (1), 40-45 (1992)
  MEDLINE   92117699
REFERENCE   2  (bases 1 to 3110)
  AUTHORS   Hashinaka,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-APR-1993) to the DDBJ/EMBL/GenBank databases. Kazuya
            Hashinaka, Medical College of Miyazaki, Department of Biochemistry;
            5200 Kihara, Kiyotake-cho, Miyazaki-gun, Miyazaki 889-16, Japan
            (Tel:0985-85-0985, Fax:0985-85-2401)
COMMENT     Submitted (16-APR-1993) to DDBJ by:
            Kazuya Hashinaka
            Department of Biochemistry
            Medical College of Miyazaki
            5200 Kihara, Kiyotake-cho
            Miyazaki-gun
            Miyazaki 889-16
            Japan
            Phone:  0985-85-0985
            Fax:    0985-85-2401.
FEATURES             Location/Qualifiers
     source          1. .3110
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="HL-60 cells"
                     /cell_type="myeloid leukemia"
                     /clone_lib="charon 4 library of T. maniatis"
     repeat_unit     8. .301
                     /rpt_family="Alu"
     misc_feature    1398
                     /note="undermethylated cytosine of a CpG in HL-60 cells"
     misc_feature    3110
                     /note="-1819 bp position from the transcription start site
                     of myeloperoxidasegene"
BASE COUNT      931 a    662 c    698 g    819 t
ORIGIN

  Query Match           11.6%;  Score 17;  DB 26;  Length 3110;
  Best Local Similarity 100.0%;  Pred. No. 1.71e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    459 ATTTATGGAGAAAAACC 475
          |||||||||||||||||
Qy     24 ATTTATGGAGAAAAACC 40


RESULT   50
LOCUS       HSU27333     3127 bp    mRNA            PRI       31-AUG-1995
```



```
DEFINITION  Human alpha (1,3) fucosyltransferase (FUT6) mRNA, major transcript
            I, complete cds.
ACCESSION   U27333
NID         g967202
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3127)
  AUTHORS   Cameron,H.S., Szczepaniak,D. and Weston,B.W.
  TITLE     Expression of human chromosome 19p alpha(1,3)-fucosyltransferase
            genes in normal tissues. Alternative splicing, polyadenylation, and
            isoforms
  JOURNAL   J. Biol. Chem. 270 (34), 20112-20122 (1995)
  MEDLINE   95378269
REFERENCE   2  (bases 1 to 3127)
  AUTHORS   Weston,B.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAY-1995) Brent W. Weston, Pediatrics, University of
            North Carolina at Chapel Hill, Division of Hematology/Oncology, 418
            MacNider Building, Chapel Hill, NC 27599-7220, USA
FEATURES             Location/Qualifiers
     source          1. .3127
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="19p"
                     /chromosome="19"
                     /clone="major transcript I"
                     /tissue_type="kidney, liver"
     gene            1154. .2233
                     /gene="FUT6"
     CDS             1154. .2233
                     /gene="FUT6"
                     /codon_start=1
                     /product="alpha (1,3) fucosyltransferase"
                     /db_xref="PID:g967203"
                     /translation="MDPLGPARPQWSWRCCLTTLLPQLLMAVCPFSYLRVSQDDPTVY
                     PNGSRFPDSTGTPAHSIPLILLWTWPFNKPIALPRCSEMVPGTADCNITADRKVYPQA
                     DAVIVHHREVMYNPSAQLPRSPRRQGQRWIWFSMESPSHCWQLKAMDGYFNLTMSYRS
                     DSDIFTPYGWLEPWSGQPAHPPLNLSAKTELVAWAVSNWGPNSARVRYYQSLQAHLKV
                     DVYGRSHKPLPQGTMMETLSRYKFYLAPENSLHPDYITEKLWRNALEAWAVPVVLGPS
                     RSNYERFLPPDAFIHVDDFQSPKDLARYLQELDKDHARYLSYFRWRETLRPRSFSWAL
                     APCKACWKLQEESRYQTRGIAAWFT"
BASE COUNT      619 a    954 c    870 g    684 t
ORIGIN

  Query Match           11.6%;  Score 17;  DB 27;  Length 3127;
  Best Local Similarity 100.0%;  Pred. No. 1.71e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```

```
Db    2286 CCTGAGTGGGGGCCTCA 2302
           |||||||||||||||||
Cp      79 CCTGAGTGGGGGCCTCA 63
```

Search completed: Tue Dec 22 19:56:56 1998
Job time : 504 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Tue Dec 22 20:18:53 1998; MasPar time 433.90 Seconds
1254.664 Million cell updates/sec

Tabular output not generated.

Title: >US-08-530-112A-1017
Description: (1-304) from US08530112A.seq
Perfect Score: 298
N.A. Sequence: 1 GATCTAAAGCAGGTNGTGTT.....ATATGTATTTTNGGGATAAA 304
Comp: CTAGATTTCGTCCANCACAA.....TATACATAAAANCCCTATTT

Scoring table: TABLE jmetric
Gap 60

Nmatch STD : Dbase 0; Query 0

Searched: 2275026 seqs, 895388244 bases x 2

Post-processing: Minimum Match 0%
Listing first 1000 summaries

Database: embl-est55
1:em_est1 2:em_gss1 3:em_gss2 4:em_gss3
Database: genbank-est107
5:gb_est1 6:gb_est10 7:gb_est11 8:gb_est12 9:gb_est13
10:gb_est14 11:gb_est15 12:gb_est16 13:gb_est17
14:gb_est18 15:gb_est19 16:gb_est2 17:gb_est20
18:gb_est21 19:gb_est3 20:gb_est4 21:gb_est5 22:gb_est6
23:gb_est7 24:gb_est8 25:gb_est9 26:gb_gss1 27:gb_gss2
28:gb_gss3 29:gb_gss4

Statistics: Mean 8.369; Variance 1.264; scale 6.621

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
	1	298	100.0	304	7	HUMGS01155	Human HL60 3'directed	0.00e+00
c	2	240	80.5	387	7	AA463716	aa07d11.s1 Soares NhHM	0.00e+00
c	3	240	80.5	421	13	AA609872	af08c06.s1 Soares test	0.00e+00
c	4	212	71.1	377	8	AA496258	aa24b10.s1 NCI_CGAP_GC	0.00e+00
c	5	181	60.7	474	12	AA760717	ns13d07.s1 NCI_CGAP_GC	0.00e+00
c	6	42	14.1	443	10	SSC11209	Sus scrofa mRNA; expre	7.84e-54
	7	27	9.1	354	24	AA273441	vc01c11.r1 Soares mous	4.44e-21
c	8	27	9.1	430	14	AA955078	UI-R-A1-ey-h-11-0-UI.s	4.44e-21
c	9	27	9.1	449	15	AI008917	EST203368 Normalized r	4.44e-21
c	10	23	7.7	363	12	AA730318	nw41f01.s1 NCI_CGAP_Ew	2.48e-13
	11	22	7.4	250	28	AQ106161	HS_3070_A2_C07_MR CIT	1.70e-11
	12	22	7.4	323	19	HSC1MG012	H. sapiens partial cDN	1.70e-11
c	13	22	7.4	418	13	AA856996	oh94b08.s1 NCI_CGAP_HN	1.70e-11
c	14	22	7.4	425	26	B45656	HS-1062-A1-G11-MF.abi	1.70e-11
c	15	22	7.4	484	6	AA043696	ak50d10.r1 Soares preg	1.70e-11
	16	22	7.4	528	16	R66601	yi32d10.r1 Homo sapien	1.70e-11
	17	22	7.4	532	12	AA744248	ny62g03.s1 NCI_CGAP_GC	1.70e-11
	18	22	7.4	562	12	AA747449	nx88h04.s1 NCI_CGAP_GC	1.70e-11
c	19	21	7.0	355	18	AI147069	ok33b12.s1 Soares_NSF_	1.04e-09
c	20	20	6.7	125	14	AA925765	UI-R-A1-ep-g-12-0-UI.s	5.68e-08
c	21	20	6.7	134	18	AI102239	EST211528 Normalized r	5.68e-08
c	22	20	6.7	239	15	AA943469	EST198968 Normalized r	5.68e-08
	23	20	6.7	272	7	AA522005	wh78a09.r1 Knowles Sol	5.68e-08
c	24	20	6.7	272	15	AI011490	EST205941 Normalized r	5.68e-08
	25	20	6.7	293	21	W18871	mc04d01.r1 Soares mous	5.68e-08
	26	20	6.7	312	17	AI072793	UI-R-T0-lw-b-06-0-UI.s	5.68e-08
	27	20	6.7	330	5	R09708	yf23e10.s1 Homo sapien	5.68e-08
c	28	20	6.7	331	6	AA293650	zt55d12.r1 Soares ovar	5.68e-08
c	29	20	6.7	339	15	AA934050	on78a12.s1 Soares_NFL_	5.68e-08
c	30	20	6.7	354	15	AI060034	UI-R-C1-kz-a-11-0-UI.s	5.68e-08
	31	20	6.7	370	28	AQ095488	HS_3012_B1_B10_MR CIT	5.68e-08
	32	20	6.7	397	11	AA672177	vk90a10.s1 Knowles Sol	5.68e-08
c	33	20	6.7	403	28	AQ211566	HS_3235_B2_F01_MR CIT	5.68e-08
c	34	20	6.7	406	14	H31634	EST105880 Rat PC-12 ce	5.68e-08
c	35	20	6.7	416	9	AA590910	vm21e04.r1 Knowles Sol	5.68e-08
	36	20	6.7	417	18	AI052454	os07g11.s1 Soares_feta	5.68e-08
c	37	20	6.7	424	29	AQ148379	HS_2211_A2_A12_MF CIT	5.68e-08
c	38	20	6.7	445	18	AI101287	EST210576 Normalized r	5.68e-08
c	39	20	6.7	447	18	AI173314	uc29c11.r1 Soares mous	5.68e-08
	40	20	6.7	449	8	AA438117	vd22f02.s1 Knowles Sol	5.68e-08
c	41	20	6.7	455	28	AQ222916	HS_2013_A1_E01_T7 CIT	5.68e-08
	42	20	6.7	467	26	B74942	RPCI11-11P8.TP RPCI11	5.68e-08

c	43	20	6.7	499	17	AI070519	UI-R-Y0-lv-f-07-0-UI.s	5.68e-08
c	44	20	6.7	510	15	AI007920	EST202371 Normalized r	5.68e-08
c	45	20	6.7	522	28	AQ184311	HS_3226_A1_D12_T7 CIT	5.68e-08
	46	20	6.7	523	29	AQ141570	HS_3154_A1_G08_T7 CIT	5.68e-08
	47	20	6.7	544	28	AQ122664	HS_3086_A2_H02_MF CIT	5.68e-08
c	48	20	6.7	587	28	AQ197092	CIT-HSP-23B4I15.TF CIT	5.68e-08
	49	20	6.7	592	17	AU024235	Mus musculus unfertili	5.68e-08
c	50	20	6.7	606	28	AQ195348	RPCI11-61L17.TJ RPCI11	5.68e-08
	51	20	6.7	624	28	AQ201089	RPCI11-45C8.TJ RPCI11	5.68e-08
	52	20	6.7	650	8	AA549979	1058m3 gmbPfHB3.1, G.	5.68e-08
c	53	19	6.4	125	6	C00200	HUMGS0005859, Human Ge	2.71e-06
c	54	19	6.4	221	8	AA496926	ae32c03.s1 Gessler Wil	2.71e-06
	55	19	6.4	234	5	T95507	ye43g11.s1 Homo sapien	2.71e-06
	56	19	6.4	237	9	W35574	TgESTay90g11.r1 TgRH T	2.71e-06
c	57	19	6.4	237	5	T18147	0722c3 Plasmodium falc	2.71e-06
	58	19	6.4	239	5	T93864	ye06a03.s1 Homo sapien	2.71e-06
c	59	19	6.4	246	15	AU008113	Schizosaccharomyces po	2.71e-06
c	60	19	6.4	314	5	R39664	yh96g01.r1 Homo sapien	2.71e-06
c	61	19	6.4	327	28	AQ094348	HS_3025_B2_F10_MR CIT	2.71e-06
c	62	19	6.4	344	6	AA437921	ve32a06.r1 Ko mouse em	2.71e-06
	63	19	6.4	357	13	AA843511	ak08a02.s1 Soares para	2.71e-06
	64	19	6.4	377	12	AA809847	oa92a05.s1 NCI_CGAP_GC	2.71e-06
	65	19	6.4	388	20	N90246	zb35e08.s1 Soares para	2.71e-06
c	66	19	6.4	396	5	R26965	yh51f06.r1 Homo sapien	2.71e-06
	67	19	6.4	397	26	B40172	HS-1051-A1-D02-MR.abi	2.71e-06
	68	19	6.4	400	21	W16661	zb94g10.s1 Soares para	2.71e-06
	69	19	6.4	413	28	AQ187931	HS_3138_B2_F06_MR CIT	2.71e-06
	70	19	6.4	428	12	AA788808	ah31d11.s1 Soares para	2.71e-06
	71	19	6.4	437	23	AA140376	mq03f07.r1 Soares 2NbM	2.71e-06
	72	19	6.4	437	10	AA158380	zo66a07.s1 Stratagene	2.71e-06
	73	19	6.4	448	22	N57632	yv54a01.s1 Soares feta	2.71e-06
c	74	19	6.4	449	8	AA438117	vd22f02.s1 Knowles Sol	2.71e-06
c	75	19	6.4	456	16	R93469	yq16e03.r1 Homo sapien	2.71e-06
	76	19	6.4	463	17	AI017527	ou35a07.x1 Soares_NFL_	2.71e-06
	77	19	6.4	467	15	AI003571	ai88c10.s1 Soares_para	2.71e-06
	78	19	6.4	468	12	AA776429	zj50g06.s1 Soares feta	2.71e-06
c	79	19	6.4	468	12	AA665208	nu82c03.s1 NCI_CGAP_Al	2.71e-06
	80	19	6.4	469	19	N51597	yy98a04.s1 Homo sapien	2.71e-06
	81	19	6.4	478	12	AA772015	ai39e07.s1 Soares para	2.71e-06
c	82	19	6.4	481	12	AA755765	vv35a07.r1 Stratagene	2.71e-06
c	83	19	6.4	498	21	W30961	zc65h12.r1 Soares feta	2.71e-06
	84	19	6.4	503	28	AQ214392	HS_3012_B1_A02_T7 CIT	2.71e-06
c	85	19	6.4	503	22	CHESTM054	C.hircus expressed seq	2.71e-06
	86	19	6.4	504	6	AA411048	zv40c09.s1 Soares ovar	2.71e-06
c	87	19	6.4	506	10	AA284307	zc65h12.T7 Soares feta	2.71e-06
	88	19	6.4	507	11	AA682887	zj16h05.s1 Soares feta	2.71e-06
	89	19	6.4	512	14	AA490163	ab06a07.s1 Stratagene	2.71e-06
	90	19	6.4	516	5	T73089	yc69h05.s1 Homo sapien	2.71e-06
	91	19	6.4	519	13	AA843685	ak01d03.s1 Soares para	2.71e-06
	92	19	6.4	521	12	AA723620	ah24h08.s1 Soares para	2.71e-06
	93	19	6.4	527	12	AA777651	zi95d07.s1 Soares feta	2.71e-06

	94	19	6.4	529	12	AA723562	ah24b08.s1 Soares para	2.71e-06
c	95	19	6.4	550	20	W39186	zb35e08.r1 Soares para	2.71e-06
c	96	19	6.4	553	29	AQ246642	HS_2062_B2_C08_T7 CIT	2.71e-06
	97	19	6.4	572	15	AI057542	oy30g06.x1 Soares_par	2.71e-06
	98	19	6.4	578	12	AA788809	ah31d12.s1 Soares para	2.71e-06
	99	19	6.4	590	17	AU024347	Mus musculus unfertili	2.71e-06
	100	19	6.4	619	27	FR0031347	Fugu rubripes GSS sequ	2.71e-06
	101	19	6.4	653	18	AI091085	qa53e12.s1 Soares_NhHM	2.71e-06
	102	19	6.4	713	13	AA843803	ak09b09.s1 Soares para	2.71e-06
	103	19	6.4	733	13	AA843802	ak09b08.s1 Soares para	2.71e-06
	104	18	6.0	120	7	HUMGS02889	Human adult lung 3'dir	1.12e-04
c	105	18	6.0	199	5	T36493	EST101438 Saccharomyce	1.12e-04
c	106	18	6.0	258	19	N47669	yy54h03.r1 Homo sapien	1.12e-04
	107	18	6.0	264	16	H13854	yj08f05.s1 Homo sapien	1.12e-04
c	108	18	6.0	289	14	H33292	EST109141 Rat PC-12 ce	1.12e-04
c	109	18	6.0	292	25	AA371481	EST83269 Prostate glan	1.12e-04
c	110	18	6.0	299	22	AA129825	zl13b12.s1 Soares preg	1.12e-04
c	111	18	6.0	299	15	AA893371	EST197174 Normalized r	1.12e-04
c	112	18	6.0	305	17	AI086226	ow90f01.s1 Soares_feta	1.12e-04
c	113	18	6.0	319	15	AI030046	UI-R-C0-ja-h-04-0-UI.s	1.12e-04
	114	18	6.0	320	24	AA299065	EST111551 Uterus Homo s	1.12e-04
	115	18	6.0	323	14	AA888411	nw74c04.s1 NCI_CGAP_Pr	1.12e-04
	116	18	6.0	324	5	T73070	yc69f05.s1 Homo sapien	1.12e-04
c	117	18	6.0	327	27	B93334	CIT-HSP-2171G8.TP CIT-	1.12e-04
c	118	18	6.0	330	15	AA910788	o148c02.s1 Soares_NFL_	1.12e-04
	119	18	6.0	335	28	AQ098209	HS_3048_A1_C03_MR CIT	1.12e-04
	120	18	6.0	336	10	AA618275	nq17c04.s1 NCI_CGAP_Th	1.12e-04
	121	18	6.0	336	13	AA885025	am41d01.s1 Soares NFL	1.12e-04
c	122	18	6.0	346	13	AA842970	T4988 MVAT4 bloodstrea	1.12e-04
	123	18	6.0	358	29	AQ228739	HS_2020_A2_E12_T7 CIT	1.12e-04
	124	18	6.0	365	26	B72048	RPCI11-7H11.TP RPCI11	1.12e-04
c	125	18	6.0	366	11	AA713427	T4579 MVAT4 bloodstrea	1.12e-04
	126	18	6.0	383	26	B41119	HS-1053-A2-A04-MR.abi	1.12e-04
c	127	18	6.0	383	27	AQ058376	CIT-HSP-2339H14.TR CIT	1.12e-04
c	128	18	6.0	391	11	AA709634	vv41d07.r1 Soares 2NbM	1.12e-04
c	129	18	6.0	403	26	B38589	HS-1047-B2-G11-MF.abi	1.12e-04
	130	18	6.0	406	28	AQ191717	HS_3228_A2_A04_MR CIT	1.12e-04
	131	18	6.0	406	26	FR0019306	F.rubripes GSS sequenc	1.12e-04
c	132	18	6.0	406	14	AA990430	ua55g11.s1 Soares 2NbM	1.12e-04
	133	18	6.0	407	26	B39676	HS-1050-A1-F12-MF.abi	1.12e-04
	134	18	6.0	412	19	H57696	yr27e10.r1 Homo sapien	1.12e-04
c	135	18	6.0	415	8	AA262614	za23a04.s1 NCI_CGAP_GC	1.12e-04
c	136	18	6.0	415	23	AA201075	mu09f02.r1 Soares 2NbM	1.12e-04
c	137	18	6.0	419	6	AA058734	zk70e06.s1 Soares preg	1.12e-04
c	138	18	6.0	419	18	AI104883	EST214172 Normalized r	1.12e-04
	139	18	6.0	424	28	AQ221443	HS_2004_B2_E04_T7 CIT	1.12e-04
	140	18	6.0	425	29	AQ242852	HS_2041_A2_D03_T7 CIT	1.12e-04
	141	18	6.0	426	27	AQ090186	HS_3009_A1_D08_T7 CIT	1.12e-04
c	142	18	6.0	426	8	AA251831	zs09d02.s1 NCI_CGAP_GC	1.12e-04
c	143	18	6.0	428	18	AI078298	oz12a05.x1 Soares_feta	1.12e-04
	144	18	6.0	438	22	N98022	2068C3 czapPFDd2.1, De	1.12e-04

	145	18	6.0	440	18	AI076043	oy60c07.x1 NCI_CGAP_Br	1.12e-04
c	146	18	6.0	441	28	AQ118020	HS_2171_A2_D04_MF CIT	1.12e-04
	147	18	6.0	442	27	AQ038257	CIT-HSP-2320E10.TR CIT	1.12e-04
c	148	18	6.0	442	8	AA542221	vk28a01.r1 Soares mous	1.12e-04
c	149	18	6.0	445	27	AQ063955	HS_2183_B2_G01_MR CIT	1.12e-04
c	150	18	6.0	452	10	AA424015	zv80b03.s1 Soares tota	1.12e-04
	151	18	6.0	455	7	AA452709	zx39c07.r1 Soares tota	1.12e-04
c	152	18	6.0	456	22	AA148889	zl27f11.s1 Soares preg	1.12e-04
c	153	18	6.0	468	26	B49798	CIT-HSP-421P4.TP CIT-H	1.12e-04
	154	18	6.0	477	26	AQ022505	HS_2177_B2_G10_MR CIT	1.12e-04
c	155	18	6.0	477	11	AA673486	vp49f04.r1 Knowles Sol	1.12e-04
	156	18	6.0	479	18	AI150898	qb48f11.x1 NCI_CGAP_Br	1.12e-04
c	157	18	6.0	490	18	AI153583	vz89a09.r1 Soares 2NbM	1.12e-04
c	158	18	6.0	491	7	AA461496	zx60a08.s1 Soares test	1.12e-04
c	159	18	6.0	492	15	AA978122	op32d09.s1 Soares_NFL_	1.12e-04
	160	18	6.0	495	28	AQ154754	HS_3027_A2_D03_MF CIT	1.12e-04
c	161	18	6.0	496	7	AA472124	vh10d05.r1 Soares mous	1.12e-04
c	162	18	6.0	500	8	AA549844	0917m3 gmbPfHB3.1, G.	1.12e-04
	163	18	6.0	505	21	W24228	zb82d09.r1 Soares sene	1.12e-04
c	164	18	6.0	506	18	AI178576	EST222255 Normalized r	1.12e-04
	165	18	6.0	508	29	AQ138233	HS_3074_B1_H10_MF CIT	1.12e-04
	166	18	6.0	511	28	AQ168093	HS_2233_A1_D08_MF CIT	1.12e-04
c	167	18	6.0	518	13	AA633695	ag87a07.s1 Stratagene	1.12e-04
c	168	18	6.0	522	6	AA404507	zw38c03.s1 Soares tota	1.12e-04
c	169	18	6.0	524	10	AA651868	ns38g03.s1 NCI_CGAP_GC	1.12e-04
	170	18	6.0	526	8	AA550100	1199m3 gmbPfHB3.1, G.	1.12e-04
c	171	18	6.0	530	13	AA868266	ak40a03.s1 Soares test	1.12e-04
c	172	18	6.0	535	11	AA669033	ab88g09.s1 Stratagene	1.12e-04
c	173	18	6.0	554	13	AA836101	od15b07.s1 NCI_CGAP_GC	1.12e-04
c	174	18	6.0	591	27	FR0033587	Fugu rubripes GSS sequ	1.12e-04
	175	18	6.0	594	28	AQ173052	HS_3200_B2_H02_T7 CIT	1.12e-04
	176	18	6.0	600	26	FR0018567	F.rubripes GSS sequenc	1.12e-04
c	177	18	6.0	605	12	AA779557	af24g09.s1 Soares tota	1.12e-04
	178	18	6.0	607	27	B86613	RPCI11-24F15.TV RPCI11	1.12e-04
c	179	18	6.0	652	28	AQ115298	RPCI11-56O7.TK.1 RPCI1	1.12e-04
c	180	18	6.0	658	8	AA550546	1709m3 gmbPfHB3.1, G.	1.12e-04
c	181	18	6.0	687	27	AQ005064	CIT-HSP-2294F4.TR CIT-	1.12e-04
c	182	18	6.0	703	27	AQ034064	l(3)04708 Drosophila m	1.12e-04
c	183	17	5.7	108	26	B38293	HS-1047-A2-F07-MR.abi	3.93e-03
	184	17	5.7	141	16	R40146	yf69b10.s1 Homo sapien	3.93e-03
	185	17	5.7	152	15	AI030400	UI-R-C0-jc-c-12-0-UI.s	3.93e-03
	186	17	5.7	164	8	AA562665	v156f11.r1 Stratagene	3.93e-03
	187	17	5.7	167	21	AA038577	mi92c07.r1 Soares mous	3.93e-03
	188	17	5.7	200	5	R43876	yg26c07.s1 Homo sapien	3.93e-03
	189	17	5.7	208	7	AA458063	vf76f05.r1 Soares mous	3.93e-03
	190	17	5.7	219	25	AA332050	EST35944 Embryo, 8 wee	3.93e-03
c	191	17	5.7	229	20	HSC0CG061	H. sapiens partial cDN	3.93e-03
	192	17	5.7	252	23	AA220270	mv67e07.r1 Soares mous	3.93e-03
	193	17	5.7	256	15	C22161	Citrus cDNA, partial s	3.93e-03
	194	17	5.7	258	19	N47669	yy54h03.r1 Homo sapien	3.93e-03
c	195	17	5.7	260	9	N82879	TgESTzy35f03.r1 TgRH T	3.93e-03

	196	17	5.7	265	11	T43770	7033 Lambda-PRL2 Arabi	3.93e-03
	197	17	5.7	265	20	T15357	IB1007 Infant brain, B	3.93e-03
	198	17	5.7	270	8	AA559768	MC210210.NA3 TBRI-Sabe	3.93e-03
	199	17	5.7	273	7	MUSGS01447	Mouse 3'-directed cDN	3.93e-03
	200	17	5.7	286	9	AA585324	TH001 HTCDLi Homo sapi	3.93e-03
	201	17	5.7	291	12	T76376	11154 Lambda-PRL2 Arab	3.93e-03
	202	17	5.7	292	11	AA720155	33348 Lambda-PRL2 Arab	3.93e-03
c	203	17	5.7	294	27	B70296	CIT-HSP-2028M18.TF CIT	3.93e-03
c	204	17	5.7	300	9	C35136	C.elegans cDNA clone y	3.93e-03
c	205	17	5.7	305	27	AQ069928	HS_3011_A2_D04_MR CIT	3.93e-03
c	206	17	5.7	308	8	AA253460	zs06f04.s1 NCI_CGAP_GC	3.93e-03
c	207	17	5.7	315	24	AA297045	EST112596 Cerebellum I	3.93e-03
	208	17	5.7	316	6	CELK066F4R	C.elegans cDNA clone y	3.93e-03
	209	17	5.7	320	25	AA377427	EST90003 Synovial memb	3.93e-03
c	210	17	5.7	320	18	AI171406	EST217365 Normalized r	3.93e-03
c	211	17	5.7	324	14	AA965283	UI-R-C0-hd-d-07-0-UI.s	3.93e-03
	212	17	5.7	332	12	T76322	11100 Lambda-PRL2 Arab	3.93e-03
c	213	17	5.7	332	5	T80409	yd18d09.r1 Homo sapien	3.93e-03
c	214	17	5.7	337	5	T09944	0492m3 Plasmodium falc	3.93e-03
c	215	17	5.7	337	27	B82192	RPCI11-27P3.TP RPCI11	3.93e-03
	216	17	5.7	342	11	T42836	6099 Lambda-PRL2 Arabi	3.93e-03
	217	17	5.7	345	9	C74073	Rice cDNA, partial seq	3.93e-03
	218	17	5.7	347	23	T03428	IB272 Infant brain, Be	3.93e-03
c	219	17	5.7	347	23	AA261087	mz65c10.r1 Soares mous	3.93e-03
c	220	17	5.7	347	12	AA142240	CK00016.Sprime CK Dros	3.93e-03
	221	17	5.7	351	26	B98343	T24A19TRB TAMU Arabido	3.93e-03
c	222	17	5.7	354	5	T02799	0137M7 Plasmodium falc	3.93e-03
	223	17	5.7	357	8	AA281075	za98b05.s1 NCI_CGAP_GC	3.93e-03
	224	17	5.7	358	17	AA407274	EST02271 Mouse 7.5 dpc	3.93e-03
c	225	17	5.7	358	6	AA134596	zl20h11.r1 Soares preg	3.93e-03
c	226	17	5.7	359	24	AA288034	vb12h12.r1 Soares mous	3.93e-03
	227	17	5.7	360	11	AA677716	zi12a12.s1 Soares feta	3.93e-03
	228	17	5.7	360	27	AQ055074	CIT-HSP-2342D7.TR CIT-	3.93e-03
c	229	17	5.7	360	6	C13449	C.elegans cDNA clone y	3.93e-03
	230	17	5.7	361	9	AA578195	nl29h03.s1 NCI_CGAP_HS	3.93e-03
	231	17	5.7	365	17	T45865	9128 Lambda-PRL2 Arabi	3.93e-03
c	232	17	5.7	367	16	R59691	yh02h08.r1 Homo sapien	3.93e-03
	233	17	5.7	371	11	T13688	1853 Lambda-PRL2 Arabi	3.93e-03
	234	17	5.7	378	28	AQ164039	HS_2270_B1_H11_MF CIT	3.93e-03
c	235	17	5.7	378	18	AI170591	EST216525 Normalized r	3.93e-03
c	236	17	5.7	378	14	AA850668	EST193436 Normalized r	3.93e-03
c	237	17	5.7	380	29	AQ150734	HS_3091_A2_D12_MR CIT	3.93e-03
c	238	17	5.7	383	6	AA386614	vb53g07.r1 Ko mouse em	3.93e-03
c	239	17	5.7	388	27	AQ072166	HS_3025_B1_C05_T7 CIT	3.93e-03
c	240	17	5.7	391	6	AA390153	mv35f11.r1 GuayWoodfor	3.93e-03
	241	17	5.7	394	10	T22116	4124 Lambda-PRL2 Arabi	3.93e-03
c	242	17	5.7	394	27	AQ072374	HS_3025_B1_D09_T7 CIT	3.93e-03
c	243	17	5.7	395	24	AA267002	mz94g02.r1 Soares mous	3.93e-03
c	244	17	5.7	397	5	T86073	yd76e05.r1 Homo sapien	3.93e-03
	245	17	5.7	398	14	AA965248	UI-R-C0-hd-a-06-0-UI.s	3.93e-03
	246	17	5.7	399	24	N45801	T1485 MVAT4 bloodstrea	3.93e-03

c	247	17	5.7	399	19	N73717	yg79h08.s1 Homo sapien	3.93e-03
c	248	17	5.7	400	29	AQ239054	RPCI11-74N5.TK RPCI11	3.93e-03
	249	17	5.7	401	29	AQ148232	HS_3116_B2_H12_MR CIT	3.93e-03
c	250	17	5.7	402	28	AQ209012	HS_3229_B2_C04_MR CIT	3.93e-03
	251	17	5.7	404	12	AA794344	vr46a05.s1 Knowles Sol	3.93e-03
	252	17	5.7	408	10	T22813	4821 Lambda-PRL2 Arabi	3.93e-03
c	253	17	5.7	408	23	AA204487	mu12h10.r1 Soares 2NbM	3.93e-03
	254	17	5.7	409	26	AQ009933	HS_2264_B1_C07_MF CIT	3.93e-03
	255	17	5.7	409	11	AA677880	zi13f04.s1 Soares feta	3.93e-03
c	256	17	5.7	411	5	T79347	yd74d06.r1 Homo sapien	3.93e-03
c	257	17	5.7	415	28	AQ176169	HS_3211_B1_H04_MR CIT	3.93e-03
	258	17	5.7	419	12	T88023	11719 Lambda-PRL2 Arab	3.93e-03
c	259	17	5.7	422	12	AA141136	CK01282.3prime CK Dros	3.93e-03
	260	17	5.7	423	27	AQ081503	RPCI11-52N18.TK RPCI11	3.93e-03
	261	17	5.7	423	23	AA217590	mu98g01.r1 Soares mous	3.93e-03
	262	17	5.7	425	26	B39857	HS-1050-B1-E03-MF.abi	3.93e-03
	263	17	5.7	427	10	AA415338	vc52f12.r1 Knowles Sol	3.93e-03
	264	17	5.7	428	28	AQ168908	HS_3174_A1_B08_MR CIT	3.93e-03
	265	17	5.7	430	5	R51446	yg72h06.r1 Homo sapien	3.93e-03
	266	17	5.7	431	14	AA963220	UI-R-E1-gj-b-10-0-UI.s	3.93e-03
	267	17	5.7	431	28	AQ176276	HS_3209_A1_B10_MR CIT	3.93e-03
c	268	17	5.7	431	12	AA739087	vv60g11.r1 Soares 2NbM	3.93e-03
	269	17	5.7	434	11	N65388	20428 Lambda-PRL2 Arab	3.93e-03
	270	17	5.7	435	8	AA542580	fa08b11.s1 Zebrafish I	3.93e-03
c	271	17	5.7	435	21	W69995	zd51a11.s1 Soares feta	3.93e-03
	272	17	5.7	437	11	T44354	7617 Lambda-PRL2 Arabi	3.93e-03
	273	17	5.7	440	18	AI141643	oy91h02.x1 Soares_feta	3.93e-03
c	274	17	5.7	446	15	AI028853	UI-R-C0-ij-f-03-0-UI.s	3.93e-03
	275	17	5.7	447	6	AA396379	vb27b03.r1 Soares mous	3.93e-03
	276	17	5.7	448	10	AA611621	vo81e08.r1 Barstead mo	3.93e-03
c	277	17	5.7	448	15	AI049104	ub35b05.r1 Soares 2NbM	3.93e-03
	278	17	5.7	449	28	AQ115812	RPCI11-52J13.TK RPCI11	3.93e-03
	279	17	5.7	449	29	AQ230059	HS_2025_A1_E09_T7 CIT	3.93e-03
c	280	17	5.7	451	15	AI059423	UI-R-C1-kw-c-09-0-UI.s	3.93e-03
c	281	17	5.7	461	23	AA124544	mp81a07.r1 Soares 2NbM	3.93e-03
c	282	17	5.7	462	15	AI020673	ua96g04.r1 Soares mous	3.93e-03
c	283	17	5.7	467	23	AA124609	mp86h02.r1 Soares 2NbM	3.93e-03
	284	17	5.7	469	28	AQ164256	HS_2207_B1_G05_MR CIT	3.93e-03
c	285	17	5.7	472	19	N70896	za33a01.s1 Homo sapien	3.93e-03
	286	17	5.7	474	27	AQ001024	CIT-HSP-2283C1.TF CIT-	3.93e-03
	287	17	5.7	475	10	T22061	4069 Lambda-PRL2 Arabi	3.93e-03
	288	17	5.7	476	18	AI155969	ue04f09.r1 Soares mous	3.93e-03
c	289	17	5.7	477	28	AQ126440	HS_3037_B1_F04_MF CIT	3.93e-03
c	290	17	5.7	479	23	AA063420	zf70b06.s1 Soares pine	3.93e-03
c	291	17	5.7	480	12	AA763979	vw09a12.r1 Soares 2NbM	3.93e-03
	292	17	5.7	481	7	AA517750	vh80h08.r1 Knowles Sol	3.93e-03
	293	17	5.7	485	11	N65568	20608 Lambda-PRL2 Arab	3.93e-03
	294	17	5.7	485	16	H10802	ym04a11.r1 Homo sapien	3.93e-03
	295	17	5.7	486	18	AI092853	qa77a08.x1 Soares_feta	3.93e-03
	296	17	5.7	492	11	N96414	20993 CD4-14 Arabidops	3.93e-03
	297	17	5.7	496	28	AQ215755	HS_3252_A2_E04_T7 CIT	3.93e-03

c	298	17	5.7	496	26	B50338	CIT-HSP-335A18.TP CIT-	3.93e-03
c	299	17	5.7	498	26	AQ020306	CIT-HSP-2307A7.TR CIT-	3.93e-03
c	300	17	5.7	501	27	B79255	CIT-HSP-335A18.TP.1 CI	3.93e-03
c	301	17	5.7	503	13	AA827860	od56e11.s1 NCI_CGAP_GC	3.93e-03
c	302	17	5.7	505	13	AA858776	UI-R-AO-bb-g-11-0-UI.s	3.93e-03
	303	17	5.7	508	23	AA174781	ms73e08.r1 Soares mous	3.93e-03
	304	17	5.7	512	15	AI048762	ub32b12.r1 Soares 2NbM	3.93e-03
c	305	17	5.7	512	23	AA195187	zr34f07.r1 Soares NhHM	3.93e-03
	306	17	5.7	513	12	R30159	12764 Lambda-PRL2 Arab	3.93e-03
c	307	17	5.7	513	16	H15989	ym20g04.r1 Homo sapien	3.93e-03
c	308	17	5.7	519	29	AQ142341	HS_3054_A2_F10_MR CIT	3.93e-03
	309	17	5.7	522	12	T88019	11715 Lambda-PRL2 Arab	3.93e-03
c	310	17	5.7	524	15	AI009362	EST203813 Normalized r	3.93e-03
c	311	17	5.7	530	11	W43725	23102 CD4-16 Arabidops	3.93e-03
	312	17	5.7	552	29	AQ232860	HS_2028_B2_E03_MR CIT	3.93e-03
	313	17	5.7	555	27	B81359	CIT-HSP-2025I18.TFB CI	3.93e-03
	314	17	5.7	559	11	R90144	16499 Lambda-PRL2 Arab	3.93e-03
	315	17	5.7	562	10	T21852	3860 Lambda-PRL2 Arabi	3.93e-03
	316	17	5.7	564	24	AA276780	vc42c03.r1 Soares mous	3.93e-03
c	317	17	5.7	570	16	H06325	yl79c08.r1 Homo sapien	3.93e-03
	318	17	5.7	575	26	FR0012919	F.rubripes GSS sequenc	3.93e-03
c	319	17	5.7	575	16	H17764	ym40g01.r1 Homo sapien	3.93e-03
	320	17	5.7	577	28	AQ179403	HS_3170_A1_C02_T7 CIT	3.93e-03
c	321	17	5.7	582	26	AQ010815	F24F13TFB IGF Arabidop	3.93e-03
c	322	17	5.7	582	29	AQ234403	HS_2058_A1_F08_MR CIT	3.93e-03
	323	17	5.7	584	17	AU024379	Mus musculus unfertili	3.93e-03
c	324	17	5.7	584	26	B16943	3430B.TVB CIT978SKA1 H	3.93e-03
c	325	17	5.7	590	26	B17453	345K24.TPB CIT978SKA1	3.93e-03
	326	17	5.7	593	11	H37570	15699 Lambda-PRL2 Arab	3.93e-03
c	327	17	5.7	595	18	AI105441	EST214730 Normalized r	3.93e-03
c	328	17	5.7	600	26	FR0018087	F.rubripes GSS sequenc	3.93e-03
	329	17	5.7	603	17	AI109239	GH08370.5prime GH Dros	3.93e-03
c	330	17	5.7	607	26	B01875	cSRL-142E10-u cSRL flo	3.93e-03
c	331	17	5.7	614	12	AA140669	CK00374.3prime CK Dros	3.93e-03
c	332	17	5.7	616	13	AA196347	zp97a09.r1 Stratagene	3.93e-03
c	333	17	5.7	635	27	AQ002540	CIT-HSP-2283I18.TF CIT	3.93e-03
	334	17	5.7	652	27	B69023	CIT-HSP-2025I17.TF CIT	3.93e-03
c	335	17	5.7	656	26	B96464	F22O18TR IGF Arabidops	3.93e-03
	336	17	5.7	669	15	AI055490	coau0004C06 Cotton Bol	3.93e-03
c	337	17	5.7	737	27	AG012443	Homo sapiens genomic D	3.93e-03
	338	17	5.7	748	8	AA550570	1735m3 gmbPfHB3.1, G.	3.93e-03
	339	17	5.7	758	26	B55044	CIT-HSP-385N14.TR CIT-	3.93e-03
	340	17	5.7	766	27	AG012421	Homo sapiens genomic D	3.93e-03
c	341	17	5.7	893	22	AA210332	mu72e09.r1 Soares mous	3.93e-03
c	342	16	5.4	123	23	AA247919	j7153.seq.F Human feta	1.15e-01
c	343	16	5.4	165	26	HSJ227842	Homo sapiens DNA for t	1.15e-01
c	344	16	5.4	174	14	AA931035	SMTBADA000210SK Schist	1.15e-01
c	345	16	5.4	174	19	H63022	EST000210 Schistosoma	1.15e-01
c	346	16	5.4	175	19	H63023	EST000211 Schistosoma	1.15e-01
c	347	16	5.4	175	14	AA931036	SMTBADA000211SK Schist	1.15e-01
c	348	16	5.4	187	18	AI155016	ud78a03.r1 Soares mous	1.15e-01

	349	16	5.4	192	7	AA457293	aa88g08.r1 Stratagene	1.15e-01
	350	16	5.4	197	5	T38618	EST104130 Saccharomyce	1.15e-01
	351	16	5.4	209	23	AA124012	mq22e10.r1 Barstead MP	1.15e-01
	352	16	5.4	216	10	AA415672	vd26d08.s1 Knowles Sol	1.15e-01
c	353	16	5.4	216	5	R05221	pk42c11.s1 Caenorhabdi	1.15e-01
c	354	16	5.4	216	5	R03614	pk09e03.r1 Caenorhabdi	1.15e-01
	355	16	5.4	228	11	AA728098	LmLv39p7/187M Leishman	1.15e-01
	356	16	5.4	229	20	HHEA35N	H. sapiens partial cDN	1.15e-01
c	357	16	5.4	234	27	B70770	CIT-HSP-2060M11.TF CIT	1.15e-01
c	358	16	5.4	245	25	AA361970	EST71332 T-cell lympho	1.15e-01
	359	16	5.4	246	13	AA834109	of26e09.s1 NCI_CGAP_Ki	1.15e-01
c	360	16	5.4	248	13	AA841327	MB3D6AA6D07T3 Brugia m	1.15e-01
c	361	16	5.4	248	5	R03590	pk09c09.s1 Caenorhabdi	1.15e-01
c	362	16	5.4	249	20	HSC3HF021	H. sapiens partial cDN	1.15e-01
c	363	16	5.4	251	8	AA484707	ne65a03.s1 NCI_CGAP_Al	1.15e-01
	364	16	5.4	259	25	AA338447	EST43376 Fetal brain I	1.15e-01
	365	16	5.4	260	15	AA430991	PMY0861 KG1-a Lambda Z	1.15e-01
c	366	16	5.4	266	28	AQ09B071	HS_3042_A1_F03_MR CIT	1.15e-01
	367	16	5.4	281	14	AA988663	oq88e07.s1 NCI_CGAP_Ki	1.15e-01
c	368	16	5.4	285	12	AA751666	96AS0408 Rice Immature	1.15e-01
c	369	16	5.4	286	28	AQ093523	HS_3026_B1_A08_MR CIT	1.15e-01
c	370	16	5.4	288	26	B43974	HS-1058-B1-F01-MR.abi	1.15e-01
	371	16	5.4	292	22	H48998	SWMPCA045SK Brugia mal	1.15e-01
	372	16	5.4	293	23	AA078367	7P02D11 Chromosome 7 P	1.15e-01
c	373	16	5.4	296	5	T18135	0697c3 Plasmodium falc	1.15e-01
c	374	16	5.4	296	5	R03805	pk12b12.s1 Caenorhabdi	1.15e-01
c	375	16	5.4	297	27	B64875	CIT-HSP-2012A5.TRC CIT	1.15e-01
	376	16	5.4	298	25	AA353740	EST61910 Activated T-c	1.15e-01
c	377	16	5.4	298	5	R25755	yh43d05.r1 Homo sapien	1.15e-01
	378	16	5.4	299	8	AA405154	zu19d10.r1 Soares NhHM	1.15e-01
	379	16	5.4	299	15	AA998663	UI-R-C0-ig-e-02-0-UI.s	1.15e-01
	380	16	5.4	300	20	W64060	md77h12.r1 Soares mous	1.15e-01
c	381	16	5.4	300	12	AA741769	LmLv39p3/444A Leishman	1.15e-01
	382	16	5.4	301	18	AI102381	EST211670 Normalized r	1.15e-01
	383	16	5.4	303	27	B99184	CIT-HSP-2287K11.TR CIT	1.15e-01
	384	16	5.4	306	29	AQ241234	RPCI11-75K19.TK RPCI11	1.15e-01
	385	16	5.4	311	24	AA329880	EST33645 Embryo, 12 we	1.15e-01
c	386	16	5.4	312	28	AQ124865	HS_2163_A2_A01_MF CIT	1.15e-01
	387	16	5.4	313	24	AA275115	vc04f02.r1 Soares mous	1.15e-01
	388	16	5.4	316	11	AA692935	vr56e11.s1 Knowles Sol	1.15e-01
	389	16	5.4	317	27	AQ014151	RPCI11-19D23.TKBF RPCI	1.15e-01
c	390	16	5.4	318	5	T82696	Hd527-f Homo sapiens c	1.15e-01
	391	16	5.4	319	8	AA562871	vl64h05.r1 Knowles Sol	1.15e-01
	392	16	5.4	321	14	H31937	EST106509 Rat PC-12 ce	1.15e-01
c	393	16	5.4	323	23	AA164016	mr18f03.r1 Soares mous	1.15e-01
c	394	16	5.4	323	27	AQ071777	HS_3020_B2_D09_MR CIT	1.15e-01
	395	16	5.4	329	28	AQ197299	CIT-HSP-2385C10.TF CIT	1.15e-01
	396	16	5.4	330	12	AA753209	97AS2292 Rice Immature	1.15e-01
c	397	16	5.4	330	27	AQ058537	CIT-HSP-2344O12.TF CIT	1.15e-01
c	398	16	5.4	332	29	AQ231871	HS_2054_A1_C12_T7 CIT	1.15e-01
c	399	16	5.4	333	24	AA309188	EST180020 Frontal cort	1.15e-01

c	400	16	5.4	336	19	H92068	ys84e04.s1 Homo sapien	1.15e-01
c	401	16	5.4	336	5	R03338	pk06a05.r1 Caenorhabdi	1.15e-01
c	402	16	5.4	339	11	AA684287	vm54b03.r1 Stratagene	1.15e-01
c	403	16	5.4	339	5	T65204	yc74d11.r1 Homo sapien	1.15e-01
c	404	16	5.4	342	11	AA705432	zj90f09.s1 Soares feta	1.15e-01
	405	16	5.4	344	14	AA981095	vx57b11.r1 Stratagene	1.15e-01
	406	16	5.4	344	20	W98362	mg14c12.r1 Soares mous	1.15e-01
c	407	16	5.4	346	21	R93880	yt72g06.r1 Soares reti	1.15e-01
	408	16	5.4	347	14	H31513	EST105607 Rat PC-12 ce	1.15e-01
	409	16	5.4	348	9	N60627	TgESTzy20h11.r1 TgRH T	1.15e-01
	410	16	5.4	352	14	AA930406	vt48b10.r1 Barstead mo	1.15e-01
c	411	16	5.4	354	10	AA584585	no12d07.s1 NCI_CGAP_Ph	1.15e-01
c	412	16	5.4	354	20	W16999	zb09d05.r1 Soares feta	1.15e-01
	413	16	5.4	355	20	W64686	md87h10.r1 Soares mous	1.15e-01
	414	16	5.4	359	27	FR0031319	Fugu rubripes GSS sequ	1.15e-01
c	415	16	5.4	359	9	N61301	TgESTzy19g06.r1 TgRH T	1.15e-01
c	416	16	5.4	359	12	AA724775	ai01g05.s1 Soares NFL	1.15e-01
c	417	16	5.4	360	18	AI177212	EST220822 Normalized r	1.15e-01
	418	16	5.4	362	26	B31514	HS-1011-A1-F11-MR.abi	1.15e-01
c	419	16	5.4	363	5	T05652	EST03541 Homo sapiens	1.15e-01
c	420	16	5.4	363	22	W84903	SWAMCA1016SK Brugia ma	1.15e-01
c	421	16	5.4	364	17	AI111044	MP001G02.NK3 S. manson	1.15e-01
c	422	16	5.4	367	17	AI072573	UI-R-C2-nn-a-05-0-UI.s	1.15e-01
c	423	16	5.4	367	26	AQ011892	HS_2172_B1_MR_D05 CIT	1.15e-01
c	424	16	5.4	368	6	AA128032	zl14a07.r1 Soares preg	1.15e-01
c	425	16	5.4	371	27	AQ064518	HS_2211_A2_D10_MR CIT	1.15e-01
	426	16	5.4	376	18	AI101941	EST211230 Normalized r	1.15e-01
	427	16	5.4	377	16	H08756	yl94a04.s1 Homo sapien	1.15e-01
c	428	16	5.4	377	21	AA063737	T3398 MVAT4 bloodstrea	1.15e-01
c	429	16	5.4	377	19	H61617	yr23b06.r1 Homo sapien	1.15e-01
	430	16	5.4	381	5	R14461	yf82c12.r1 Homo sapien	1.15e-01
	431	16	5.4	382	5	T98968	ye66c03.r1 Homo sapien	1.15e-01
	432	16	5.4	383	18	AI088355	qb07c07.x1 Soares_pree	1.15e-01
c	433	16	5.4	385	8	AA527421	ng41e01.s1 NCI_CGAP_Co	1.15e-01
	434	16	5.4	387	16	R44519	yq24h04.s1 Homo sapien	1.15e-01
	435	16	5.4	387	28	AQ199484	RPCI11-47G20.TJ RPCI11	1.15e-01
c	436	16	5.4	387	12	AA263449	LD07221.3prime LD Dros	1.15e-01
	437	16	5.4	388	29	AQ139497	HS_3088_A2_G02_MR CIT	1.15e-01
c	438	16	5.4	388	17	AI120561	ue89f06.x1 Sugano mous	1.15e-01
c	439	16	5.4	392	6	AA413215	ve52a05.r1 Beddington	1.15e-01
	440	16	5.4	393	20	W75361	me54b06.r1 Soares mous	1.15e-01
c	441	16	5.4	393	16	R96151	yt84a12.r1 Homo sapien	1.15e-01
c	442	16	5.4	393	6	AA045848	zk76f06.s1 Soares preg	1.15e-01
	443	16	5.4	396	8	AA474453	vd52c12.s1 Soares mous	1.15e-01
c	444	16	5.4	399	17	AU016947	Mus musculus 2-cell em	1.15e-01
c	445	16	5.4	399	28	AQ111750	CIT-HSP-2379H7.TR CIT-	1.15e-01
c	446	16	5.4	400	5	T92076	ye01a11.r1 Homo sapien	1.15e-01
c	447	16	5.4	400	8	AA132903	zo21h01.r1 Stratagene	1.15e-01
c	448	16	5.4	400	15	AI008916	EST203367 Normalized r	1.15e-01
c	449	16	5.4	401	8	AA134190	zo20b11.r1 Stratagene	1.15e-01
	450	16	5.4	403	28	AQ215804	HS_3217_B2_H10_MR CIT	1.15e-01

	451	16	5.4	403	8	AA514245	nf56f09.s1 NCI_CGAP_Co	1.15e-01
c	452	16	5.4	403	13	AA855659	vw70g09.r1 Stratagene	1.15e-01
c	453	16	5.4	403	29	AQ139131	HS_3080_B2_C05_MR CIT	1.15e-01
	454	16	5.4	406	15	AI031041	UI-R-C0-jp-d-10-0-UI.s	1.15e-01
c	455	16	5.4	406	19	H11512	ym59a05.r1 Homo sapien	1.15e-01
c	456	16	5.4	407	26	B00468	cSRL-1113g7-u cSRL flow	1.15e-01
c	457	16	5.4	408	27	B93488	CIT-HSP-2172H9.TR CIT-	1.15e-01
	458	16	5.4	409	13	AA821794	vp72a07.r1 Knowles Sol	1.15e-01
c	459	16	5.4	409	19	H59068	yr39b04.s1 Homo sapien	1.15e-01
c	460	16	5.4	409	28	AQ131069	HS_3034_A1_H03_MR CIT	1.15e-01
c	461	16	5.4	409	12	AA754871	vu55a03.r1 Soares mous	1.15e-01
c	462	16	5.4	411	15	AA997528	UI-R-C0-hw-h-11-0-UI.s	1.15e-01
c	463	16	5.4	412	14	AA925942	UI-R-A1-eo-a-08-0-UI.s	1.15e-01
c	464	16	5.4	412	26	B47279	HS-1056-A2-D04-MF.abi	1.15e-01
c	465	16	5.4	413	13	AA855658	vw70g08.r1 Stratagene	1.15e-01
	466	16	5.4	414	13	AA839534	uc99a05.r1 Soares mous	1.15e-01
c	467	16	5.4	414	5	T77596	yc91e02.r1 Homo sapien	1.15e-01
	468	16	5.4	416	9	AA590910	vm21e04.r1 Knowles Sol	1.15e-01
c	469	16	5.4	416	9	AA579675	nm72d02.s1 NCI_CGAP_Co	1.15e-01
c	470	16	5.4	416	25	AA342577	EST48201 Fetal spleen	1.15e-01
	471	16	5.4	417	10	AA395117	26900 Lambda-PRL2 Arab	1.15e-01
c	472	16	5.4	417	13	AA863662	vx06e06.r1 Soares 2NbM	1.15e-01
	473	16	5.4	419	13	AA167217	zp13b11.s1 Stratagene	1.15e-01
c	474	16	5.4	419	22	AA121866	zn96c12.r1 Stratagene	1.15e-01
	475	16	5.4	422	28	AQ183341	HS_3138_B2_D10_MR CIT	1.15e-01
c	476	16	5.4	422	19	N68841	za23d02.s1 Homo sapien	1.15e-01
	477	16	5.4	423	19	N39114	yy46a04.s1 Homo sapien	1.15e-01
c	478	16	5.4	423	28	AQ187001	HS_3122_B2_E03_MR CIT	1.15e-01
c	479	16	5.4	424	24	AA309370	EST180243 Liver III Ho	1.15e-01
c	480	16	5.4	425	26	B40218	HS-1051-A1-F02-MF.abi	1.15e-01
c	481	16	5.4	426	27	AQ035794	CIT-HSP-2333J24.TR CIT	1.15e-01
c	482	16	5.4	427	26	B48640	RPCI11-2M16.TP RPCI11	1.15e-01
c	483	16	5.4	428	18	AI150279	qf34h08.x1 Soares_test	1.15e-01
c	484	16	5.4	428	28	AQ212904	HS_3173_A2_E04_T7 CIT	1.15e-01
	485	16	5.4	429	5	H20328	yg35f11.r1 Homo sapien	1.15e-01
	486	16	5.4	430	18	AI079479	ox39g07.s1 Soares_NhHM	1.15e-01
c	487	16	5.4	430	15	AI037795	ub51f11.s1 Soares mous	1.15e-01
c	488	16	5.4	433	8	AA550689	1926m3 gmbPfHB3.1, G.	1.15e-01
	489	16	5.4	434	23	AA144872	mr74b01.r1 Stratagene	1.15e-01
c	490	16	5.4	434	28	AQ223972	HS_2259_B2_H01_T7 CIT	1.15e-01
c	491	16	5.4	436	26	FR0016116	F.rubripes GSS sequenc	1.15e-01
c	492	16	5.4	437	27	FR0029045	Fugu rubripes GSS sequ	1.15e-01
	493	16	5.4	438	20	W59072	md67g12.r1 Soares mous	1.15e-01
	494	16	5.4	438	28	AQ202669	RPCI11-67L1.TJ RPCI11	1.15e-01
c	495	16	5.4	438	17	AI077974	oy39d01.s1 Soares_para	1.15e-01
c	496	16	5.4	439	29	AQ238931	RPCI11-7A7.TK RPCI11	1.15e-01
	497	16	5.4	440	20	W04980	za45e11.r1 Soares feta	1.15e-01
c	498	16	5.4	440	27	B59895	CIT-HSP-347A20.TVC CIT	1.15e-01
c	499	16	5.4	441	5	R24799	yg43c11.r1 Homo sapien	1.15e-01
c	500	16	5.4	444	29	AQ142262	HS_3174_B2_C02_MR CIT	1.15e-01
c	501	16	5.4	444	7	AA512798	vg13c07.r1 Soares mous	1.15e-01

c	502	16	5.4	444	7	AA520532	TgESTzz62g03.r1 TgME49	1.15e-01
	503	16	5.4	445	7	AA458098	vg45d06.r1 Soares mous	1.15e-01
c	504	16	5.4	445	23	W85829	zh55d11.s1 Soares feta	1.15e-01
c	505	16	5.4	448	24	AA310015	EST180928 Jurkat T-cel	1.15e-01
	506	16	5.4	449	28	AQ178352	HS_2219_B1_H12_MR CIT	1.15e-01
	507	16	5.4	450	23	AA240324	mv31d09.r1 GuayWoodfor	1.15e-01
	508	16	5.4	451	22	N54342	yv35h05.s1 Soares feta	1.15e-01
c	509	16	5.4	451	10	AA626445	ab49h02.r1 Stratagene	1.15e-01
	510	16	5.4	452	27	AQ035409	CIT-HSP-2329D18.TF CIT	1.15e-01
c	511	16	5.4	452	12	AA820224	LD23726.5prime LD Dros	1.15e-01
c	512	16	5.4	453	29	AQ238583	RPCI11-70P6.TK RPCI11	1.15e-01
c	513	16	5.4	453	19	N22488	yw35e09.s1 Homo sapien	1.15e-01
	514	16	5.4	454	14	AA984819	am93e08.s1 Stratagene	1.15e-01
c	515	16	5.4	454	26	B75680	RPCI11-11P14.TV RPCI11	1.15e-01
	516	16	5.4	455	8	AA259996	za35c10.r1 NCI_CGAP_GC	1.15e-01
c	517	16	5.4	455	10	AA611138	vo61g10.r1 Soares mous	1.15e-01
c	518	16	5.4	456	29	AQ242525	HS_2060_A1_A07_MR CIT	1.15e-01
	519	16	5.4	458	27	AQ054838	CIT-HSP-2338K12.TF CIT	1.15e-01
	520	16	5.4	459	27	FR0028493	Fugu rubripes GSS sequ	1.15e-01
	521	16	5.4	459	11	AA694169	zi35h10.s1 Soares feta	1.15e-01
	522	16	5.4	459	29	AQ226794	HS_2016_B1_A01_T7 CIT	1.15e-01
	523	16	5.4	460	18	AI154294	ud31b01.r1 Soares mous	1.15e-01
c	524	16	5.4	460	14	AA886876	oj25h06.s1 NCI_CGAP_Ki	1.15e-01
c	525	16	5.4	461	16	H14186	ym62b09.r1 Homo sapien	1.15e-01
c	526	16	5.4	461	28	AQ179179	HS_3120_A1_H09_MR CIT	1.15e-01
c	527	16	5.4	463	28	AQ222299	HS_2007_A2_A07_MR CIT	1.15e-01
	528	16	5.4	464	23	AA146126	mq88e01.r1 Stratagene	1.15e-01
c	529	16	5.4	465	5	R02606	ye76a07.r1 Homo sapien	1.15e-01
	530	16	5.4	466	21	AA041740	mj07h03.r1 Soares mous	1.15e-01
c	531	16	5.4	467	8	AA526513	ni60f01.s1 NCI_CGAP_Ov	1.15e-01
	532	16	5.4	468	16	H53313	yq84d09.r1 Homo sapien	1.15e-01
	533	16	5.4	469	17	AA934264	ka07c06.s1 Onchocerca	1.15e-01
c	534	16	5.4	469	27	B85721	RPCI11-20M12.TV RPCI11	1.15e-01
	535	16	5.4	470	16	R88153	ym90b02.r1 Homo sapien	1.15e-01
	536	16	5.4	470	18	AI153536	zc51h09.r1 Soares 2NbM	1.15e-01
c	537	16	5.4	470	13	AA839251	vw95e10.r1 Stratagene	1.15e-01
c	538	16	5.4	470	19	H58144	yr19h09.r1 Homo sapien	1.15e-01
c	539	16	5.4	471	28	AQ224763	HS_2013_A1_A03_MR CIT	1.15e-01
	540	16	5.4	472	22	AA212635	mu80d03.r1 Stratagene	1.15e-01
	541	16	5.4	472	13	C86104	Mus musculus fertilize	1.15e-01
c	542	16	5.4	472	5	T38918	EST104484 Saccharomyce	1.15e-01
c	543	16	5.4	473	29	AQ241332	CITBI-E1-2505G1.TF.1 C	1.15e-01
c	544	16	5.4	475	13	AA843400	aj16h08.s1 Soares para	1.15e-01
c	545	16	5.4	476	23	AA137732	mq89h11.r1 Stratagene	1.15e-01
	546	16	5.4	478	11	AA672193	vk90c05.s1 Knowles Sol	1.15e-01
	547	16	5.4	478	13	AA434478	aa84d05.s1 Stratagene	1.15e-01
	548	16	5.4	479	26	AP017579	Homo sapiens YAC 307A1	1.15e-01
c	549	16	5.4	479	19	N68717	za20a06.s1 Homo sapien	1.15e-01
	550	16	5.4	481	21	AA032934	mi22e06.r1 Soares mous	1.15e-01
	551	16	5.4	482	28	AQ203636	HS_3095_B2_A10_T7 CIT	1.15e-01
	552	16	5.4	482	27	B85524	RPCI11-19B13.TP RPCI11	1.15e-01

	553	16	5.4	483	16	H10247	ym02c02.r1 Homo sapien	1.15e-01
	554	16	5.4	483	8	AA562799	vl44h07.r1 Stratagene	1.15e-01
	555	16	5.4	483	21	W89741	mf77a10.r1 Soares mous	1.15e-01
c	556	16	5.4	485	29	AQ144251	HS_3085_A2_F07_MF CIT	1.15e-01
c	557	16	5.4	485	12	AA539319	LD19014.5prime LD Dros	1.15e-01
c	558	16	5.4	485	9	AA590856	vn60e08.r1 Barstead mo	1.15e-01
	559	16	5.4	486	10	AA415671	vd26d07.s1 Knowles Sol	1.15e-01
	560	16	5.4	486	20	W25213	zb68g12.r1 Soares feta	1.15e-01
	561	16	5.4	486	16	R55849	yg89e02.r1 Homo sapien	1.15e-01
c	562	16	5.4	486	19	N68744	za20e05.s1 Homo sapien	1.15e-01
c	563	16	5.4	486	28	AQ222498	HS_3252_B1_A09_T7 CIT	1.15e-01
	564	16	5.4	487	23	AA170627	ms92h10.r1 Soares mous	1.15e-01
	565	16	5.4	488	17	AI004292	ou11h02.x1 Soares_NFL_	1.15e-01
c	566	16	5.4	488	14	AA830804	oc56d07.s1 NCI_CGAP_GC	1.15e-01
c	567	16	5.4	488	28	AQ123297	HS_3090_B1_G01_T7 CIT	1.15e-01
	568	16	5.4	490	29	AQ138200	HS_3074_B1_E08_MF CIT	1.15e-01
c	569	16	5.4	490	15	AI057618	oy31f08.x1 Soares_para	1.15e-01
	570	16	5.4	491	18	AI151346	qc79f04.x1 Soares_plac	1.15e-01
c	571	16	5.4	491	7	AA444207	ve34e03.r1 Ko mouse em	1.15e-01
	572	16	5.4	492	15	AA908777	ol06b06.s1 NCI_CGAP_Lu	1.15e-01
	573	16	5.4	496	19	N66580	yy69f01.s1 Homo sapien	1.15e-01
	574	16	5.4	498	13	AA223943	zr13e05.s1 Stratagene	1.15e-01
c	575	16	5.4	498	18	AI172571	UI-R-C2p-nz-g-03-0-UI.	1.15e-01
c	576	16	5.4	498	13	AA206621	zq56b05.r1 Stratagene	1.15e-01
c	577	16	5.4	498	23	AA170195	ms73a09.r1 Soares mous	1.15e-01
c	578	16	5.4	500	7	AA516632	vh83g10.r1 Knowles Sol	1.15e-01
c	579	16	5.4	500	6	AA047104	zk74b05.r1 Soares preg	1.15e-01
	580	16	5.4	501	29	AQ148393	HS_3123_A2_F04_MR CIT	1.15e-01
c	581	16	5.4	501	28	AQ191411	HS_3225_B1_H06_T7 CIT	1.15e-01
	582	16	5.4	502	28	AQ157649	nbxb0009J02f CUGI Rice	1.15e-01
	583	16	5.4	502	28	AQ197041	CIT-HSP-2380A6.TR CIT-	1.15e-01
c	584	16	5.4	503	5	R35158	yg62b04.r1 Homo sapien	1.15e-01
c	585	16	5.4	504	28	AQ204569	HS_3229_B1_A06_T7 CIT	1.15e-01
	586	16	5.4	505	6	C06779	similar to none, mRNA	1.15e-01
	587	16	5.4	505	24	AA286629	vb87a05.r1 Soares mous	1.15e-01
	588	16	5.4	505	10	AA419814	vf48f05.r1 Soares mous	1.15e-01
c	589	16	5.4	505	18	AI088924	qa21f08.x1 NCI_CGAP_Br	1.15e-01
	590	16	5.4	507	28	AQ173084	HS_3204_A1_H07_T7 CIT	1.15e-01
c	591	16	5.4	511	26	B62461	T21N16TR TAMU Arabidop	1.15e-01
c	592	16	5.4	511	18	AI176598	EST220186 Normalized r	1.15e-01
c	593	16	5.4	512	27	AQ009204	CIT-HFP-2025P22.TF CIT	1.15e-01
c	594	16	5.4	513	12	AA772081	ai41g08.s1 Soares para	1.15e-01
c	595	16	5.4	514	29	AQ140021	HS_3106_A1_G01_T7 CIT	1.15e-01
c	596	16	5.4	515	15	AA922108	om41g04.s1 Soares_NFL_	1.15e-01
c	597	16	5.4	516	28	AQ180937	HS_3221_A2_B11_T7 CIT	1.15e-01
	598	16	5.4	517	12	AA742277	ny28e09.s1 NCI_CGAP_GC	1.15e-01
	599	16	5.4	520	28	AQ183417	HS_3164_A1_D01_MR CIT	1.15e-01
c	600	16	5.4	520	8	AA523516	ni45f08.s1 NCI_CGAP_Lu	1.15e-01
	601	16	5.4	521	14	AA900588	UI-R-E0-dn-c-06-0-UI.s	1.15e-01
c	602	16	5.4	527	7	AA520479	TgESTzz61d12.r1 TgME49	1.15e-01
c	603	16	5.4	527	29	AQ235942	HS_2056_A1_B07_T7 CIT	1.15e-01

c	604	16	5.4	527	8	AA523708	ni50c09.s1 NCI_CGAP_Ov	1.15e-01
c	605	16	5.4	528	12	AA762518	vv84c10.r1 Stratagene	1.15e-01
	606	16	5.4	532	12	AA740763	nz03c04.s1 NCI_CGAP_GC	1.15e-01
c	607	16	5.4	534	8	AA292206	zt50g02.r1 Soares ovar	1.15e-01
	608	16	5.4	537	14	D42853	Rice callus cDNA, 366,	1.15e-01
	609	16	5.4	537	7	SCMRAP0331	Schistosoma mansoni (c	1.15e-01
	610	16	5.4	537	28	AQ203045	RPCI11-67N17.TJ RPCI11	1.15e-01
	611	16	5.4	541	26	AQ023806	CpGR0172B Cryptosporid	1.15e-01
	612	16	5.4	545	28	AQ170694	HS_2270_B2_B10_MR CIT	1.15e-01
c	613	16	5.4	546	12	AA141079	CK01178.3prime CK Dros	1.15e-01
	614	16	5.4	549	12	AA441371	LD16278.5prime LD Dros	1.15e-01
	615	16	5.4	550	11	AA131369	zo08c06.s1 Stratagene	1.15e-01
	616	16	5.4	555	17	AU024423	Mus musculus unfertili	1.15e-01
	617	16	5.4	555	18	AI178155	EST221820 Normalized r	1.15e-01
	618	16	5.4	558	14	AA987135	uc80g01.x1 Sugano mous	1.15e-01
	619	16	5.4	563	9	AA552971	nk83b05.s1 NCI_CGAP_Sc	1.15e-01
c	620	16	5.4	563	6	AA128056	zl14b06.r1 Soares preg	1.15e-01
	621	16	5.4	566	27	B58894	CIT-HSP-2014E24.TF CIT	1.15e-01
c	622	16	5.4	567	23	AA245177	mw28f05.r1 Soares mous	1.15e-01
	623	16	5.4	568	15	C77981	Mus musculus 3.5-dpc b	1.15e-01
c	624	16	5.4	568	11	AA728426	33605 CD4-6 Arabidopsi	1.15e-01
c	625	16	5.4	569	16	H17773	ym40h02.r1 Homo sapien	1.15e-01
c	626	16	5.4	570	11	H36197	14719 Lambda-PRL2 Arab	1.15e-01
	627	16	5.4	571	26	B78198	T31K13TF TAMU Arabidop	1.15e-01
	628	16	5.4	579	28	AQ120044	HS_3022_B2_E04_T7 CIT	1.15e-01
c	629	16	5.4	579	22	AA203595	zx59e08.r1 Soares feta	1.15e-01
	630	16	5.4	582	17	AA556418	273 Loblolly pine C Pi	1.15e-01
	631	16	5.4	583	27	FR0028520	Fugu rubripes GSS sequ	1.15e-01
	632	16	5.4	584	15	C81520	Mus musculus 3.5-dpc b	1.15e-01
	633	16	5.4	585	17	AU015349	Mus musculus 2-cell em	1.15e-01
c	634	16	5.4	587	15	C79426	Mus musculus 3.5-dpc b	1.15e-01
	635	16	5.4	600	27	FR0033801	Fugu rubripes GSS sequ	1.15e-01
c	636	16	5.4	608	23	W94988	zh47h06.r1 Soares feta	1.15e-01
	637	16	5.4	609	6	AA127645	zk92g02.s1 Soares preg	1.15e-01
	638	16	5.4	614	18	AI177100	EST220707 Normalized r	1.15e-01
c	639	16	5.4	616	17	AI109619	GH08858.5prime GH Dros	1.15e-01
	640	16	5.4	619	27	FR0032485	Fugu rubripes GSS sequ	1.15e-01
	641	16	5.4	619	26	FR0016074	F.rubripes GSS sequenc	1.15e-01
c	642	16	5.4	619	27	B79886	CIT-HSP-2050M8.TR CIT-	1.15e-01
c	643	16	5.4	623	12	AA390904	LD11749.5prime LD Dros	1.15e-01
c	644	16	5.4	628	27	AQ057314	CIT-HSP-2337K23.TR CIT	1.15e-01
	645	16	5.4	630	29	AF005806	Arabidopsis thaliana 2	1.15e-01
	646	16	5.4	634	15	AI053062	BSBmMFSZ01H22SK Brugia	1.15e-01
c	647	16	5.4	634	17	AI124501	SWOvL3CAN50F04SK Oncho	1.15e-01
c	648	16	5.4	634	18	AI178158	EST221823 Normalized r	1.15e-01
	649	16	5.4	637	28	AG014456	Homo sapiens genomic D	1.15e-01
	650	16	5.4	637	26	B76181	RPCI11-13N24.TF RPCI11	1.15e-01
c	651	16	5.4	645	29	AQ247528	HS_2062_B2_D09_MR CIT	1.15e-01
c	652	16	5.4	648	28	AG014414	Homo sapiens genomic D	1.15e-01
c	653	16	5.4	650	27	AQ038849	CIT-HSP-2321J17.TF CIT	1.15e-01
	654	16	5.4	653	15	AI008661	EST203112 Normalized r	1.15e-01

c	655	16	5.4	654	14	AA968292	uc71g03.y1 Sugano mous	1.15e-01
	656	16	5.4	655	29	AQ238474	RPCI11-66P10.TJ RPCI11	1.15e-01
c	657	16	5.4	658	27	AQ051299	nbxb0001cH02f CUGI Ric	1.15e-01
c	658	16	5.4	659	28	AG014413	Homo sapiens genomic D	1.15e-01
c	659	16	5.4	663	27	AQ027438	CIT-HSP-2315I6.TR CIT-	1.15e-01
c	660	16	5.4	669	12	AA820605	LD24419.5prime LD Dros	1.15e-01
	661	16	5.4	676	27	AQ048786	cLM-10a10-u cLM Giardi	1.15e-01
c	662	16	5.4	676	10	AA646243	vn14d07.r1 Stratagene	1.15e-01
c	663	16	5.4	679	27	B94133	CIT-HSP-2163A2.TF CIT-	1.15e-01
c	664	16	5.4	681	26	B97218	T32L16TR TAMU Arabidop	1.15e-01
	665	16	5.4	691	26	AQ018076	CIT-HSP-2307D2.TR CIT-	1.15e-01
	666	16	5.4	698	26	B71786	RPCI11-10N3.TP RPCI11	1.15e-01
	667	16	5.4	710	28	AQ163424	nbxb0006P10f CUGI Rice	1.15e-01
	668	16	5.4	720	27	AG011609	Homo sapiens genomic D	1.15e-01
	669	16	5.4	720	26	AG004929	Homo sapiens genomic D	1.15e-01
c	670	16	5.4	723	28	AQ158370	nbxb0011A02f CUGI Rice	1.15e-01
c	671	16	5.4	739	27	B91087	CIT-HSP-2173I2.TR CIT-	1.15e-01
	672	16	5.4	742	27	AG012423	Homo sapiens genomic D	1.15e-01
c	673	16	5.4	764	28	AQ114907	CIT-HSP-2376N11.TR CIT	1.15e-01
	674	16	5.4	773	14	AA876129	ob93g01.s1 NCI_CGAP_GC	1.15e-01
c	675	16	5.4	944	20	W26634	34b10 Human retina cDN	1.15e-01
c	676	16	5.4	989	21	W08460	mb50b12.r1 Soares mous	1.15e-01
	677	16	5.4	1406	26	B10648	T10I5-T7 TAMU Arabidop	1.15e-01
c	678	15	5.0	139	14	AA990132	ua60a01.r1 Soares 2NbM	2.76e+00
c	679	15	5.0	148	9	AA585203	KTH109 HTCDL1 Homo sap	2.76e+00
	680	15	5.0	183	16	H49017	EST000069 Schistosoma	2.76e+00
c	681	15	5.0	194	5	T11834	A1055R Homo sapiens cD	2.76e+00
c	682	15	5.0	196	20	HSC03H122	H. sapiens partial cDN	2.76e+00
c	683	15	5.0	199	22	N97842	2134C3 czapPFDd2.1, De	2.76e+00
	684	15	5.0	237	25	AA333358	EST37420 Embryo, 8 wee	2.76e+00
c	685	15	5.0	249	14	AA889063	am34h04.s1 Soares_NFL_	2.76e+00
c	686	15	5.0	256	22	N58158	yv65g05.s1 Soares feta	2.76e+00
	687	15	5.0	260	19	N55036	yz20c12.r1 Homo sapien	2.76e+00
	688	15	5.0	264	15	T35898	EST93232 Homo sapiens	2.76e+00
c	689	15	5.0	266	8	AA256063	zs29e11.r1 NCI_CGAP_GC	2.76e+00
	690	15	5.0	268	27	B69631	CIT-HSP-2059M24.TR CIT	2.76e+00
	691	15	5.0	269	8	AA547517	vk83a05.s1 Knowles Sol	2.76e+00
	692	15	5.0	272	19	N36516	yx85a07.r1 Homo sapien	2.76e+00
	693	15	5.0	281	8	AA251202	zs03g10.r1 NCI_CGAP_GC	2.76e+00
c	694	15	5.0	287	19	H80738	yv78b12.r1 Homo sapien	2.76e+00
c	695	15	5.0	289	15	AI026176	MEEG0412.M1F Schistoso	2.76e+00
	696	15	5.0	295	12	AA761377	nz25g05.s1 NCI_CGAP_GC	2.76e+00
c	697	15	5.0	295	20	N77921	za93c01.s1 Homo sapien	2.76e+00
	698	15	5.0	298	25	AA362968	EST72751 Ovary II Homo	2.76e+00
c	699	15	5.0	304	15	AI044745	UI-R-C1-ki-f-10-0-UI.s	2.76e+00
c	700	15	5.0	304	12	AA816378	LD01319.5prime LD Dros	2.76e+00
	701	15	5.0	306	5	R16891	yf44d05.s2 Homo sapien	2.76e+00
	702	15	5.0	308	19	H75688	yr77h02.r1 Homo sapien	2.76e+00
c	703	15	5.0	308	5	T23562	seq3817 Homo sapiens c	2.76e+00
c	704	15	5.0	314	14	AA975098	on06b01.s1 NCI_CGAP_Le	2.76e+00
c	705	15	5.0	314	5	R41519	yf96e03.s1 Homo sapien	2.76e+00

	706	15	5.0	318	28	AQ119717	HS_2175_A1_A11_MF CIT	2.76e+00
	707	15	5.0	321	15	AA906798	ok73a03.s1 NCI_CGAP_GC	2.76e+00
	708	15	5.0	326	15	AI010404	EST204855 Normalized r	2.76e+00
	709	15	5.0	327	27	AQ083459	RPCI11-55H23.TJ RPCI11	2.76e+00
	710	15	5.0	327	20	HSAAAAVGP	H. sapiens putatively	2.76e+00
	711	15	5.0	328	11	AA074557	sm76b01.s1 Stratagene	2.76e+00
	712	15	5.0	330	26	B34405	HS-1024-B1-H05-MF.abi	2.76e+00
c	713	15	5.0	330	16	T33241	EST57183 Homo sapiens	2.76e+00
c	714	15	5.0	332	27	B89647	CIT-HSP-2170B6.TR CIT-	2.76e+00
c	715	15	5.0	334	19	HSC1DG012	H. sapiens partial cDN	2.76e+00
	716	15	5.0	337	13	AA859140	UI-R-A0-bn-b-01-0-UI.s	2.76e+00
	717	15	5.0	337	14	AA928746	on98a06.s1 Soares_NFL_	2.76e+00
	718	15	5.0	339	28	AQ197693	CIT-HSP-2380K6.TR CIT-	2.76e+00
	719	15	5.0	340	17	AI083052	JAA00051.QA3 Adult SjC	2.76e+00
	720	15	5.0	343	13	AA868375	ak41d09.s1 Soares test	2.76e+00
	721	15	5.0	345	15	AI005899	ua84a05.r1 Soares mous	2.76e+00
c	722	15	5.0	345	26	B31843	HS-1012-B1-G04-MR.abi	2.76e+00
c	723	15	5.0	345	14	AA904292	od89b05.s1 NCI_CGAP_Br	2.76e+00
	724	15	5.0	347	19	N26754	yx96b04.s1 Homo sapien	2.76e+00
	725	15	5.0	348	21	T16164	IB3615 Infant brain, B	2.76e+00
	726	15	5.0	349	19	HSC2EF091	H. sapiens partial cDN	2.76e+00
	727	15	5.0	352	29	AQ238098	RPCI11-73J16.TK RPCI11	2.76e+00
	728	15	5.0	354	16	M78546	EST00694 Homo sapiens	2.76e+00
c	729	15	5.0	355	7	AA461713	vg65g05.r1 Soares mous	2.76e+00
	730	15	5.0	356	15	AA910629	ok51d04.s1 NCI_CGAP_GC	2.76e+00
c	731	15	5.0	358	16	R81702	yj04e12.s1 Homo sapien	2.76e+00
c	732	15	5.0	359	7	RICS11930A	Rice cDNA, partial seq	2.76e+00
c	733	15	5.0	364	22	AA154592	mt64a03.r1 Soares mous	2.76e+00
c	734	15	5.0	364	18	AI168102	oz91e10.x1 Soares_para	2.76e+00
	735	15	5.0	367	27	B81256	CIT-HSP-2022M24.TFB CI	2.76e+00
c	736	15	5.0	367	27	AQ088377	HS_3000_A2_E03_MR CIT	2.76e+00
	737	15	5.0	368	26	B46264	HS-1063-A2-F05-MF.abi	2.76e+00
c	738	15	5.0	369	12	AA819454	UI-R-A0-bk-b-10-0-UI.s	2.76e+00
c	739	15	5.0	369	17	AI117780	uc39f12.r1 Soares mous	2.76e+00
c	740	15	5.0	374	13	AA883407	al56e08.s1 Soares NFL_	2.76e+00
c	741	15	5.0	374	27	B91010	CIT-HSP-2174O16.TR CIT	2.76e+00
	742	15	5.0	375	14	AA982265	ua49f04.r1 Stratagene	2.76e+00
	743	15	5.0	375	15	AA953286	oo87h03.s1 NCI_CGAP_Ki	2.76e+00
c	744	15	5.0	375	26	B40876	HS-1052-B1-E11-MF.abi	2.76e+00
c	745	15	5.0	376	10	AA666756	vm49h12.r1 Stratagene	2.76e+00
	746	15	5.0	377	18	AI184731	qd64c10.x1 Soares_test	2.76e+00
c	747	15	5.0	378	22	AA035695	ze87f06.r1 Soares feta	2.76e+00
	748	15	5.0	379	27	AQ080782	CIT-HSP-2358F15.TR CIT	2.76e+00
	749	15	5.0	379	6	AA433265	EST79 Drosophila melan	2.76e+00
	750	15	5.0	379	29	AQ146614	HS_2233_B2_A01_MF CIT	2.76e+00
c	751	15	5.0	379	5	R53480	yg83h10.s1 Homo sapien	2.76e+00
	752	15	5.0	381	28	AQ094410	HS_3023_B1_A07_T7 CIT	2.76e+00
c	753	15	5.0	383	18	AI159115	vz81h07.r1 Soares mous	2.76e+00
	754	15	5.0	384	11	AA698566	HL04936.5prime HL Dros	2.76e+00
	755	15	5.0	384	24	AA275134	vc05c04.r1 Soares mous	2.76e+00
c	756	15	5.0	385	13	AA173481	zp47f10.s1 Stratagene	2.76e+00

	757	15	5.0	387	23	AA253633	mv36e09.r1 GuayWoodfor	2.76e+00
c	758	15	5.0	387	29	AQ138400	HS_3075_B1_A07_MR CIT	2.76e+00
c	759	15	5.0	388	5	R07438	ye96c12.s1 Homo sapien	2.76e+00
c	760	15	5.0	389	11	AA083084	zn08g08.s1 Stratagene	2.76e+00
c	761	15	5.0	390	18	AI169996	EST215912 Normalized r	2.76e+00
	762	15	5.0	391	5	T80304	yd07a07.r1 Homo sapien	2.76e+00
	763	15	5.0	393	17	AI067299	EST208977 Schistosoma	2.76e+00
c	764	15	5.0	393	28	AQ101093	HS_3067_A2_G04_MR CIT	2.76e+00
c	765	15	5.0	393	5	R43058	yg17b03.s1 Homo sapien	2.76e+00
	766	15	5.0	394	11	AA718659	vt86h09.r1 Soares mous	2.76e+00
c	767	15	5.0	394	21	N95835	zb66f06.s1 Soares feta	2.76e+00
c	768	15	5.0	394	23	AA177333	mt17g09.r1 Soares mous	2.76e+00
c	769	15	5.0	395	28	AQ134604	HS_3058_A1_H01_MR CIT	2.76e+00
	770	15	5.0	396	25	AA351635	EST59432 Infant brain	2.76e+00
c	771	15	5.0	396	17	AA994058	ou37c04.s1 Soares_NFL_	2.76e+00
	772	15	5.0	398	26	AQ010040	HS_2262_A1_MR_G03 CIT	2.76e+00
c	773	15	5.0	398	5	R22809	yh20f02.s1 Homo sapien	2.76e+00
c	774	15	5.0	398	24	AA314502	EST186337 Colon carcin	2.76e+00
	775	15	5.0	399	8	AA252082	zr63f09.s1 Soares NhHM	2.76e+00
c	776	15	5.0	399	28	AQ200652	RPCI11-48F4.TJ RPCI11	2.76e+00
	777	15	5.0	400	16	H03062	yj46f09.r1 Homo sapien	2.76e+00
c	778	15	5.0	400	5	T84838	yd54g06.r1 Homo sapien	2.76e+00
	779	15	5.0	402	22	AA182938	zp51f02.r1 Stratagene	2.76e+00
c	780	15	5.0	402	27	B88139	RPCI11-18O11.TPB RPCI1	2.76e+00
c	781	15	5.0	402	5	R14845	yf42a03.s1 Homo sapien	2.76e+00
	782	15	5.0	404	28	AQ180320	HS_3225_B1_E07_MR CIT	2.76e+00
	783	15	5.0	404	22	AA020947	ze64d10.r1 Soares reti	2.76e+00
	784	15	5.0	405	8	AA236558	zs39e01.s1 Soares NhHM	2.76e+00
c	785	15	5.0	406	28	AQ170018	HS_3178_B1_A05_MR CIT	2.76e+00
c	786	15	5.0	410	14	AA959083	ua08e12.r1 Soares 2NbM	2.76e+00
	787	15	5.0	411	22	AA088393	zl82g03.r1 Stratagene	2.76e+00
	788	15	5.0	413	27	AQ034790	RPCI11-31N11.TJ CIT-HS	2.76e+00
	789	15	5.0	413	24	AA301628	EST14813 Aorta endothe	2.76e+00
	790	15	5.0	413	23	AA063495	zf70b06.r1 Soares pine	2.76e+00
	791	15	5.0	413	27	AQ033225	HS_2226_A2_C12_MF CIT	2.76e+00
c	792	15	5.0	415	29	AQ153651	HS_2243_B2_A10_MR CIT	2.76e+00
c	793	15	5.0	417	29	AQ153677	HS_2243_B2_D05_MR CIT	2.76e+00
c	794	15	5.0	420	11	AA101146	zn25d09.s1 Stratagene	2.76e+00
	795	15	5.0	421	17	AI118714	uc11g12.r1 Soares mous	2.76e+00
c	796	15	5.0	421	5	T84108	yd46g04.s1 Homo sapien	2.76e+00
c	797	15	5.0	421	27	AQ089797	HS_3008_B2_D11_T7 CIT	2.76e+00
c	798	15	5.0	421	26	AQ011749	HS_2260_A1_MF_F08 CIT	2.76e+00
c	799	15	5.0	423	28	AQ117654	HS_2188_B1_G02_MF CIT	2.76e+00
	800	15	5.0	424	17	AI126280	qd77f09.x1 Soares_test	2.76e+00
c	801	15	5.0	425	11	AA688298	nv14a07.s1 NCI_CGAP_Pr	2.76e+00
c	802	15	5.0	426	19	N29982	yx82a06.s1 Homo sapien	2.76e+00
c	803	15	5.0	427	26	AQ015543	CIT-HSP-2303J23.TR CIT	2.76e+00
c	804	15	5.0	428	11	AA082994	zn08c04.s1 Stratagene	2.76e+00
c	805	15	5.0	429	27	B80740	CIT-HSP-2050O17.TF CIT	2.76e+00
c	806	15	5.0	429	19	H65421	yu63e09.r1 Homo sapien	2.76e+00
	807	15	5.0	430	26	AQ022838	HS_2183_A2_D03_MR CIT	2.76e+00

c	808	15	5.0	430	22	W95241	ze01d03.r1 Soares feta	2.76e+00
	809	15	5.0	432	24	AA316090	EST187963 HCC cell lin	2.76e+00
c	810	15	5.0	432	28	AQ092514	HS_3003_B2_B09_T7 CIT	2.76e+00
c	811	15	5.0	432	13	AA874124	vw86h07.r1 Stratagene	2.76e+00
	812	15	5.0	434	28	AQ210468	HS_2202_B2_C01_MF CIT	2.76e+00
c	813	15	5.0	434	16	H55761	yq94c04.s1 Homo sapien	2.76e+00
c	814	15	5.0	438	13	AA874963	UI-R-E0-cf-a-07-0-UI.s	2.76e+00
c	815	15	5.0	438	18	AI094287	qa72f04.x1 Soares_feta	2.76e+00
	816	15	5.0	439	8	AA549131	vk74d11.s1 Knowles Sol	2.76e+00
c	817	15	5.0	439	19	H97703	yw07h03.s1 Homo sapien	2.76e+00
	818	15	5.0	440	26	B46009	HS-1062-B2-D04-MR.abi	2.76e+00
c	819	15	5.0	440	15	AI009000	EST203451 Normalized r	2.76e+00
c	820	15	5.0	441	28	AQ176938	HS_3208_A1_C11_MR CIT	2.76e+00
c	821	15	5.0	442	13	AA882142	vx36g06.r1 Stratagene	2.76e+00
c	822	15	5.0	442	10	AA613635	no39e02.s1 NCI_CGAP_Pr	2.76e+00
c	823	15	5.0	443	23	AA220030	mv72a11.r1 Soares mous	2.76e+00
	824	15	5.0	446	16	H11867	ym15f01.r1 Homo sapien	2.76e+00
c	825	15	5.0	447	10	AA635675	nr33b04.s1 NCI_CGAP_Pr	2.76e+00
c	826	15	5.0	448	16	R96376	yq37e06.s1 Homo sapien	2.76e+00
	827	15	5.0	449	15	AI045067	UI-R-C1-kf-b-11-0-UI.s	2.76e+00
	828	15	5.0	449	26	B52010	CIT-HSP-385I6.TF CIT-H	2.76e+00
c	829	15	5.0	449	7	HUM101A03A	Human fetal brain cDNA	2.76e+00
c	830	15	5.0	449	19	N70627	za84d09.s1 Homo sapien	2.76e+00
c	831	15	5.0	450	18	AI168077	oz89c07.x1 Soares_sene	2.76e+00
	832	15	5.0	452	26	B24878	P22J12TR IGF Arabidops	2.76e+00
c	833	15	5.0	452	17	AI083709	qf19b06.x1 NCI_CGAP_Br	2.76e+00
	834	15	5.0	453	26	B38851	HS-1048-A2-H10-MF.abi	2.76e+00
c	835	15	5.0	453	6	AA033600	zk21e09.s1 Soares preg	2.76e+00
c	836	15	5.0	454	18	AI075739	oz30e12.x1 Soares_tota	2.76e+00
c	837	15	5.0	454	21	W19227	zb90a04.s1 Soares sene	2.76e+00
	838	15	5.0	456	26	B46036	HS-1062-B2-F04-MF.abi	2.76e+00
c	839	15	5.0	456	7	AA460530	zx68c09.s1 Soares tota	2.76e+00
c	840	15	5.0	456	12	AA539752	LD18074.5prime LD Dros	2.76e+00
	841	15	5.0	457	5	R18909	yg17c11.r1 Homo sapien	2.76e+00
	842	15	5.0	457	8	AA545360	vj95c07.s1 Knowles Sol	2.76e+00
c	843	15	5.0	459	18	AI174226	an01e01.x1 Stratagene	2.76e+00
c	844	15	5.0	459	28	AQ129652	HS_2255_B1_H02_MR CIT	2.76e+00
	845	15	5.0	460	6	AA431783	zw80f03.s1 Soares test	2.76e+00
c	846	15	5.0	460	26	B41503	HS-1053-B2-E07-MR.abi	2.76e+00
c	847	15	5.0	462	17	AI051340	ow21h06.x1 Soares_para	2.76e+00
c	848	15	5.0	465	10	AA588873	no22h02.s1 NCI_CGAP_Pr	2.76e+00
	849	15	5.0	468	15	C79631	Mus musculus 3.5-dpc b	2.76e+00
	850	15	5.0	469	22	AA211767	zq84a04.r1 Stratagene	2.76e+00
	851	15	5.0	470	20	W75746	me55c08.r1 Soares mous	2.76e+00
c	852	15	5.0	470	27	AQ039836	CIT-HSP-2324A20.TF CIT	2.76e+00
c	853	15	5.0	470	17	AI015856	Mus musculus 2-cell em	2.76e+00
c	854	15	5.0	470	6	AA152184	zl42e08.r1 Soares preg	2.76e+00
c	855	15	5.0	471	14	AA940711	LD22848.5prime LD Dros	2.76e+00
c	856	15	5.0	471	12	AA539392	LD17479.5prime LD Dros	2.76e+00
c	857	15	5.0	471	5	T57671	yc01a11.s1 Homo sapien	2.76e+00
c	858	15	5.0	473	21	T17387	NIB834 Normalized infa	2.76e+00

	859	15	5.0	474	11	T42297	5560 Lambda-PRL2 Arabi	2.76e+00
	860	15	5.0	475	20	W02526	za48e05.r1 Soares feta	2.76e+00
c	861	15	5.0	475	12	AA392561	LD11465.5prime LD Dros	2.76e+00
	862	15	5.0	479	12	AA538981	LD18666.5prime LD Dros	2.76e+00
	863	15	5.0	480	21	AA014419	mi66g12.r1 Soares mous	2.76e+00
	864	15	5.0	481	16	R59330	yh16f11.r1 Homo sapien	2.76e+00
	865	15	5.0	481	12	AA715552	nv53h04.r1 NCI_CGAP_Ew	2.76e+00
	866	15	5.0	482	23	AA145828	mq68d12.r1 Soares 2NbM	2.76e+00
c	867	15	5.0	482	6	AA152268	zl04c04.r1 Soares preg	2.76e+00
c	868	15	5.0	483	11	AA635393	nt68a06.s1 NCI_CGAP_Pr	2.76e+00
c	869	15	5.0	483	26	B29080	T25L13TR TAMU Arabidop	2.76e+00
	870	15	5.0	486	28	AQ132522	HS_3017_A2_B11_MR CIT	2.76e+00
	871	15	5.0	486	8	AA459205	aa25a04.r1 NCI_CGAP_GC	2.76e+00
	872	15	5.0	486	18	AI128753	qa63c07.s1 Soares_feta	2.76e+00
c	873	15	5.0	486	22	AA191045	zq43c01.r1 Stratagene	2.76e+00
	874	15	5.0	487	24	AA275595	vc10h10.r1 Barstead MP	2.76e+00
c	875	15	5.0	487	17	AU021865	Mus musculus unfertili	2.76e+00
	876	15	5.0	488	20	N94182	za26d11.r1 Homo sapien	2.76e+00
c	877	15	5.0	488	17	AA994057	ou37c03.s1 Soares_NFL_	2.76e+00
c	878	15	5.0	488	12	AA438337	LD12291.5prime LD Dros	2.76e+00
	879	15	5.0	491	24	AA310749	EST181558 Jurkat T-cel	2.76e+00
	880	15	5.0	491	13	AA825065	ETN385 Trypanosoma bru	2.76e+00
c	881	15	5.0	491	27	AQ039200	CIT-HSP-233SE17.TF CIT	2.76e+00
	882	15	5.0	492	26	AQ022015	CIT-HSP-2304P19.TF CIT	2.76e+00
c	883	15	5.0	492	28	AQ178559	HS_2242_B1_H01_MR CIT	2.76e+00
c	884	15	5.0	492	8	AA481152	za34f12.s1 NCI_CGAP_GC	2.76e+00
c	885	15	5.0	495	18	AI140118	qa95c10.x1 Soares_feta	2.76e+00
	886	15	5.0	497	8	AA291003	zs45e04.r1 NCI_CGAP_GC	2.76e+00
	887	15	5.0	498	20	W05509	za84d09.r1 Soares feta	2.76e+00
	888	15	5.0	498	27	AQ074728	CIT-HSP-2357D16.TR CIT	2.76e+00
	889	15	5.0	498	15	AA923723	om39c08.s1 Soares_NFL_	2.76e+00
c	890	15	5.0	498	17	AI115024	ui46d06.x1 Sugano mous	2.76e+00
c	891	15	5.0	499	10	AA634345	zu78h04.s1 Soares test	2.76e+00
	892	15	5.0	500	28	AQ121659	HS_3089_A1_F01_MF CIT	2.76e+00
	893	15	5.0	500	7	AA509963	vh53a01.r1 Soares mous	2.76e+00
	894	15	5.0	504	18	AI169581	EST215464 Normalized r	2.76e+00
	895	15	5.0	504	29	AQ240816	CIT-HSP-2387L3.TP.1 CI	2.76e+00
c	896	15	5.0	505	20	N91078	za17h10.r1 Soares feta	2.76e+00
	897	15	5.0	506	14	AA901726	NCM4A4T3 Mycelial Neur	2.76e+00
c	898	15	5.0	506	26	B49907	CIT-HSP-518O2.TV CIT-H	2.76e+00
	899	15	5.0	508	13	AA161218	zo59e07.s1 Stratagene	2.76e+00
c	900	15	5.0	508	14	AA976285	oq67c12.s1 NCI_CGAP_Ki	2.76e+00
	901	15	5.0	510	14	AA967227	vs70a07.r1 Soares mous	2.76e+00
c	902	15	5.0	512	17	AI039157	ox31c10.s1 Soares_tota	2.76e+00
c	903	15	5.0	513	14	AA984318	am83h10.s1 Stratagene	2.76e+00
c	904	15	5.0	514	13	AA859719	UI-R-E0-br-h-06-0-UI.s	2.76e+00
	905	15	5.0	517	23	AA250379	mw16c12.r1 Soares mous	2.76e+00
c	906	15	5.0	519	8	AA438018	vd21f01.s1 Knowles Sol	2.76e+00
c	907	15	5.0	519	16	H24470	ym53f07.s1 Homo sapien	2.76e+00
c	908	15	5.0	522	12	AA781472	aj26h08.s1 Soares test	2.76e+00
	909	15	5.0	527	24	AA285553	vb86c10.r1 Soares mous	2.76e+00

c	910	15	5.0	528	18	AI089534	oz22a12.x1 Soares_tota	2.76e+00
c	911	15	5.0	529	29	AQ148925	HS_3140_B2_H07_T7 CIT	2.76e+00
	912	15	5.0	531	14	AA908118	TENS0960 T. cruzi epim	2.76e+00
c	913	15	5.0	531	14	AA940849	LD23118.5prime LD Dros	2.76e+00
c	914	15	5.0	531	14	AA949835	LD29871.5prime LD Dros	2.76e+00
c	915	15	5.0	531	23	AA209994	mu38e12.r1 Soares 2NbM	2.76e+00
	916	15	5.0	539	11	N65755	20795 Lambda-PRL2 Arab	2.76e+00
c	917	15	5.0	540	10	AA633723	ag87d06.s1 Stratagene	2.76e+00
c	918	15	5.0	541	11	AA704011	ag80a09.r1 Stratagene	2.76e+00
c	919	15	5.0	541	17	AI077243	TENU3468 T. cruzi epim	2.76e+00
	920	15	5.0	542	12	AA141029	CK01299.3prime.contig	2.76e+00
	921	15	5.0	543	14	AA989752	ue16f06.x1 Sugano mous	2.76e+00
c	922	15	5.0	543	16	R88234	ym90b11.r1 Homo sapien	2.76e+00
c	923	15	5.0	545	13	AA883862	am26e04.s1 Soares NFL	2.76e+00
c	924	15	5.0	547	14	C90068	Dictyostelium discoide	2.76e+00
c	925	15	5.0	548	27	AQ081998	RPCI11-54G5.TJ RPCI11	2.76e+00
c	926	15	5.0	549	18	AI185855	qe33g11.s1 Soares_feta	2.76e+00
	927	15	5.0	552	19	N23397	yx83f11.s1 Homo sapien	2.76e+00
c	928	15	5.0	555	14	AA893494	EST197297 Normalized r	2.76e+00
	929	15	5.0	557	12	AA141231	CK01415.5prime CK Dros	2.76e+00
c	930	15	5.0	559	15	AI063548	GH03421.5prime GH Dros	2.76e+00
c	931	15	5.0	560	27	B66374	CIT-HSP-2005G10.TFB CI	2.76e+00
	932	15	5.0	562	26	B18132	A-58B8.TP CIT978SK Hom	2.76e+00
	933	15	5.0	562	15	AI064488	GH04933.3prime GH Dros	2.76e+00
c	934	15	5.0	562	7	AA530369	yj48b06.r1 Stratagene	2.76e+00
	935	15	5.0	563	11	AA651500	31614 Lambda-PRL2 Arab	2.76e+00
	936	15	5.0	566	6	AA115693	zk96g01.s1 Soares preg	2.76e+00
	937	15	5.0	567	5	T64992	yd11f09.s1 Homo sapien	2.76e+00
c	938	15	5.0	568	18	AI178800	EST222482 Normalized r	2.76e+00
	939	15	5.0	569	24	AA308727	EST179536 HCC cell lin	2.76e+00
	940	15	5.0	575	18	AU035983	Mus musculus cDNA, clo	2.76e+00
	941	15	5.0	578	23	AA089182	ml85f04.r1 Stratagene	2.76e+00
	942	15	5.0	578	27	AQ056708	CIT-HSP-2345C4.TF CIT-	2.76e+00
	943	15	5.0	581	19	N52795	yy65d10.s1 Homo sapien	2.76e+00
	944	15	5.0	583	14	AA901689	NCP5A7T3 Perithecial N	2.76e+00
c	945	15	5.0	589	26	B60996	T17M8TR TAMU Arabidops	2.76e+00
c	946	15	5.0	589	8	AA546940	vk67h12.s1 Knowles Sol	2.76e+00
	947	15	5.0	594	14	AA901722	NCM1A4T3 Mycelial Neur	2.76e+00
c	948	15	5.0	594	19	N29998	yx82d06.s1 Homo sapien	2.76e+00
	949	15	5.0	595	14	AA986398	uc82f11.x1 Sugano mous	2.76e+00
c	950	15	5.0	599	17	AI114231	GH11011.5prime GH Dros	2.76e+00
	951	15	5.0	600	11	AA671843	vl02d01.r1 Soares mous	2.76e+00
c	952	15	5.0	611	15	AI064646	GH02862.5prime GH Dros	2.76e+00
	953	15	5.0	612	26	B24816	P22F11TR IGF Arabidops	2.76e+00
	954	15	5.0	612	23	AA140253	mq04g04.r1 Soares 2NbM	2.76e+00
c	955	15	5.0	612	12	AA820521	LD24295.5prime LD Dros	2.76e+00
c	956	15	5.0	612	14	AA950169	LD30111.5prime LD Dros	2.76e+00
c	957	15	5.0	613	26	FR0018578	F.rubripes GSS sequenc	2.76e+00
c	958	15	5.0	615	17	AI113756	GH10205.5prime GH Dros	2.76e+00
c	959	15	5.0	616	27	B95236	CIT-HSP-2171L1.TF CIT-	2.76e+00
c	960	15	5.0	625	28	AQ157007	nbxb0008F08r CUGI Rice	2.76e+00

c	961	15	5.0	627	15	AI062712	GH02015.5prime GH Dros 2.76e+00
	962	15	5.0	631	28	AQ201191	RPCI11-46M14.TK RPCI11 2.76e+00
	963	15	5.0	632	23	AA218437	my02f06.r1 Barstead mo 2.76e+00
c	964	15	5.0	632	26	AQ011291	F26K21TFB IGF Arabidop 2.76e+00
	965	15	5.0	634	26	B24705	F21N10TF IGF Arabidops 2.76e+00
c	966	15	5.0	652	27	BE1471	CIT-HSP-2059M22.TPC CI 2.76e+00
	967	15	5.0	653	28	AQ115028	CIT-HSP-2374E4.TR CIT- 2.76e+00
c	968	15	5.0	661	29	AQ237257	RPCI11-69E24.TJ RPCI11 2.76e+00
	969	15	5.0	664	26	AQ013092	CIT-HSP-2298I13.TF CIT 2.76e+00
	970	15	5.0	665	9	AA596256	vo38f08.r1 Barstead mo 2.76e+00
c	971	15	5.0	665	26	B20432	T22J2-Sp6 TAMU Arabido 2.76e+00
c	972	15	5.0	673	26	AG007262	Homo sapiens genomic D 2.76e+00
	973	15	5.0	676	20	W27606	35b11 Human retina cDN 2.76e+00
c	974	15	5.0	678	28	AQ196694	CIT-HSP-2380J12.TR CIT 2.76e+00
c	975	15	5.0	679	14	AA951868	LD32720.5prime LD Dros 2.76e+00
	976	15	5.0	683	27	AQ082352	RPCI11-50D21.TJ RPCI11 2.76e+00
	977	15	5.0	683	24	AA274639	vc15a10.r1 Barstead MP 2.76e+00
	978	15	5.0	685	28	AF046287	Mus musculus clone OST 2.76e+00
c	979	15	5.0	688	26	AG006362	Homo sapiens genomic D 2.76e+00
c	980	15	5.0	688	29	AG014647	Homo sapiens genomic D 2.76e+00
	981	15	5.0	694	27	AQ049977	nbxb0003aE10r CUGI Ric 2.76e+00
	982	15	5.0	695	29	AQ237266	RPCI11-69G6.TK RPCI11 2.76e+00
	983	15	5.0	695	12	AA753485	96BS0577 Rice Immature 2.76e+00
c	984	15	5.0	696	8	AA126464	zn84d12.s1 Stratagene 2.76e+00
	985	15	5.0	702	28	AQ195905	RPCI11-58C21.TJ RPCI11 2.76e+00
c	986	15	5.0	707	23	AA177410	mt16b12.r1 Soares mous 2.76e+00
c	987	15	5.0	707	27	AQ047278	RPCI11-32K22.TK RPCI11 2.76e+00
	988	15	5.0	711	27	AQ051188	nbxb0005bD05f CUGI Ric 2.76e+00
c	989	15	5.0	721	27	AG012476	Homo sapiens genomic D 2.76e+00
	990	15	5.0	732	27	AG012422	Homo sapiens genomic D 2.76e+00
c	991	15	5.0	744	28	AQ202823	RPCI11-65M16.TK RPCI11 2.76e+00
c	992	15	5.0	778	14	AA952017	LD29023.5prime LD Dros 2.76e+00
c	993	15	5.0	790	14	AA978424	LD30540.5prime LD Dros 2.76e+00
c	994	15	5.0	810	14	AA940709	LD22846.5prime LD Dros 2.76e+00
c	995	15	5.0	814	27	AG010456	Homo sapiens genomic D 2.76e+00
	996	15	5.0	832	26	B04869	cSRL-47G11-u cSRL flow 2.76e+00
c	997	15	5.0	924	7	AA455540	aa16b05.r1 Soares NhHM 2.76e+00
	998	15	5.0	953	26	B08102	T27L1-T7 TAMU Arabidop 2.76e+00
	999	15	5.0	1099	26	B09453	T1K16-T7 TAMU Arabidop 2.76e+00
	1000	15	5.0	1166	26	B12399	F12B6-Sp6 IGF Arabidop 2.76e+00

ALIGNMENTS

```
RESULT    1
LOCUS       HUMGS01155     304 bp    mRNA            EST       17-JUN-1996
DEFINITION  Human HL60 3'directed MboI cDNA, HUMGS01155, clone pm2184, mRNA
            sequence.
ACCESSION   D20181
NID         g501278
KEYWORDS    EST; EST(expressed sequence tag); HL60; gene signature(GS);
```



```
            granulocyte; macrophage; monocyte; promyelocyte.
  SOURCE    Homo sapiens adult female promyelocyte cell_line:HL60 cDNA to mRNA.
    ORGANISM  Homo sapiens
              Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
              Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
              Homo.
  REFERENCE 1  (sites)
    AUTHORS   Okubo,K., Itoh,K., Fukushima,A., Yoshii,J. and Matsubara,K.
    TITLE     Monitoring cell physiology by expression profiles and discovering
              cell type-specific genes by compiled expression profiles
    JOURNAL   Genomics 30 (2), 178-186 (1995)
    MEDLINE   96163870
  REFERENCE 2  (bases 1 to 304)
    AUTHORS   Okubo,K., Fukushima,A., Yoshii,J., Niiyama,T., Kojima,Y.,
              Yoshinari,H., Arimoto,J. and Matsubara,K.
    TITLE     Gene expression of human promyelocytic cell line HL60before and
              after induction of differentiation. A new application of
              3'directedcDNA sequencing
    JOURNAL   Unpublished (1993)
  REFERENCE 3  (bases 1 to 304)
    AUTHORS   Okubo,K.
    TITLE     Direct Submission
    JOURNAL   Submitted (10-SEP-1993) to the DDBJ/EMBL/GenBank databases. Kousaku
              Okubo, Osaka University, Institute for Molecular and Cellular Bio;
              1-3,Yamada-oka, Suita, Osaka Pref. 565, Japan
              (E-mail:kousaku@imcb.osaka-u.ac.jp, Tel:06-877-5111(ex.3315),
              Fax:06-877-1922)
  COMMENT   Submitted (10-Sep-1993) to DDBJ by:
              Kousaku Okubo
              Institute for Molecular and Cellular Biology
              Osaka University
              3-1, Yamadaoka
              Suita, Osaka, 565
              Japan
              Phone: 06-877-5111
              Fax  : 06-875-1922.
  FEATURES             Location/Qualifiers
       source          1..304
                       /organism="Homo sapiens"
                       /db_xref="taxon:9606"
                       /cell_line="HL60"
                       /cell_type="promyelocyte"
                       /dev_stage="adult"
                       /sex="female"
  BASE COUNT       84 a      41 c      67 g     106 t       6 others
  ORIGIN

    Query Match          100.0%;  Score 298;  DB 7;  Length 304;
    Best Local Similarity 100.0%;  Pred. No. 0.00e+00;
    Matches   304;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```



```
Db       1 GATCTAAAGCAGGTNGTGTTGTTTACATGTTTCTACACATTTCATCCTTTAAAAAGTTGT 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy       1 GATCTAAAGCAGGTNGTGTTGTTTACATGTTTCTACACATTTCATCCTTTAAAAAGTTGT 60

Db      61 TGAGAGAGGTTGTATTTACCTTCCCAAGGTTGGAAAGCAGGGGAATTTCCCAGTGTCCTA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      61 TGAGAGAGGTTGTATTTACCTTCCCAAGGTTGGAAAGCAGGGGAATTTCCCAGTGTCCTA 120

Db     121 GTTTTCCACCAGAGGAATATGTGTAAGTAGCAAAGTATTTGCTGCTTACATATAGTGTGT 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     121 GTTTTCCACCAGAGGAATATGTGTAAGTAGCAAAGTATTTGCTGCTTACATATAGTGTGT 180

Db     181 ATGTATGTATATATGTAAATNGTGTGTTAAAGAGCTGATACTGATTTTCATATGNCAATG 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     181 ATGTATGTATATATGTAAATNGTGTGTTAAAGAGCTGATACTGATTTTCATATGNCAATG 240

Db     241 TTAAGGCAAAGGCCTCCCTGCATTTGANGAGCAGGTNTTCATTTATATGTATTTTNGGGA 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     241 TTAAGGCAAAGGCCTCCCTGCATTTGANGAGCAGGTNTTCATTTATATGTATTTTNGGGA 300

Db     301 TAAA 304
           ||||
Qy     301 TAAA 304


RESULT    2
LOCUS       AA463716       387 bp    mRNA            EST       10-JUN-1997
DEFINITION  aa07d11.s1 Soares NhHMPu S1 Homo sapiens cDNA clone 812565 3', mRNA
            sequence.
ACCESSION   AA463716
NID         g2188600
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 387)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J.,
            Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-Merck EST Project 1997
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
```



```
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -41m13 fwd. ET from Amersham
            High quality sequence stop: 345.
  FEATURES             Location/Qualifiers
       source          1..387
                       /organism="Homo sapiens"
                       /note="Organ: mixed (see below); Vector: pT7T3D-Pac
                       (Pharmacia) with a modified polylinker; Site_1: Not I;
                       Site_2: Eco RI; Equal amounts of plasmid DNA from three
                       normalized libraries (melanocyte 2NbHM, pregnant uterus
                       NbHPU, and fetal heart NbHH19W) were mixed, and ss circles
                       were made in vitro. Following HAP purification, this DNA
                       was used as tracer in a subtractive hybridization
                       reaction. The driver was PCR-amplified cDNAs from pools of
                       5,000 clones made from the same 3 libraries. The pools
                       consisted of I.M.A.G.E. clones 260232-265223,
                       340488-345479, and 484488-489479."
                       /db_xref="taxon:9606"
                       /clone="812565"
                       /clone_lib="Soares NhHMPu S1"
                       /tissue_type="Pooled human melanocyte, fetal heart, and
                       pregnant uterus"
                       /lab_host="DH10B"
       mRNA            complement(<1..>387)
                       /db_xref="GDB:6043475"
  BASE COUNT       146 a      77 c      52 g     112 t
  ORIGIN

    Query Match          80.5%;  Score 240;  DB 7;  Length 387;
    Best Local Similarity  98.8%;  Pred. No. 0.00e+00;
    Matches   240;  Conservative    0;  Mismatches   3;  Indels   0;  Gaps   0;

Db      75 TAACATTGTCATATGAAAATCAGTATCAGCTCTTTAACACACAATTTACATATATACATA 134
           |||||||| ||||||||||||||||||||||||||||||||| ||||||||||||||||||
Cp     243 TAACATTGNCATATGAAAATCAGTATCAGCTCTTTAACACACNATTTACATATATACATA 184

Db     135 CATACACACTATATGTAAGCAGCAAATACTTTGCTACTTACACATATTCCTCTGGTGGAA 194
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     183 CATACACACTATATGTAAGCAGCAAATACTTTGCTACTTACACATATTCCTCTGGTGGAA 124

Db     195 AACTAGGACACTGGGAAATTCCCCTGCTTTCCAACCTTGGGAAGGTAAATACAACCTCTC 254
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     123 AACTAGGACACTGGGAAATTCCCCTGCTTTCCAACCTTGGGAAGGTAAATACAACCTCTC 64

Db     255 TCAACAACTTTTTAAAGGATGAAATGTGTAGAAACATGTAAACAACACAACCTGCTTTAG 314
           |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Cp      63 TCAACAACTTTTTAAAGGATGAAATGTGTAGAAACATGTAAACAACACNACCTGCTTTAG 4

Db     315 ATC 317
```

```
          |||
Cp      3 ATC 1

RESULT    3
LOCUS       AA609872      421 bp    mRNA            EST       02-MAR-1998
DEFINITION  af08c06.s1 Soares testis NHT Homo sapiens cDNA clone 1031050 3',
            mRNA sequence.
ACCESSION   AA609872
NID         g2458300
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 421)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
            Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
            Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1117   Std Error: 0.00
            Seq primer: -40ml3 fwd. ET from Amersham
            High quality sequence stop: 351.
FEATURES              Location/Qualifiers
     source           1. .421
                      /organism="Homo sapiens"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                      was prepared from mRNA obtained from Clontech
                      Laboratories, Inc., and primed with  a Not I - oligo(dT)
                      primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3'].
                      Double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Not I and cloned into the Not I
                      and Eco RI sites of the modified pT7T3 vector. Library
                      went through one round of normalization to Cot5, and was
                      constructed  by Bento Soares and M. Fatima Bonaldo."
                      /db_xref="taxon:9606"
                      /clone="1031050"
                      /clone_lib="Soares testis NHT"
```



```
                      /sex="male"
                      /lab_host="DH10B"
BASE COUNT      160 a      80 c      55 g     126 t
ORIGIN

  Query Match           80.5%;  Score 240;  DB 13;  Length 421;
  Best Local Similarity  98.8%;  Pred. No. 0.00e+00;
  Matches   240;  Conservative    0;  Mismatches   3;  Indels   0;  Gaps   0;

Db     86 TAACATTGTCATATGAAAATCAGTATCAGCTCTTTAACACACAATTTACATATATACATA 145
          ||||||| ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Cp    243 TAACATTGNCATATGAAAATCAGTATCAGCTCTTTAACACACNATTTACATATATACATA 184

Db    146 CATACACACTATATGTAAGCAGCAAATACTTTGCTACTTACACATATTCCTCTGGTGGAA 205
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp    183 CATACACACTATATGTAAGCAGCAAATACTTTGCTACTTACACATATTCCTCTGGTGGAA 124

Db    206 AACTAGGACACTGGGAAATTCCCCCTGCTTTCCAACCTTGGGAAGGTAAATACAACCTCTC 265
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp    123 AACTAGGACACTGGGAAATTCCCCCTGCTTTCCAACCTTGGGAAGGTAAATACAACCTCTC 64

Db    266 TCAACAACTTTTTAAAGGATGAAATGTGTAGAAACATGTAAACAACACAACCTGCTTTAG 325
          |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Cp     63 TCAACAACTTTTTAAAGGATGAAATGTGTAGAAACATGTAAACAACACNACCTGCTTTAG 4

Db    326 ATC 328
          |||
Cp      3 ATC 1


RESULT    4
LOCUS       AA496258      377 bp    mRNA            EST       18-AUG-1997
DEFINITION  aa24b10.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:814171 3',
            mRNA sequence.
ACCESSION   AA496258
NID         g2229579
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 377)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
```



```
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
            Ph.D., Gerald Marti, M.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40ml3 fwd. ET from Amersham
            High quality sequence stop: 331.
FEATURES              Location/Qualifiers
     source           1. .377
                      /organism="Homo sapiens"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                      was prepared from human tonsillar cells enriched for
                      germinal center B cells by flow sorting (CD20+, IgD-),
                      provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                      (NCI) and Dr. Gerald Marti (CBER).  cDNA synthesis was
                      primed with a Not I - oligo(dT) primer
                      [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTT-
                      3']. Double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Not I and cloned into the Not I
                      and Eco RI sites of the modified pT7T3 vector. Library
                      went through one round of normalization, and was
                      constructed by Bento Soares and M. Fatima Bonaldo."
                      /db_xref="taxon:9606"
                      /clone="IMAGE:814171"
                      /clone_lib="NCI_CGAP_GCB1"
                      /tissue_type="germinal center B cell"
                      /lab_host="DH10B"
     mRNA             complement(<1. .>377)
                      /db_xref="GDB:6032124"
BASE COUNT      117 a      70 c      44 g     146 t
ORIGIN

  Query Match           71.1%;  Score 212;  DB 8;  Length 377;
  Best Local Similarity  99.1%;  Pred. No. 0.00e+00;
  Matches   212;  Conservative    0;  Mismatches   2;  Indels   0;  Gaps   0;

Db    163 CTCTTTAACACACAATTTACATATATACATACATACACACTATATGTAAGCAGCAAATAC 222
          |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Cp    214 CTCTTTAACACACNATTTACATATATACATACATACACACTATATGTAAGCAGCAAATAC 155

Db    223 TTTGCTACTTACACATATTCCTCTGGTGGAAAACTAGGACACTGGGAAATTCCCCTGCTT 282
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp    154 TTTGCTACTTACACATATTCCTCTGGTGGAAAACTAGGACACTGGGAAATTCCCCTGCTT 95

Db    283 TCCAACCTTGGGAAGGTAAATACAACCTCTCTCAACAACTTTTTAAAGGATGAAATGTGT 342
```



```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     94 TCCAACCTTGGGAAGGTAAATACAACCTCTCTCAACAACTTTTTAAAGGATGAAATGTGT 35

Db    343 AGAAACATGTAAACAACACAACCTGCTTTAGATC 376
          ||||||||||||||||||| ||||||||||||||
Cp     34 AGAAACATGTAAACAACACNACCTGCTTTAGATC 1


RESULT    5
LOCUS       AA760717      474 bp    mRNA            EST       07-FEB-1998
DEFINITION  nz13d07.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1287661,
            mRNA sequence.
ACCESSION   AA760717
NID         g2809647
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 474)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
            Ph.D., Gerald Marti, M.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 2006   Std Error: 0.00
            Seq primer: -40ml3 fwd. ET from Amersham
            High quality sequence stop: 447.
FEATURES              Location/Qualifiers
     source           1. .474
                      /organism="Homo sapiens"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                      was prepared from human tonsillar cells enriched for
                      germinal center B cells by flow sorting (CD20+, IgD-),
                      provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                      (NCI) and Dr. Gerald Marti (CBER).  cDNA synthesis was
                      primed with a Not I - oligo(dT) primer
```

```
                     [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTTT-
                     3']. Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo."
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1287661"
                     /clone_lib="NCI_CGAP_GCB1"
                     /tissue_type="germinal center B cell"
                     /lab_host="DH10B"
BASE COUNT      174 a     85 c     61 g    154 t
ORIGIN

  Query Match           60.7%;  Score 181;  DB 12;  Length 474;
  Best Local Similarity  98.6%;  Pred. No. 0.00e+00;
  Matches   211;  Conservative    0;  Mismatches   3;  Indels   0;  Gaps   0;

Db    145 CTCTTTAACACACAATTTACATATATACATACATACACACTATATGTAAGCAGCGAATAC 204
        ||||||||||||| |||||||||||||||||||||||||||||||||||||||||| |||||
Cp    214 CTCTTTAACACACNATTTACATATATACATACATACACACTATATGTAAGCAGCAAATAC 155

Db    205 TTTGCTACTTACACATATTCCTCTGGTGGAAAACTAGGACACTGGGAAAATTCCCCTGCTT 264
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp    154 TTTGCTACTTACACATATTCCTCTGGTGGAAAACTAGGACACTGGGAAAATTCCCCTGCTT 95

Db    265 TCCAACCTTGGGAAGGTAAATACAACCTCTCTCAACAACTTTTTAAAGGATGAAATGTGT 324
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     94 TCCAACCTTGGGAAGGTAAATACAACCTCTCTCAACAACTTTTTAAAGGATGAAATGTGT 35

Db    325 AGAAACATGTAAACAACACAACCTGCTTTAGATC 358
          |||||||||||||||||| ||||||||||||||
Cp     34 AGAAACATGTAAACAACACNACCTGCTTTAGATC 1


RESULT    6
LOCUS       SSC11G09      443 bp    RNA             EST       29-OCT-1997
DEFINITION  Sus scrofa mRNA; expressed sequence tag (3'; clone c11g09), mRNA
            sequence.
ACCESSION   F22894
NID         g2578461
KEYWORDS    EST; expressed sequence tag.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Artiodactyla; Suiformes; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 443)
  AUTHORS   Winteroe,A.K. and Fredholm,M.
  TITLE     Evaluation and characterization of a porcine small intestine cDNA
            library
  JOURNAL   Unpublished
```



```
REFERENCE   2  (bases 1 to 443)
  AUTHORS   Winteroe,A.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-OCT-1997) Winteroe A.K., The Royal Veterinary and
            Agricultural University, Department of Animal Science and Animal
            Health, Division of Animal Genetics, Bulowsvej 13, 1870
            Frederiksberg C, DENMARK
FEATURES             Location/Qualifiers
     source          1. .443
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /tissue_type="small intestine"
                     /clone_lib="directionally cloned cDNA in XL1-blue MRF'"
                     /clone="c11g09"
     mRNA            1. .>443
                     /note="expressed sequence tag"
BASE COUNT      156 a     93 c     61 g    132 t      1 others
ORIGIN

  Query Match           14.1%;  Score 42;  DB 10;  Length 443;
  Best Local Similarity  93.3%;  Pred. No. 7.84e-54;
  Matches    42;  Conservative    0;  Mismatches   3;  Indels   0;  Gaps   0;

Db     41 ATATAAATGAAAACCNGCTCTTCAAATGCAGGGAGGCCTTTGCCT 85
          |||||||||| ||| |||| |||||||||||||||||||||||||
Cp    288 ATATAAATGAANACCTGCTCNTCAAATGCAGGGAGGCCTTTGCCT 244


RESULT    7
LOCUS       AA273441      354 bp    mRNA            EST       28-MAR-1997
DEFINITION  vc01c11.r1 Soares mouse lymph node NbMLN Mus musculus cDNA clone
            765236 5', mRNA sequence.
ACCESSION   AA273441
NID         g1913436
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Mus.
REFERENCE   1  (bases 1 to 354)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
```



```
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:466156
            Seq primer: -28m13 rev2 ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .354
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGATACTTTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not
                     I and Eco RI sites of the modified pT7T3 vector. RNA
                     provided by Dr. Bertrand Jordan. Library constructed and
                     normalized by Bento Soares and M.Fatima Bonaldo."
                     /db_xref="taxon:10090"
                     /clone="765236"
                     /clone_lib="Soares mouse lymph node NbMLN"
                     /sex="male"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
     mRNA            <1. .>354
BASE COUNT       87 a     69 c     85 g    113 t
ORIGIN

  Query Match           9.1%;  Score 27;  DB 24;  Length 354;
  Best Local Similarity  93.1%;  Pred. No. 4.44e-21;
  Matches    27;  Conservative    0;  Mismatches   2;  Indels   0;  Gaps   0;

Db    316 GAGCAGGTTTTCATTTATATGTATTTTTG 344
          |||||||| |||||||||||||||||| |
Qy    269 GAGCAGGTNTTCATTTATATGTATTTTNG 297


RESULT    8
LOCUS       AA955078      430 bp    mRNA            EST       07-MAY-1998
DEFINITION  UI-R-A1-ey-h-11-0-UI.s1 UI-R-A1 Rattus norvegicus cDNA clone
            UI-R-A1-ey-h-11-0-UI 3' similar to gi|1913436|gb|AA273441|AA273441
            vc01c11.r1 Soares mouse lymph node NbMLN Mus musculus cDNA clone
            765236 5', mRNA sequence.
ACCESSION   AA955078
NID         g3119127
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
```



```
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 430)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT
            Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence tag present in the cDNA between the NotI site and the
            oligo-dT track (not shown) served to identify it as a clone from
            the normalized adult Lung library. cDNA Library Preparation: M.
            Fatima Bonaldo, Ph.D. Clone distribution: clones will be available
            through Research Genetics
            Seq primer: M13 Forward.
FEATURES             Location/Qualifiers
     source          1. .430
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-A1
                     library is a subtracted library derived from the UI-R-A0
                     library. The UI-R-A0 library consisted of a mixture of
                     individually tagged normalized libraries constructed from
                     rat placenta, adult lung, brain, liver, kidney, heart,
                     spleen, ovary, and muscle. The tag is a string of 3-5
                     nucleotides present between the Not I site and the
                     oligo-dT track which allows identification of the library
                     of origin of a clone within the mixture. The subtracted
                     library (UI-R-A1) was constructed as follows: PCR
                     amplified cDNA inserts from a pool of approximately 3,840
                     UI-R-A0 clones from which 3' ESTs had been derived was
                     used as a driver in a hybridization with the UI-R-A0
                     library in the form of single-stranded circles. The
                     remaining single-stranded circles (subtracted library) was
                     purified by hydroxyapatite column chromatography,
                     converted to double-stranded circles and electroporated
                     into DH10B bacteria (Life Technologies) to generate the
                     UI-R-A1 library. This procedure has been previously
                     described (Bonaldo, Lennon and Soares, Genome Research 6:
                     791-806, 1996)"
                     /db_xref="taxon:10116"
                     /clone="UI-R-A1-ey-h-11-0-UI"
                     /clone_lib="UI-R-A1"
```

```
                     /dev_stage="adult"
                     /lab_host="DH10B (Life Technologies)"
BASE COUNT       155 a      95 c      79 g     101 t
ORIGIN

  Query Match           9.1%;  Score 27;  DB 14;  Length 430;
  Best Local Similarity  93.1%;  Pred. No. 4.44e-21;
  Matches    27;  Conservative    0;  Mismatches   2;  Indels   0;  Gaps   0;

Db      47 CAAAAATACATATAAATGAAAACCTGCTC 75
           | ||||||||||||||||||| ||||||||
Cp     297 CNAAAATACATATAAATGAANACCTGCTC 269


RESULT    9
LOCUS       AI008917      449 bp    mRNA            EST       15-JUN-1998
DEFINITION  EST203368 Normalized rat embryo, Bento Soares Rattus sp. cDNA clone
            REMBF37 3' end, mRNA sequence.
ACCESSION   AI008917
NID         g3222749
KEYWORDS    EST.
SOURCE      Rattus sp.
  ORGANISM  Rattus sp.
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 449)
  AUTHORS   Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
            Kerlavage,A.R. and Adams,M.D.
  TITLE     Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
            Gene Index
  JOURNAL   Unpublished (1998)
COMMENT
            Contact: Lee, NH
            ATCC
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
            Seq primer: M13-21.
FEATURES             Location/Qualifiers
     source          1. .449
                     /organism="Rattus sp."
                     /note="Vector: pT7T3Pac; Site_1: EcoRI; Site_2: NotI"
                     /db_xref="taxon:10118"
                     /clone="REMBF37"
                     /clone_lib="Normalized rat embryo, Bento Soares"
                     /dev_stage="embryo 8, 12, 18 dpc"
BASE COUNT       159 a     101 c      93 g      96 t
ORIGIN
```



```
  Query Match           9.1%;  Score 27;  DB 15;  Length 449;
  Best Local Similarity  93.1%;  Pred. No. 4.44e-21;
  Matches    27;  Conservative    0;  Mismatches   2;  Indels   0;  Gaps   0;

Db       6 CAAAAATACATATAAATGAAAACCTGCTC 34
           | ||||||||||||||||||| ||||||||
Cp     297 CNAAAATACATATAAATGAANACCTGCTC 269


RESULT   10
LOCUS       AA730318      363 bp    mRNA            EST       27-JAN-1998
DEFINITION  nw41f01.s1 NCI_CGAP_Ew1 Homo sapiens cDNA clone IMAGE:1249177
            similar to contains Alu repetitive element;contains element MER10
            repetitive element ;, mRNA sequence.
ACCESSION   AA730318
NID         g2751522
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 363)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Lee Helman, M.D., Michael R. Emmert-Buck, M.D.,
            Ph.D.
             cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 744   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 363.
FEATURES             Location/Qualifiers
     source          1. .363
                     /organism="Homo sapiens"
                     /note="Vector: pAMP10; mRNA made from Ewing's sarcoma,
                     cDNA made by oligo-dT priming.  Non-directionally cloned.
                     Size-selected on agarose gel, average insert  size 600 bp.
                     Reference:  Krizman et al. (1996) Cancer Research
                     56:5380-5383."
                     /db_xref="taxon:9606"
```



```
                     /clone="IMAGE:1249177"
                     /clone_lib="NCI_CGAP_Ew1"
                     /tissue_type="Ewing's sarcoma"
                     /lab_host="DH10B"
BASE COUNT       110 a      65 c      67 g     121 t
ORIGIN

  Query Match           7.7%;  Score 23;  DB 12;  Length 363;
  Best Local Similarity  95.8%;  Pred. No. 2.48e-13;
  Matches    23;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db     184 ACAATTTACATATATACATACATA 207
           || |||||||||||||||||||||
Cp     203 ACNATTTACATATATACATACATA 180


RESULT   11
LOCUS       AQ106161      250 bp    DNA             GSS       27-AUG-1998
DEFINITION  HS_3070_A2_C07_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3070 Col=14 Row=E, genomic survey
            sequence.
ACCESSION   AQ106161
NID         g3481517
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 250)
  AUTHORS   Mahairas,G.G., Zackrone,K.D., Smith,T., Tipton,S., Schmidt,S.,
            Traicoff,R., Abajian,C., Blanchard,A., West,A. and Hood,L.E.
  TITLE     Construction of a Characterized Clone Resource for Genomic
            Sequencing: Generation and Preliminary Analysis of 20,000 Sequence
            Tagged Connectors
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3070  row: E  column: 14
            Class: BAC ends
            High quality sequence stop: 250.
FEATURES             Location/Qualifiers
     source          1. .250
                     /organism="Homo sapiens"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
```



```
                     E-Coli DH10B"
                     /db_xref="taxon:9606"
                     /clone="Plate=3070 Col=14 Row=E"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
BASE COUNT       79 a      33 c      37 g     101 t
ORIGIN

  Query Match           7.4%;  Score 22;  DB 28;  Length 250;
  Best Local Similarity 100.0%;  Pred. No. 1.70e-11;
  Matches    22;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     131 TGTGTATGTATGTATATATGTA 152
           ||||||||||||||||||||||
Qy     176 TGTGTATGTATGTATATATGTA 197


RESULT   12
LOCUS       HSC1MG012     323 bp    RNA             EST       21-SEP-1995
DEFINITION  H. sapiens partial cDNA sequence; clone c-img01, mRNA sequence.
ACCESSION   Z39934
NID         g564550
KEYWORDS    EST; partial cDNA sequence; transcribed sequence fragment.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 323)
  AUTHORS   Genexpress.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-OCT-1994) Genethon, B.P. 60, 91002 Evry Cedex France
            and Genetique Moleculaire et Biologie du developpement, CNRS UPR420
            B.P. 8, 94801 Villejuif Cedex France.E-mail: genexpress@genethon.fr
REFERENCE   2  (bases 1 to 323)
  AUTHORS   Genexpress.
  TITLE     The Genexpress cDNA program
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 323)
  AUTHORS   Auffray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C.,
            Devignes,M.D., Duprat,S., Houlgatte,R., Jumeau,M.N., Lamy,B.,
            Lorenzo,F., Mitchell,H., Mariage-Samson,R., Pietu,G., Pouliot,Y.,
            Sebastiani-Kabaktchis,C. and Tessier,A.
  TITLE     IMAGE: molecular integration of the analysis of the human genome
            and its expression
  JOURNAL   C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
  MEDLINE   95277534
COMMENT     Clone library from B.Soares, Psychiatry Dept. Columbia University
            USA;

            Cloning_method: total mRNA was oligo-(dT) primed and directionally
            cloned 5' -> 3' into the HindIII -> NotI sites of the lafmid BA
```

```
            vector;
            Sequencing_method: single read, full automatic;
            Primer: (-21)M13_universal;
            cDNA sequence complementary to mRNA (3'end)
            Stretch_removed: 30 T removed at sequence 5'end
            Normalization_method: Bento Soares, P.N.A.S in press;
            Genexpress_library_idt: C;
            Genexpress_sequence_idt: alc-lmg01;

            No significant homology found with :
            genbank release 81 swissprot release 28.
FEATURES              Location/Qualifiers
     source           1. .323
                      /organism="Homo sapiens"
                      /isolate="muscular atrophy patient"
                      /db_xref="taxon:9606"
                      /dev_stage="3 months old"
                      /tissue_type="total brain"
                      /clone_lib="normalized infant brain cDNA"
                      /sex="Female"
BASE COUNT      111 a     52 c     57 g     99 t      4 others
ORIGIN

  Query Match             7.4%;  Score 22;  DB 19;  Length 323;
  Best Local Similarity 100.0%;  Pred. No. 1.70e-11;
  Matches    22;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    176 TGTATGTATGTATATATGTAAA 197
          |||||||||||||||||||||
Qy    178 TGTATGTATGTATATATGTAAA 199


RESULT   13
LOCUS       AA856996       418 bp    mRNA            EST       09-MAR-1998
DEFINITION  oh94b08.s1 NCI_CGAP_HN4 Homo sapiens cDNA clone IMAGE:1474647 3'
            similar to contains element MER7 repetitive element ;, mRNA
            sequence.
ACCESSION   AA856996
NID         g2945298
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 418)
  AUTHORS   NCI/NIDR-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute / National Institute of Dental Research,
            Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
```



```
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: John Ensley, M.D.
             cDNA Library Preparation: Stratagene, Inc.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 397.
FEATURES              Location/Qualifiers
     source           1. .418
                      /organism="Homo sapiens"
                      /note="Organ: pharynx; Vector: Bluescript SK-; Site_1:
                      EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                      Oligo dT. Average insert size 1.5 kb. 5' adaptor sequence:
                      5' GAATTCGGCACGAG 3' 3' adaptor sequence: 5'
                      (GA)10ACTAGTCTCGAGTTTTTTTTTTTTTTTT 3'"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1474647"
                      /clone_lib="NCI_CGAP_HN4"
                      /tissue_type="squamous cell carcinoma"
                      /lab_host="SOLR (kanamycin resistant)"
BASE COUNT      153 a     67 c     61 g    137 t
ORIGIN

  Query Match             7.4%;  Score 22;  DB 13;  Length 418;
  Best Local Similarity 100.0%;  Pred. No. 1.70e-11;
  Matches    22;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    183 TACATATATACATACATACACA 204
          ||||||||||||||||||||||
Cp    197 TACATATATACATACATACACA 176


RESULT   14
LOCUS       B45656         425 bp    DNA             GSS       20-OCT-1997
DEFINITION  HS-1062-A1-G11-MF.abi CIT Human Genomic Sperm Library C Homo
            sapiens genomic clone Plate=CT 784 Col=21 Row=M, genomic survey
            sequence.
ACCESSION   B45656
NID         g2550490
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 425)
  AUTHORS   Mahairas,G.G., Zackrone,K.D., Smith,T., Tipton,S., Schmidt,S.,
```



```
            Traicoff,R., Abajian,C., Blanchard,A., West,A. and Hood,L.E.
  TITLE     Construction of a Characterized Clone Resource for Genomic
            Sequencing: Generation and Preliminary Analysis of 20,000 Sequence
            Tagged Connectors
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Mahairas GG, Zackrone KD, Hood L
            University of Washington
            Seattle, WA 98195, USA
            Tel: (206) 616-8744
            Fax: (206) 685-7301
            Email: kzackron@u.washington.edu
            Sequence Tagged Connector
            Plate: CT 784  row: M  column: 21
            Class: BAC ends
            High quality sequence stop: 425.
FEATURES              Location/Qualifiers
     source           1. .425
                      /organism="Homo sapiens"
                      /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                      E-Coli DH10B"
                      /db_xref="taxon:9606"
                      /clone="Plate=CT 784 Col=21 Row=M"
                      /clone_lib="CIT Human Genomic Sperm Library C"
                      /sex="M"
BASE COUNT      132 a     68 c     75 g    148 t      2 others
ORIGIN

  Query Match             7.4%;  Score 22;  DB 26;  Length 425;
  Best Local Similarity 100.0%;  Pred. No. 1.70e-11;
  Matches    22;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    388 TTTACATATATACATACATACA 409
          ||||||||||||||||||||||
Cp    199 TTTACATATATACATACATACA 178


RESULT   15
LOCUS       AA043696       484 bp    mRNA            EST       10-MAY-1997
DEFINITION  zk50d10.r1 Soares pregnant uterus NbHPU Homo sapiens cDNA clone
            486259 5', mRNA sequence.
ACCESSION   AA043696
NID         g1521552
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 484)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
```



```
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
            Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 867   Std Error: 0.00
            Seq primer: -28M13 rev2 from Amersham.
FEATURES              Location/Qualifiers
     source           1. .484
                      /organism="Homo sapiens"
                      /note="Organ: uterus; Vector: pT7T3-Pac; Site_1: Not I;
                      Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                      oligo(dT) primer (5'
                      AACTGGAAGAATTCGCGGCCGCCTTTTTTTTTTTTTTTT 3'),
                      double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Not I and cloned into the Not I
                      and Eco RI sites of the modified pT7T3 vector. Library
                      went through one round of normalization.  Library
                      constructed by M. Fatima Bonaldo."
                      /db_xref="taxon:9606"
                      /clone="486259"
                      /clone_lib="Soares pregnant uterus NbHPU"
                      /sex="female"
                      /dev_stage="adult"
                      /lab_host="DH10B"
     mRNA             <1. .>484
                      /db_xref="GDB:3759973"
BASE COUNT      134 a     80 c     90 g    178 t      2 others
ORIGIN

  Query Match             7.4%;  Score 22;  DB 6;  Length 484;
  Best Local Similarity 100.0%;  Pred. No. 1.70e-11;
  Matches    22;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    361 TTTACATATATACATACATACA 382
          ||||||||||||||||||||||
Cp    199 TTTACATATATACATACATACA 178


RESULT   16
LOCUS       R66601         528 bp    mRNA            EST       30-MAY-1995
```

```
DEFINITION  yi32d10.r1 Homo sapiens cDNA clone 140947 5' similar to contains
            MER32 repetitive element ;.
ACCESSION   R66601
NID         g839239
KEYWORDS    EST.
SOURCE      human clone=140947 library=Soares placenta Nb2HP vector=pT7T3D
            (Pharmacia) with a modified polylinker host=DH10B (ampicillin
            resistant) primer=M13RP1 Rsite1=Not I Rsite2=Eco RI Female placenta
            obtained at birth (full term). 1st strand cDNA was primed with a
            Not I - oligo(dT) primer [5'
            AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTTTT 3'], double-stranded
            cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not
            I and cloned into the Not I and Eco RI sites of the modified pT7T3
            vector. Library went through one round of normalization. Library
            constructed by Bento Soares and M.Fatima Bonaldo.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 528)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 296
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES            Location/Qualifiers
     source         1. .528
                    /organism="Homo sapiens"
                    /clone="140947"
BASE COUNT      147 a      88 c     105 g     178 t       9 others
ORIGIN

  Query Match             7.4%;  Score 22;  DB 16;  Length 528;
  Best Local Similarity 100.0%;  Pred. No. 1.70e-11;
  Matches    22;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```



```
Db    328 TGTGTATGTATGTATATATGTA 349
          |||||||||||||||||||||
Qy    176 TGTGTATGTATGTATATATGTA 197


RESULT  17
LOCUS       AA744248      532 bp    mRNA            EST       23-JAN-1998
DEFINITION  ny62g03.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1282900,
            mRNA sequence.
ACCESSION   AA744248
NID         g2784998
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 532)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
            Ph.D., Gerald Marti, M.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 1721  Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 485.
FEATURES            Location/Qualifiers
     source         1. .532
                    /organism="Homo sapiens"
                    /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                    polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                    was prepared from human tonsillar cells enriched for
                    germinal center B cells by flow sorting (CD20+, IgD-),
                    provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                    (NCI) and Dr. Gerald Marti (CBER).  cDNA synthesis was
                    primed with a Not I - oligo(dT) primer
                    [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTTT-
                    3']. Double-stranded cDNA was ligated to Eco RI adaptors
                    (Pharmacia), digested with Not I and cloned into the Not I
```



```
                    and Eco RI sites of the modified pT7T3 vector. Library
                    went through one round of normalization, and was
                    constructed by Bento Soares and M. Fatima Bonaldo."
                    /db_xref="taxon:9606"
                    /clone="IMAGE:1282900"
                    /clone_lib="NCI_CGAP_GCB1"
                    /tissue_type="germinal center B cell"
                    /lab_host="DH10B"
BASE COUNT      195 a      91 c      88 g     158 t
ORIGIN

  Query Match             7.4%;  Score 22;  DB 12;  Length 532;
  Best Local Similarity 100.0%;  Pred. No. 1.70e-11;
  Matches    22;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    458 TGTATGTATGTATATATGTAAA 479
          |||||||||||||||||||||
Qy    178 TGTATGTATGTATATATGTAAA 199


RESULT  18
LOCUS       AA747449      562 bp    mRNA            EST       16-JAN-1998
DEFINITION  nx88h04.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1269367,
            mRNA sequence.
ACCESSION   AA747449
NID         g2787407
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 562)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
            Ph.D., Gerald Marti, M.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40m13 fwd. ET from Amersham
```



```
            High quality sequence stop: 480.
FEATURES            Location/Qualifiers
     source         1. .562
                    /organism="Homo sapiens"
                    /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                    polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                    was prepared from human tonsillar cells enriched for
                    germinal center B cells by flow sorting (CD20+, IgD-),
                    provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                    (NCI) and Dr. Gerald Marti (CBER).  cDNA synthesis was
                    primed with a Not I - oligo(dT) primer
                    [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTTT-
                    3']. Double-stranded cDNA was ligated to Eco RI adaptors
                    (Pharmacia), digested with Not I and cloned into the Not I
                    and Eco RI sites of the modified pT7T3 vector. Library
                    went through one round of normalization, and was
                    constructed by Bento Soares and M. Fatima Bonaldo."
                    /db_xref="taxon:9606"
                    /clone="IMAGE:1269367"
                    /clone_lib="NCI_CGAP_GCB1"
                    /tissue_type="germinal center B cell"
                    /lab_host="DH10B"
BASE COUNT      207 a      96 c      91 g     168 t
ORIGIN

  Query Match             7.4%;  Score 22;  DB 12;  Length 562;
  Best Local Similarity 100.0%;  Pred. No. 1.70e-11;
  Matches    22;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    460 TGTATGTATGTATATATGTAAA 481
          |||||||||||||||||||||
Qy    178 TGTATGTATGTATATATGTAAA 199


RESULT  19
LOCUS       AI147069      355 bp    mRNA            EST       29-SEP-1998
DEFINITION  ok33b12.s1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone
            IMAGE:1509599 3' similar to contains element MER19 repetitive
            element ;, mRNA sequence.
ACCESSION   AI147069
NID         g3674751
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 355)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
```

```
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 268.
FEATURES              Location/Qualifiers
     source           1. .355
                      /organism="Homo sapiens"
                      /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                      a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                      Equal amounts of plasmid DNA from five normalized
                      libraries were mixed, and ss circles were made in vitro.
                      Following HAP purification, this DNA was used as tracer in
                      a subtractive hybridization reaction. The driver was
                      PCR-amplified cDNAs from pools of 5,000 clones made from
                      the same 5 libraries. The pools consisted of the following
                      libraries and cloneIDs: Soares NbHSF pool 1:
                      309384-310919, 323208-325895 Soares Nb2HP pool 1:
                      145032-147335, 147720-148103, 148872-149255, 15002 -
                      150407, 151176-152327 Soares Nb2HF8-9W pool 1:
                      758280-760583, 772104-774407 Soares NbHPA pool 1:
                      304776-306311, 320136-322823, 326280-326663 Soares NbHOT
                      pool 1:  723720-726407, 739080-740999 Subtraction by Bento
                      Soares and M. Fatima Bonaldo."
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1509599"
                      /clone_lib="Soares_NSF_F8_9W_OT_PA_P_S1"
                      /lab_host="DH10B"
BASE COUNT      125 a     67 c     47 g    116 t
ORIGIN

  Query Match           7.0%;  Score 21;  DB 18;  Length 355;
  Best Local Similarity 100.0%;  Pred. No. 1.04e-09;
  Matches    21;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    195 ACATATATACATACATACACA 215
          |||||||||||||||||||||
Cp    196 ACATATATACATACATACACA 176


RESULT   20
LOCUS       AA925765      125 bp    mRNA            EST       21-APR-1998
DEFINITION  UI-R-A1-ep-g-12-0-UI.s1 UI-R-A1 Rattus norvegicus cDNA clone
            UI-R-A1-ep-g-12-0-UI 3', mRNA sequence.
ACCESSION   AA925765
NID         g3072901
KEYWORDS    EST.
SOURCE      Norway rat.
```



```
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 125)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT
            Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence tag present in the cDNA between the NotI site and the
            oligo-dT track (not shown) served to identify it as a clone from
            the normalized adult Lung library. cDNA Library Preparation: M.
            Fatima Bonaldo, Ph.D. Clone distribution: clones will be available
            through Research Genetics
            Seq primer: M13 Forward.
FEATURES              Location/Qualifiers
     source           1. .125
                      /organism="Rattus norvegicus"
                      /strain="Sprague-Dawley"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-A1
                      library is a subtracted library derived from the UI-R-A0
                      library. The UI-R-A0 library  consisted of a mixture of
                      individually tagged normalized libraries constructed from
                      rat placenta, adult lung, brain, liver, kidney, heart,
                      spleen, ovary, and muscle. The tag is a string of 3-5
                      nucleotides present between the Not I site and the
                      oligo-dT track which allows identification of the library
                      of origin of a clone within the mixture. The subtracted
                      library (UI-R-A1) was constructed as follows: PCR
                      amplified cDNA inserts from a pool of approximately 3,840
                      UI-R-A0 clones from which 3' ESTs had been derived was
                      used as a driver in a hybridization with the UI-R-A0
                      library in the form of single-stranded circles. The
                      remaining single-stranded circles (subtracted library) was
                      purified by hydroxyapatite column chromatography,
                      converted to double-stranded circles and electroporated
                      into DH10B bacteria (Life Technologies) to generate the
                      UI-R-A1 library. This procedure has been previously
                      described (Bonaldo, Lennon and Soares, Genome Research 6:
                      791-806, 1996)"
                      /db_xref="taxon:10116"
                      /clone="UI-R-A1-ep-g-12-0-UI"
```



```
                      /clone_lib="UI-R-A1"
                      /dev_stage="adult"
                      /lab_host="DH10B (Life Technologies)"
BASE COUNT       47 a     23 c      9 g     46 t
ORIGIN

  Query Match           6.7%;  Score 20;  DB 14;  Length 125;
  Best Local Similarity 100.0%;  Pred. No. 5.68e-08;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     66 TACATATATACATACATACA 85
          ||||||||||||||||||||
Cp    197 TACATATATACATACATACA 178


RESULT   21
LOCUS       AI102239      134 bp    mRNA            EST       07-OCT-1998
DEFINITION  EST211528 Normalized rat brain, Bento Soares Rattus sp. cDNA clone
            RBRCA93 3' end, mRNA sequence.
ACCESSION   AI102239
NID         g3707041
KEYWORDS    EST.
SOURCE      Rattus sp.
  ORGANISM  Rattus sp.
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 134)
  AUTHORS   Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
            Kerlavage,A.R. and Adams,M.D.
  TITLE     Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
            Gene Index
  JOURNAL   Unpublished (1998)
COMMENT     On Aug 21, 1998 this sequence version replaced.

            Contact: Lee, NH
            ATCC
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
            Seq primer: M13-21.
FEATURES              Location/Qualifiers
     source           1. .134
                      /organism="Rattus sp."
                      /note="Organ: brain; Vector: pT7T3Pac; Site_1: EcoRI;
                      Site_2: NotI"
                      /db_xref="taxon:10118"
                      /clone="RBRCA93"
                      /clone_lib="Normalized rat brain, Bento Soares"
BASE COUNT       28 a     46 c     41 g     19 t
```



```
ORIGIN

  Query Match           6.7%;  Score 20;  DB 18;  Length 134;
  Best Local Similarity 100.0%;  Pred. No. 5.68e-08;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    106 TACATATATACATACATACA 125
          ||||||||||||||||||||
Cp    197 TACATATATACATACATACA 178


RESULT   22
LOCUS       AA943469      239 bp    mRNA            EST       16-JUN-1998
DEFINITION  EST198968 Normalized rat brain, Bento Soares Rattus sp. cDNA clone
            RBRAK23 3' end, mRNA sequence.
ACCESSION   AA943469
NID         g3103385
KEYWORDS    EST.
SOURCE      Rattus sp.
  ORGANISM  Rattus sp.
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 239)
  AUTHORS   Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
            Kerlavage,A.R. and Adams,M.D.
  TITLE     Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
            Gene Index
  JOURNAL   Unpublished (1998)
COMMENT
            Contact: Lee, NH
            ATCC
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
            Seq primer: M13-21.
FEATURES              Location/Qualifiers
     source           1. .239
                      /organism="Rattus sp."
                      /note="Organ: brain; Vector: pT7T3Pac; Site_1: EcoRI;
                      Site_2: NotI"
                      /db_xref="ATCC (inhost):2012985"
                      /db_xref="taxon:10118"
                      /clone="RBRAK23"
                      /clone_lib="Normalized rat brain, Bento Soares"
BASE COUNT       72 a     45 c     51 g     71 t
ORIGIN

  Query Match           6.7%;  Score 20;  DB 15;  Length 239;
  Best Local Similarity 100.0%;  Pred. No. 5.68e-08;
```

```
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      60 TACATATATACATACATACA 79
           ||||||||||||||||||||
Cp     197 TACATATATACATACATACA 178


RESULT   23
LOCUS       AA522005      272 bp    mRNA            EST       17-JUL-1997
DEFINITION  vh78a09.r1 Knowles Solter mouse E6 5d whole embryo Mus musculus
            cDNA clone 893080 3', mRNA sequence.
ACCESSION   AA522005
NID         g2262750
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae;
            Murinae; Mus.
REFERENCE   1  (bases 1 to 272)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:521040.
FEATURES             Location/Qualifiers
     source          1. .272
                     /organism="Mus musculus"
                     /strain="B6D2 F1/J"
                     /note="Vector: pBluescribe (modified); Site_1: MluI;
                     Site_2: SalI; Cloned unidirectionally from mRNA prepared
                     from 13,500 2-cell stage  embryos.  Primer: SalI(dT):
                     5'-CGGTCGACCGTCGACCGTTTTTTTTTTTTTTT-3'.   cDNAs were
                     cloned into the MluI/SalI sites of a modified pBluescribe
                     vector using commercial linkers (NEB)."
                     /db_xref="taxon:10090"
                     /clone="893080"
                     /clone_lib="Knowles Solter mouse E6 5d whole embryo"
```



```
                     /dev_stage="embryo (post-implantation)"
                     /lab_host="DH10B"
     mRNA            complement(<1. .>272)
BASE COUNT       66 a      50 c      56 g     100 t
ORIGIN

  Query Match           6.7%;  Score 20;  DB 7;  Length 272;
  Best Local Similarity 100.0%;  Pred. No. 5.68e-08;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      98 GTGTGTATGTATGTATATAT 117
           ||||||||||||||||||||
Qy     175 GTGTGTATGTATGTATATAT 194


RESULT   24
LOCUS       AI011490      272 bp    mRNA            EST       15-JUN-1998
DEFINITION  EST205941 Normalized rat ovary, Bento Soares Rattus sp. cDNA clone
            ROVAV64 3' end, mRNA sequence.
ACCESSION   AI011490
NID         g3225322
KEYWORDS    EST.
SOURCE      Rattus sp.
  ORGANISM  Rattus sp.
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 272)
  AUTHORS   Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
            Kerlavage,A.R. and Adams,M.D.
  TITLE     Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
            Gene Index
  JOURNAL   Unpublished (1998)
COMMENT
            Contact: Lee, NH
            ATCC
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
            Seq primer: M13-21.
FEATURES             Location/Qualifiers
     source          1. .272
                     /organism="Rattus sp."
                     /note="Organ: ovary; Vector: pT7T3Pac; Site_1: EcoRI;
                     Site_2: NotI"
                     /db_xref="taxon:10118"
                     /clone="ROVAV64"
                     /clone_lib="Normalized rat ovary, Bento Soares"
BASE COUNT       86 a      50 c      57 g      79 t
ORIGIN
```



```
  Query Match           6.7%;  Score 20;  DB 15;  Length 272;
  Best Local Similarity 100.0%;  Pred. No. 5.68e-08;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      60 TACATATATACATACATACA 79
           ||||||||||||||||||||
Cp     197 TACATATATACATACATACA 178


RESULT   25
LOCUS       W18871        293 bp    mRNA            EST       10-SEP-1996
DEFINITION  mc04d01.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone 337921
            5', mRNA sequence.
ACCESSION   W18871
NID         g1294581
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Mus.
REFERENCE   1  (bases 1 to 293)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:219321
            Seq primer: mob.REGA+ET
            High quality sequence stop: 286.
FEATURES             Location/Qualifiers
     source          1. .293
                     /organism="Mus musculus"
                     /note="Vector: pT7T3D (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
```



```
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by Bento
                     Soares and M.Fatima Bonaldo. RNA was kindly provided by
                     Dr. Minoru Ko (Wayne State University)."
                     /db_xref="taxon:10090"
                     /clone="337921"
                     /clone_lib="Soares mouse p3NMF19.5"
                     /dev_stage="19.5 dpc total fetus"
                     /lab_host="DH10B (ampicillin resistant)"
     mRNA            <1. .>293
BASE COUNT       81 a      56 c      65 g      91 t
ORIGIN

  Query Match           6.7%;  Score 20;  DB 21;  Length 293;
  Best Local Similarity 100.0%;  Pred. No. 5.68e-08;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     258 TGTATGTATGTATATATGTA 277
           ||||||||||||||||||||
Qy     178 TGTATGTATGTATATATGTA 197


RESULT   26
LOCUS       AI072793      312 bp    mRNA            EST       05-AUG-1998
DEFINITION  UI-R-Y0-lw-b-06-0-UI.s1 UI-R-Y0 Rattus norvegicus cDNA clone
            UI-R-Y0-lw-b-06-0-UI 3', mRNA sequence.
ACCESSION   AI072793
NID         g3398987
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 312)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT
            Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            cDNA Library Preparation: M. Fatima Bonaldo, Ph.D. Clone
            distribution: clones will be available through Research Genetics
```

```
            The following repetitive elements were found in this cDNA sequence:
            20-234, B3#SINE/B2
            Seq primer: M13 Forward.
FEATURES                Location/Qualifiers
     source             1. .312
                        /organism="Rattus norvegicus"
                        /strain="Sprague-Dawley"
                        /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                        polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-Y0
                        library is a subtracted library derived from an
                        individually-tagged normalized whole-eye (minus the lens)
                        library. The driver for the subtraction consisted of a
                        pool of all previous libraries (UI-R-A0, UI-R-A1, UI-R-E0,
                        UI-R-E1, UI-R-C0, and UI-R-C1). The tag is a string of
                        3-5 nucleotides present between the Not I site and the
                        oligo-dT track which allows identification of the library
                        of origin of a clone within the mixture. The subtracted
                        library (UI-R-Y0) was constructed as follows: PCR
                        amplified cDNA inserts from previous library clones from
                        which 3' ESTs had  been derived were used as a driver in a
                        hybridization with the normalized  whole-eye library in
                        the form of single-stranded circles. The remaining
                        single-stranded circles (subtracted library) was purified
                        by hydroxyapatite column chromatography, converted to
                        double-stranded circles and electroporated into DH10B
                        bacteria (Life Technologies) to generate the UI-R-Y0
                        library. This procedure has been previously described
                        (Bonaldo, Lennon and Soares, Genome Research 6: 791-806,
                        1996)"
                        /db_xref="taxon:10116"
                        /clone="UI-R-Y0-lw-b-06-0-UI"
                        /clone_lib="UI-R-Y0"
                        /dev_stage="adult"
                        /lab_host="DH10B (Life Technologies)"
BASE COUNT       90 a      59 c      61 g     102 t
ORIGIN

  Query Match             6.7%;  Score 20;  DB 17;  Length 312;
  Best Local Similarity 100.0%;  Pred. No. 5.68e-08;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     241 TGTATGTATGTATATATGTA 260
           ||||||||||||||||||||
Qy     178 TGTATGTATGTATATATGTA 197


RESULT   27
LOCUS       R09708          330 bp    mRNA            EST       05-APR-1995
DEFINITION  yf23e10.s1 Homo sapiens cDNA clone 127722 3' similar to contains
            Alu repetitive element;.
ACCESSION   R09708
```



```
NID         g761631
KEYWORDS    EST.
SOURCE      human clone=127722 library=Soares fetal liver spleen 1NFLS
            vector=pT7T3D (Pharmacia) with a modified polylinker host=DH10B
            (ampicillin resistant) primer=-21m13 Rsite1=Pac I Rsite2=Eco RI
            Liver and spleen from a 20 week-post conception male fetus. 1st
            strand cDNA was primed with a Pac I - oligo(dT) primer [5'
            AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTT 3'], double-stranded
            cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac
            I and cloned into the Pac I and Eco RI sites of the modified pT7T3
            vector.  Library went through one round of normalization. Library
            constructed by Bento Soares and M.Fatima Bonaldo.
  ORGANISM  Homo sapiens
            Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 330)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High qality sequence stops: 293
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES                Location/Qualifiers
     source             1. .330
                        /organism="Homo sapiens"
                        /clone="127722"
BASE COUNT       70 a      81 c      83 g      93 t       3 others
ORIGIN

  Query Match             6.7%;  Score 20;  DB 5;  Length 330;
  Best Local Similarity 100.0%;  Pred. No. 5.68e-08;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     166 TGTATGTATGTATATATGTA 185
           ||||||||||||||||||||
Qy     178 TGTATGTATGTATATATGTA 197
```



```
RESULT   28
LOCUS       AA293650        331 bp    mRNA            EST       16-MAY-1997
DEFINITION  zt55d12.r1 Soares ovary tumor NbHOT Homo sapiens cDNA clone 726263
            5' similar to contains L1.t1 L1 repetitive element ;, mRNA
            sequence.
ACCESSION   AA293650
NID         g1941556
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 331)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J.,
            Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-Merck EST Project 1997
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 329.
FEATURES                Location/Qualifiers
     source             1. .331
                        /organism="Homo sapiens"
                        /note="Organ: ovary; Vector: pT7T3D (Pharmacia) with a
                        modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                        strand cDNA was primed with a Not I - oligo(dT) primer [5'
                        TGTTACCAATCTGAAGTGGGAGGCGGCCGCGGTTTTTTTTTTTTTTTT 3'],
                        double-stranded cDNA was size selected, ligated to Eco RI
                        adapters (Pharmacia), digested with Not I and cloned into
                        the Not I and Eco RI sites of a modified pT7T3 vector
                        (Pharmacia). Library constructed by Bento Soares and
                        M.Fatima Bonaldo."
                        /db_xref="taxon:9606"
                        /clone="726263"
                        /clone_lib="Soares ovary tumor NbHOT"
                        /sex="Female"
                        /tissue_type="ovarian tumor"
                        /lab_host="DH10B (ampicillin resistant)"
     mRNA               <1. .>331
```



```
                        /db_xref="GDB:5938341"
BASE COUNT      128 a      47 c      73 g      83 t
ORIGIN

  Query Match             6.7%;  Score 20;  DB 6;  Length 331;
  Best Local Similarity 100.0%;  Pred. No. 5.68e-08;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     104 ATATATACATACATACACAC 123
           ||||||||||||||||||||
Cp     194 ATATATACATACATACACAC 175


RESULT   29
LOCUS       AA934050        339 bp    mRNA            EST       23-JUN-1998
DEFINITION  on78a12.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
            IMAGE:1562782 3', mRNA sequence.
ACCESSION   AA934050
NID         g3090318
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 339)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 596   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 326.
FEATURES                Location/Qualifiers
     source             1. .339
                        /organism="Homo sapiens"
                        /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                        a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                        Equal amounts of plasmid DNA from three normalized
                        libraries (fetal lung NbHL19W, testis NHT, and B-cell
                        NCI_CGAP_GCB1) were mixed, and ss circles were made in
                        vitro. Following HAP purification, this DNA was used as
                        tracer in a subtractive hybridization reaction. The driver
                        was PCR-amplified cDNAs from pools of 5,000 clones made
                        from the same 3 libraries. The pools consisted of
                        I.M.A.G.E. clones 297480-302087, 682632-687239,
```

```
                     726408-728711, and 729096-731399. Subtraction by Bento
                     Soares and M. Fatima Bonaldo."
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1562782"
                     /clone_lib="Soares_NFL_T_GBC_S1"
                     /lab_host="DH10B"
BASE COUNT       99 a      86 c      60 g      94 t
ORIGIN

  Query Match           6.7%;  Score 20;  DB 15;  Length 339;
  Best Local Similarity 100.0%;  Pred. No. 5.68e-08;
  Matches   20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     274 ATATATACATACATACACAC 293
           ||||||||||||||||||||
Cp     194 ATATATACATACATACACAC 175


RESULT   30
LOCUS       AI060034      354 bp    mRNA            EST       20-JUL-1998
DEFINITION  UI-R-C1-kz-a-11-0-UI.s1 UI-R-C1 Rattus norvegicus cDNA clone
            UI-R-C1-kz-a-11-0-UI 3', mRNA sequence.
ACCESSION   AI060034
NID         g3333811
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 354)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT
            Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence tag present in the cDNA between the NotI site and the
            oligo-dT track served to identify it as a clone from the normalized
            adult Lung library. cDNA Library Preparation: M. Fatima Bonaldo,
            Ph.D. Clone distribution: clones will be available through Research
            Genetics
            Seq primer: M13 Forward.
FEATURES             Location/Qualifiers
     source          1. .354
```



```
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-C1
                     library is a subtracted library derived from the UI-R-C0
                     library, which is a subtracted library derived from the
                     UI-R-A1 and UI-R-E1 libraries. The UI-R-A1 library
                     consisted of a mixture of  individually tagged normalized
                     libraries constructed from rat placenta,  adult lung,
                     brain, liver, kidney, heart, spleen, ovary, and muscle.
                     The UI-R-E1 library  consisted of a mixture of
                     individually tagged normalized libraries constructed from
                     8, 12 and 18-day embryo. The tag is a string of 3-5
                     nucleotides present between the Not I site and the
                     oligo-dT track which allows identification of the library
                     of origin of a clone within the mixture. The subtracted
                     library (UI-R-C1) was constructed as follows: PCR
                     amplified cDNA inserts from  UI-R-C0 clones from which 3'
                     ESTs had been derived was used as a driver in a
                     hybridization with the UI-R-C0 library in the form of
                     single-stranded circles. The remaining single-stranded
                     circles (subtracted library) was purified by
                     hydroxyapatite column chromatography, converted to
                     double-stranded circles and electroporated into DH10B
                     bacteria (Life Technologies) to generate the UI-R-C1
                     library. This procedure has been previously described
                     (Bonaldo, Lennon and Soares, Genome Research 6: 791-806,
                     1996)"
                     /db_xref="taxon:10116"
                     /clone="UI-R-C1-kz-a-11-0-UI"
                     /clone_lib="UI-R-C1"
                     /dev_stage="Adult"
                     /lab_host="DH10B (Life Technologies)"
BASE COUNT      114 a      56 c      58 g     126 t
ORIGIN

  Query Match           6.7%;  Score 20;  DB 15;  Length 354;
  Best Local Similarity 100.0%;  Pred. No. 5.68e-08;
  Matches   20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     274 ATATATACATACATACACAC 293
           ||||||||||||||||||||
Cp     194 ATATATACATACATACACAC 175


RESULT   31
LOCUS       AQ095488      370 bp    DNA             GSS       26-AUG-1998
DEFINITION  HS_3012_B1_B10_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3012 Col=19 Row=D, genomic survey
            sequence.
ACCESSION   AQ095488
```



```
NID         g3463223
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 370)
  AUTHORS   Mahairas,G.G., Zackrone,K.D., Smith,T., Tipton,S., Schmidt,S.,
            Traicoff,R., Abajian,C., Blanchard,A., West,A. and Hood,L.E.
  TITLE     Construction of a Characterized Clone Resource for Genomic
            Sequencing: Generation and Preliminary Analysis of 20,000 Sequence
            Tagged Connectors
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3012  row: D  column: 19
            Class: BAC ends
            High quality sequence stop: 370.
FEATURES             Location/Qualifiers
     source          1. .370
                     /organism="Homo sapiens"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
                     /db_xref="taxon:9606"
                     /clone="Plate=3012 Col=19 Row=D"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
BASE COUNT      103 a      78 c      70 g     119 t
ORIGIN

  Query Match           6.7%;  Score 20;  DB 28;  Length 370;
  Best Local Similarity 100.0%;  Pred. No. 5.68e-08;
  Matches   20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     123 TGTGTTAAAGAGCTGATACT 142
           ||||||||||||||||||||
Qy     203 TGTGTTAAAGAGCTGATACT 222


RESULT   32
LOCUS       AA672177      397 bp    mRNA            EST       25-NOV-1997
DEFINITION  vk90a10.s1 Knowles Solter mouse 2 cell Mus musculus cDNA clone
            961914 5', mRNA sequence.
ACCESSION   AA672177
```



```
NID         g2644394
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 397)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:550706
            High quality sequence stop: 355.
FEATURES             Location/Qualifiers
     source          1. .397
                     /organism="Mus musculus"
                     /strain="B6D2 F1/J"
                     /note="Organ: embryo; Vector: pBluescribe (modified);
                     Site_1: MluI; Site_2: SalI; Cloned unidirectionally from
                     mRNA prepared from 13,500 2-cell stage  embryos.  Primer:
                     SalI(dT): 5'-CGGTCGACCGTCGACCGTTTTTTTTTTTTTT-3',  cDNAs
                     were cloned into the MluI/SalI sites of a modified
                     pBluescribe  vector using commercial linkers (NEB).
                     Average insert size:  1.2 kb."
                     /db_xref="taxon:10090"
                     /clone="961914"
                     /clone_lib="Knowles Solter mouse 2 cell"
                     /tissue_type="embryo"
                     /dev_stage="2-cell"
                     /lab_host="DH10B"
BASE COUNT      107 a      51 c      94 g     145 t
ORIGIN

  Query Match           6.7%;  Score 20;  DB 11;  Length 397;
  Best Local Similarity 100.0%;  Pred. No. 5.68e-08;
  Matches   20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     282 TGTATGTATGTATATATGTA 301
```

```
          ||||||||||||||||||||
Qy    178 TGTATGTATGTATATATGTA 197


RESULT   33
LOCUS       AQ211566      403 bp    DNA              GSS       17-SEP-1998
DEFINITION  HS_3235_B2_F01_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3235 Col=2 Row=L, genomic survey
            sequence.
ACCESSION   AQ211566
NID         g3622767
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 403)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Furlong,J.,
            Shaker,R., Schmidt,S., Traicoff,R. and Hood,L.E.
  TITLE     Construction of a Characterized Clone Resource for Genomic
            Sequencing
  JOURNAL   Unpublished (1998)
COMMENT
            Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3235  row: L  column: 2
            Class: BAC ends
            High quality sequence stop: 403.
FEATURES             Location/Qualifiers
     source          1. .403
                     /organism="Homo sapiens"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
                     /db_xref="taxon:9606"
                     /clone="Plate=3235 Col=2 Row=L"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
BASE COUNT      112 a     90 c     88 g    113 t
ORIGIN

  Query Match           6.7%;  Score 20;  DB 28;  Length 403;
  Best Local Similarity 100.0%;  Pred. No. 5.68e-08;
  Matches   20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    118 TACATATATACATACATACA 137
```



```
          ||||||||||||||||||||
Cp    197 TACATATATACATACATACA 178


RESULT   34
LOCUS       H31634        406 bp    mRNA             EST       02-APR-1998
DEFINITION  EST105880 Rat PC-12 cells, untreated Rattus sp. cDNA clone RPCAV05
            3' end, mRNA sequence.
ACCESSION   H31634
NID         g977051
KEYWORDS    EST.
SOURCE      Rattus sp.
  ORGANISM  Rattus sp.
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 406)
  AUTHORS   Lee,N.H., Weinstock,K.G., Kirkness,E.F., Earle-Hughes,J.A.,
            Fuldner,R.A., Marmaras,S., Glodek,A., Gocayne,J.D., Adams,M.D.,
            Kerlavage,A.R., Fraser,C.M. and Venter,J.C.
  TITLE     Comparative expressed-sequence-tag analysis of differential gene
            expression profiles in PC-12 cells before and after nerve growth
            factor treatment
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 92, 8303-8307 (1995)
  MEDLINE   95396786
COMMENT     Other_ESTs: EST105881
            Contact: Lee, NH
            ATCC
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
            For clone availability please contact  the TIGR Database
            (tdbinfo@tdb.tigr.org)
            Seq primer: M13 - 21.
FEATURES             Location/Qualifiers
     source          1. .406
                     /organism="Rattus sp."
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; poly(A)+ RNA was purified from untreated PC12 cells
                     cultured for 9 days.  cDNA was constructed using an
                     oligo-dT primer and directionally cloned using the Lambda
                     ZAP II Vector Kit by Stratagene"
                     /db_xref="ATCC (inhost):2000854"
                     /db_xref="taxon:10118"
                     /clone="RPCAV05"
                     /clone_lib="Rat PC-12 cells, untreated"
BASE COUNT      114 a     84 c     92 g    113 t      3 others
ORIGIN

  Query Match           6.7%;  Score 20;  DB 14;  Length 406;
```



```
  Best Local Similarity 100.0%;  Pred. No. 5.68e-08;
  Matches   20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     68 TACATATATACATACATACA 87
          ||||||||||||||||||||
Cp    197 TACATATATACATACATACA 178


RESULT   35
LOCUS       AA590910      416 bp    mRNA             EST       16-SEP-1997
DEFINITION  vm21e04.r1 Knowles Solter mouse blastocyst B1 Mus musculus cDNA
            clone 990846 5', mRNA sequence.
ACCESSION   AA590910
NID         g2404223
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 416)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:563126
            High quality sequence stop: 384.
FEATURES             Location/Qualifiers
     source          1. .416
                     /organism="Mus musculus"
                     /strain="B6D2 F1/J"
                     /note="Organ: embryo; Vector: pSPORT; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally from mRNA prepared
                     from 800 blastocysts.  Primer: SalI(dT):
                     5'-CGGTCGACCGTCGACCGTTTTTTTTTTTTTTT-3'.  cDNAs  were
                     cloned into the NotI/SalI sites of a pSPORT vector (Life
                     Technologies).  Two different size selections:  B1 (larger
                     inserts) and B3."
                     /db_xref="taxon:10090"
```



```
                     /clone="990846"
                     /clone_lib="Knowles Solter mouse blastocyst B1"
                     /tissue_type="blastocyst"
                     /dev_stage="embryo (pre-implantation)"
                     /lab_host="DH10B"
     mRNA            <1. .>416
BASE COUNT      150 a     84 c     64 g    118 t
ORIGIN

  Query Match           6.7%;  Score 20;  DB 9;  Length 416;
  Best Local Similarity 100.0%;  Pred. No. 5.68e-08;
  Matches   20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    384 TACATATATACATACATACA 403
          ||||||||||||||||||||
Cp    197 TACATATATACATACATACA 178


RESULT   36
LOCUS       AI052454      417 bp    mRNA             EST       29-SEP-1998
DEFINITION  oz07g11.x1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA
            clone IMAGE:1674692 3' similar to contains Alu repetitive element;,
            mRNA sequence.
ACCESSION   AI052454
NID         g3308445
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 417)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 636  Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 333.
FEATURES             Location/Qualifiers
     source          1. .417
                     /organism="Homo sapiens"
                     /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                     with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                     This is a subtracted version of the original Soares fetal
                     liver spleen 1NFLS library.  1st strand cDNA was primed
```

```
                     with a Pac I - oligo(dT) primer [5'
                     AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Pac I and cloned into the Pac I
                     and Eco RI sites of the modified pT7T3 vector.  Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1674692"
                     /clone_lib="Soares_fetal_liver_spleen_1NFLS_S1"
                     /sex="male"
                     /dev_stage="20 week-post conception fetus"
                     /lab_host="DH10B (ampicillin resistant)"
BASE COUNT      104 a     109 c      92 g     112 t
ORIGIN

  Query Match             6.7%;  Score 20;  DB 18;  Length 417;
  Best Local Similarity 100.0%;  Pred. No. 5.68e-08;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    164 TGTATGTATGTATATATGTA 183
          ||||||||||||||||||||
Qy    178 TGTATGTATGTATATATGTA 197


RESULT   37
LOCUS       AQ148379      424 bp    DNA             GSS       07-OCT-1998
DEFINITION  HS_2211_A2_A12_MF CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=2211 Col=24 Row=A, genomic survey
            sequence.
ACCESSION   AQ148379
NID         g3541837
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 424)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Furlong,J.,
            Shaker,R., Schmidt,S., Traicoff,R. and Hood,L.E.
  TITLE     Construction of a Characterized Clone Resource for Genomic
            Sequencing
  JOURNAL   Unpublished (1998)
COMMENT
            Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
```



```
            Sequence Tagged Connector
            Plate: 2211  row: A  column: 24
            Class: BAC ends
            High quality sequence stop: 424.
FEATURES              Location/Qualifiers
     source           1. .424
                      /organism="Homo sapiens"
                      /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                      E-Coli DH10B"
                      /db_xref="taxon:9606"
                      /clone="Plate=2211 Col=24 Row=A"
                      /clone_lib="CIT Approved Human Genomic Sperm Library D"
                      /sex="male"
BASE COUNT      158 a      72 c      50 g     138 t      6 others
ORIGIN

  Query Match             6.7%;  Score 20;  DB 29;  Length 424;
  Best Local Similarity 100.0%;  Pred. No. 5.68e-08;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    126 ATATATACATACATACACAC 145
          ||||||||||||||||||||
Cp    194 ATATATACATACATACACAC 175


RESULT   38
LOCUS       AI101287      445 bp    mRNA            EST       06-OCT-1998
DEFINITION  EST210576 Normalized rat brain, Bento Soares Rattus sp. cDNA clone
            RBRBK80 3' end, mRNA sequence.
ACCESSION   AI101287
NID         g3704624
KEYWORDS    EST.
SOURCE      Rattus sp.
  ORGANISM  Rattus sp.
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 445)
  AUTHORS   Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
            Kerlavage,A.R. and Adams,M.D.
  TITLE     Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
            Gene Index
  JOURNAL   Unpublished (1998)
COMMENT
            Contact: Lee, NH
            ATCC
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
            Seq primer: M13-21.
```



```
FEATURES              Location/Qualifiers
     source           1. .445
                      /organism="Rattus sp."
                      /note="Organ: brain; Vector: pT7T3Pac; Site_1: EcoRI;
                      Site_2: NotI"
                      /db_xref="taxon:10118"
                      /clone="RBRBK80"
                      /clone_lib="Normalized rat brain, Bento Soares"
BASE COUNT      177 a      76 c      66 g     126 t
ORIGIN

  Query Match             6.7%;  Score 20;  DB 18;  Length 445;
  Best Local Similarity 100.0%;  Pred. No. 5.68e-08;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    160 ATATATACATACATACACAC 179
          ||||||||||||||||||||
Cp    194 ATATATACATACATACACAC 175


RESULT   39
LOCUS       AI173314      447 bp    mRNA            EST       07-OCT-1998
DEFINITION  uc29c11.r1 Soares mouse mammary gland NbMMG Mus musculus cDNA clone
            1399412 5', mRNA sequence.
ACCESSION   AI173314
NID         g3720894
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 447)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,P., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:911128
            Seq primer: -28m13 rev2 ET from Amersham
```



```
            High quality sequence stop: 360.
FEATURES              Location/Qualifiers
     source           1. .447
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /note="Organ: mammary gland; Vector: pT7T3D-Pac
                      (Pharmacia) with a modified polylinker; Site_1: Not I;
                      Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                      oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATGGTTTTTTTTTTTTTTTTTTTTT
                      T 3']; double-stranded cDNA was ligated to Eco RI
                      adaptors (Pharmacia), digested with Not I and cloned into
                      the Not I and Eco RI sites of the modified pT7T3 vector.
                      RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
                      constructed and normalized by Bento Soares and M.Fatima
                      Bonaldo."
                      /db_xref="taxon:10090"
                      /clone="1399412"
                      /clone_lib="Soares mouse mammary gland NbMMG"
                      /sex="male"
                      /tissue_type="mammary gland"
                      /dev_stage="4 weeks"
                      /lab_host="DH10B"
BASE COUNT      111 a      66 c      95 g     175 t
ORIGIN

  Query Match             6.7%;  Score 20;  DB 18;  Length 447;
  Best Local Similarity 100.0%;  Pred. No. 5.68e-08;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    211 TACATATATACATACATACA 230
          ||||||||||||||||||||
Cp    197 TACATATATACATACATACA 178


RESULT   40
LOCUS       AA438117      449 bp    mRNA            EST       04-AUG-1997
DEFINITION  vd22f02.s1 Knowles Solter mouse 2 cell Mus musculus cDNA clone
            793275 5', mRNA sequence.
ACCESSION   AA438117
NID         g2143031
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae;
            Murinae; Mus.
REFERENCE   1  (bases 1 to 449)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
```

```
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:477683
            Seq primer: -40ml3 fwd. ET from Amersham
            High quality sequence stop: 390.
FEATURES             Location/Qualifiers
     source          1. .449
                     /organism="Mus musculus"
                     /strain="B6D2 F1/J"
                     /note="Organ: embryo; Vector: pBluescribe (modified);
                     Site_1: MluI; Site_2: SalI; Cloned unidirectionally from
                     mRNA prepared from 13,500 2-cell stage embryos. Primer:
                     SalI(dT): 5'-CGGTCGACCGTCGACCGTTTTTTTTTTTTTT-3'. cDNAs
                     were cloned into the MluI/SalI sites of a modified
                     pBluescribe vector using commercial linkers (NEB).
                     Average insert size: 1.2 kb."
                     /db_xref="taxon:10090"
                     /clone="793275"
                     /clone_lib="Knowles Solter mouse 2 cell"
                     /tissue_type="embryo"
                     /dev_stage="2-cell"
                     /lab_host="DH10B"
     mRNA            <1. .>449
BASE COUNT      123 a     81 c     73 g    172 t
ORIGIN

  Query Match           6.7%;  Score 20;  DB 8;  Length 449;
  Best Local Similarity 100.0%;  Pred. No. 5.68e-08;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    304 GTGTGTATGTATGTATATAT 323
          ||||||||||||||||||||
Qy    175 GTGTGTATGTATGTATATAT 194


RESULT   41
LOCUS       AQ222916      455 bp    DNA              GSS       18-SEP-1998
DEFINITION  HS_2013_A1_E01_T7 CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=2013 Col=1 Row=I, genomic survey
```



```
            sequence.
ACCESSION   AQ222916
NID         g3636529
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 455)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Furlong,J.,
            Shaker,R., Schmidt,S., Traicoff,R. and Hood,L.E.
  TITLE     Construction of a Characterized Clone Resource for Genomic
            Sequencing
  JOURNAL   Unpublished (1998)
COMMENT
            Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 2013  row: I  column: 1
            Class: BAC ends
            High quality sequence stop: 455.
FEATURES             Location/Qualifiers
     source          1. .455
                     /organism="Homo sapiens"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
                     /db_xref="taxon:9606"
                     /clone="Plate=2013 Col=1 Row=I"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
BASE COUNT      165 a     84 c    111 g     94 t      1 others
ORIGIN

  Query Match           6.7%;  Score 20;  DB 28;  Length 455;
  Best Local Similarity 100.0%;  Pred. No. 5.68e-08;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    320 TACATATATACATACATACA 339
          ||||||||||||||||||||
Cp    197 TACATATATACATACATACA 178


RESULT   42
LOCUS       B74942        467 bp    DNA              GSS       06-JUN-1998
DEFINITION  RPCI11-11P8.TP RPCI11 Homo sapiens genomic clone R-11P8, genomic
            survey sequence.
```



```
ACCESSION   B74942
NID         g2770629
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 467)
  AUTHORS   Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K.,
            Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences for Sequence-Ready Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: RPCI11-11P8.TV
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            For clone availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .467
                     /organism="Homo sapiens"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
                     /db_xref="taxon:9606"
                     /clone="R-11P8"
                     /clone_lib="RPCI11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
BASE COUNT      171 a     81 c     46 g    169 t
ORIGIN

  Query Match           6.7%;  Score 20;  DB 26;  Length 467;
  Best Local Similarity 100.0%;  Pred. No. 5.68e-08;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    196 GTGTGTATGTATGTATATAT 215
          ||||||||||||||||||||
Qy    175 GTGTGTATGTATGTATATAT 194


RESULT   43
LOCUS       AI070519      499 bp    mRNA             EST       05-AUG-1998
```



```
DEFINITION  UI-R-Y0-1v-f-07-0-UI.s1 UI-R-Y0 Rattus norvegicus cDNA clone
            UI-R-Y0-1v-f-07-0-UI 3', mRNA sequence.
ACCESSION   AI070519
NID         g3396770
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 499)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT
            Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            cDNA Library Preparation: M. Fatima Bonaldo, Ph.D. Clone
            distribution: clones will be available through Research Genetics
            Seq primer: M13 Forward.
FEATURES             Location/Qualifiers
     source          1. .499
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-Y0
                     library is a subtracted library derived from an
                     individually-tagged normalized whole-eye (minus the lens)
                     library. The driver for the subtraction consisted of a
                     pool of all previous libraries (UI-R-A0, UI-R-A1, UI-R-E0,
                      UI-R-E1, UI-R-C0, and UI-R-C1). The tag is a string of
                     3-5 nucleotides present between the Not I site and the
                     oligo-dT track which allows identification of the library
                     of origin of a clone within the mixture. The subtracted
                     library (UI-R-Y0) was constructed as follows: PCR
                     amplified cDNA inserts from previous library clones from
                     which 3' ESTs had been derived were used as a driver in a
                     hybridization with the normalized whole-eye library in
                     the form of single-stranded circles. The remaining
                     single-stranded circles (subtracted library) was purified
                     by hydroxyapatite column chromatography, converted to
                     double-stranded circles and electroporated into DH10B
                     bacteria (Life Technologies) to generate the UI-R-Y0
                     library. This procedure has been previously described
                     (Bonaldo, Lennon and Soares, Genome Research 6: 791-806,
```

```
                    1996)"
                    /db_xref="taxon:10116"
                    /clone="UI-R-Y0-lv-f-07-0-UI"
                    /clone_lib="UI-R-Y0"
                    /dev_stage="adult"
                    /lab_host="DH10B (Life Technologies)"
BASE COUNT      133 a    110 c    108 g    148 t
ORIGIN

  Query Match           6.7%;  Score 20;  DB 17;  Length 499;
  Best Local Similarity 100.0%;  Pred. No. 5.68e-08;
  Matches   20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      83 TACATATATACATACATACA 102
           ||||||||||||||||||||
Cp     197 TACATATATACATACATACA 178


RESULT   44
LOCUS       AI007920      510 bp    mRNA            EST       15-JUN-1998
DEFINITION  EST202371 Normalized rat brain, Bento Soares Rattus sp. cDNA clone
            RBRAZ11 3' end, mRNA sequence.
ACCESSION   AI007920
NID         g3221752
KEYWORDS    EST.
SOURCE      Rattus sp.
  ORGANISM  Rattus sp.
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 510)
  AUTHORS   Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
            Kerlavage,A.R. and Adams,M.D.
  TITLE     Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
            Gene Index
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Lee, NH
            ATCC
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
            Seq primer: M13-21.
FEATURES             Location/Qualifiers
     source          1. .510
                     /organism="Rattus sp."
                     /note="Organ: brain; Vector: pT7T3Pac; Site_1: EcoRI;
                     Site_2: NotI"
                     /db_xref="taxon:10118"
                     /clone="RBRAZ11"
```



```
                    /clone_lib="Normalized rat brain, Bento Soares"
BASE COUNT      138 a    117 c    117 g    138 t
ORIGIN

  Query Match           6.7%;  Score 20;  DB 15;  Length 510;
  Best Local Similarity 100.0%;  Pred. No. 5.68e-08;
  Matches   20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      60 TACATATATACATACATACA 79
           ||||||||||||||||||||
Cp     197 TACATATATACATACATACA 178


RESULT   45
LOCUS       AQ184311      522 bp    DNA             GSS       11-SEP-1998
DEFINITION  HS_3226_A1_D12_T7 CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3226 Col=23 Row=G, genomic survey
            sequence.
ACCESSION   AQ184311
NID         g3582923
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 522)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Furlong,J. and
            Shaker,R.
  TITLE     Construction of a Characterized Clone Resource for Genomic
            Sequencing
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3226  row: G  column: 23
            Class: BAC ends
            High quality sequence stop: 522.
FEATURES             Location/Qualifiers
     source          1. .522
                     /organism="Homo sapiens"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
                     /db_xref="taxon:9606"
                     /clone="Plate=3226 Col=23 Row=G"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
```



```
                    /sex="male"
BASE COUNT      179 a     86 c     80 g    169 t      8 others
ORIGIN

  Query Match           6.7%;  Score 20;  DB 28;  Length 522;
  Best Local Similarity 100.0%;  Pred. No. 5.68e-08;
  Matches   20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     214 ACATATATACATACATACAC 233
           ||||||||||||||||||||
Cp     196 ACATATATACATACATACAC 177


RESULT   46
LOCUS       AQ141570      523 bp    DNA             GSS       23-SEP-1998
DEFINITION  HS_3154_A1_G08_T7 CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3154 Col=15 Row=M, genomic survey
            sequence.
ACCESSION   AQ141570
NID         g3532223
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 523)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Furlong,J.,
            Shaker,R., Schmidt,S., Traicoff,R. and Hood,L.E.
  TITLE     Construction of a Characterized Clone Resource for Genomic
            Sequencing
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3154  row: M  column: 15
            Class: BAC ends
            High quality sequence stop: 523.
FEATURES             Location/Qualifiers
     source          1. .523
                     /organism="Homo sapiens"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
                     /db_xref="taxon:9606"
                     /clone="Plate=3154 Col=15 Row=M"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
```



```
                    /sex="male"
BASE COUNT      163 a     95 c     95 g    166 t      4 others
ORIGIN

  Query Match           6.7%;  Score 20;  DB 29;  Length 523;
  Best Local Similarity 100.0%;  Pred. No. 5.68e-08;
  Matches   20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     487 GTGTGTATGTATGTATATAT 506
           ||||||||||||||||||||
Qy     175 GTGTGTATGTATGTATATAT 194


RESULT   47
LOCUS       AQ122664      544 bp    DNA             GSS       21-SEP-1998
DEFINITION  HS_3086_A2_H02_MP CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3086 Col=4 Row=O, genomic survey
            sequence.
ACCESSION   AQ122664
NID         g3499830
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 544)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Furlong,J.,
            Shaker,R., Schmidt,S., Traicoff,R. and Hood,L.E.
  TITLE     Construction of a Characterized Clone Resource for Genomic
            Sequencing
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3086  row: O  column: 4
            Class: BAC ends
            High quality sequence stop: 544.
FEATURES             Location/Qualifiers
     source          1. .544
                     /organism="Homo sapiens"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
                     /db_xref="taxon:9606"
                     /clone="Plate=3086 Col=4 Row=O"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
```

```
                     /sex="male"
BASE COUNT      142 a    118 c     96 g    176 t     12 others
ORIGIN

  Query Match             6.7%;  Score 20;  DB 28;  Length 544;
  Best Local Similarity 100.0%;  Pred. No. 5.68e-08;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    329 TGTATGTATGTATATATGTA 348
          ||||||||||||||||||||
Qy    178 TGTATGTATGTATATATGTA 197


RESULT   48
LOCUS       AQ197092      587 bp    DNA             GSS       15-SEP-1998
DEFINITION  CIT-HSP-2384I15.TF CIT-HSP Homo sapiens genomic clone 2384I15,
            genomic survey sequence.
ACCESSION   AQ197092
NID         g3604454
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 587)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
            Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
            Venter,J.C.
  TITLE     Use of a random human BAC End Sequence Database for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1998)
COMMENT     Other_GSSs: CIT-HSP-2384I15.TR
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13-21
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .587
                     /organism="Homo sapiens"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII"
                     /db_xref="taxon:9606"
                     /clone="2384I15"
```



```
                     /clone_lib="CIT-HSP"
                     /sex="Male"
                     /cell_type="Sperm"
BASE COUNT      197 a    112 c    128 g    149 t      1 others
ORIGIN

  Query Match             6.7%;  Score 20;  DB 28;  Length 587;
  Best Local Similarity 100.0%;  Pred. No. 5.68e-08;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    215 TACATATATACATACATACA 234
          ||||||||||||||||||||
Cp    197 TACATATATACATACATACA 178


RESULT   49
LOCUS       AU024235      592 bp    mRNA            EST       04-AUG-1998
DEFINITION  Mus musculus unfertilized egg cDNA 3'-end sequence, clone J0449A10,
            mRNA sequence.
ACCESSION   AU024235
NID         g3394582
KEYWORDS    EST; EST(expressed sequence tag).
SOURCE      Mus musculus (strain:C57BL/6J) unfertilized egg cDNA to mRNA,
            clone:J0449A10.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 592)
  AUTHORS   Doi,H. and Ko,M.S.
  TITLE     Systematic analyses of genes expressed in unfertilized mouse eggs
            (The ERATO/Doi Project at Wayne State University)
  JOURNAL   Published Only in DataBase (1998) In press
REFERENCE   2  (bases 1 to 592)
  AUTHORS   Doi,H. and Ko,M.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUN-1998) to the DDBJ/EMBL/GenBank databases.
            Hirofumi Doi, Japan Science and Technology Corporation (JST), Doi
            Bicasymmetry Project, ERATO; WBG Marive East 12F, 2-6 Nakase,
            Mihama-ku, Chiba, Chiba 261-7112, Japan (E-mail:hd@bioa.jst.go.jp,
            Tel:+81-43-299-1351, Fax:+81-43-297-7530)
FEATURES             Location/Qualifiers
     source          1. .592
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="J0449A10"
                     /dev_stage="unfertilized egg"
BASE COUNT      186 a     65 c    110 g    223 t      8 others
ORIGIN

  Query Match             6.7%;  Score 20;  DB 17;  Length 592;
```



```
  Best Local Similarity 100.0%;  Pred. No. 5.68e-08;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    487 GTATGTATGTATATATGTAA 506
          ||||||||||||||||||||
Qy    179 GTATGTATGTATATATGTAA 198


RESULT   50
LOCUS       AQ195348      606 bp    DNA             GSS       15-SEP-1998
DEFINITION  RPCI11-61L17.TJ RPCI11 Homo sapiens genomic clone R-61L17, genomic
            survey sequence.
ACCESSION   AQ195348
NID         g3606960
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 606)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
            Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
  TITLE     Use of human BAC End Sequences for Sequence-Ready Map Building
  JOURNAL   Unpublished (1998)
COMMENT
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            For clone availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .606
                     /organism="Homo sapiens"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
                     /db_xref="taxon:9606"
                     /clone="R-61L17"
                     /clone_lib="RPCI11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
BASE COUNT      213 a     95 c    107 g    191 t
ORIGIN

  Query Match             6.7%;  Score 20;  DB 28;  Length 606;
```



```
  Best Local Similarity 100.0%;  Pred. No. 5.68e-08;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    542 TACATATATACATACATACA 561
          ||||||||||||||||||||
Cp    197 TACATATATACATACATACA 178


Search completed: Tue Dec 22 20:27:59 1998
Job time : 546 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Tue Dec 22 20:28:20 1998; MasPar time 57.70 Seconds
716.756 Million cell updates/sec

Tabular output not generated.

Title: >US-08-530-112A-1017
Description: (1-304) from US08530112A.seq
Perfect Score: 298
N.A. Sequence: 1 GATCTAAAGCAGGTNGTGTT.....ATATGTATTTTNGGGATAAA 304
Comp: CTAGATTTCGTCCANCACAA.....TATACATAAAANCCCTATTT

Scoring table: TABLE jmetric
Gap 60

Nmatch STD : Dbase 0; Query 0

Searched: 188442 seqs, 68026449 bases x 2

Post-processing: Minimum Match 0%
Listing first 1000 summaries

Database: n-geneseq32
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39 40:part40

Statistics: Mean 6.722; Variance 2.861; scale 2.349

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
	1	298	100.0	304	19	T20017	Human gene signature	5.87e-269
c	2	20	6.7	264	29	T65661	Polymorphic repeat se	1.55e-02
c	3	20	6.7	5511	24	T41853	cDNA encoding Plasmod	1.55e-02
c	4	20	6.7	9789	24	T41852	cDNA encoding Plasmod	1.55e-02
c	5	19	6.4	128	20	T23920	Human gene signature	7.80e-02
	6	18	6.0	122	20	T21516	Human gene signature	3.77e-01
	7	18	6.0	2891	10	Q57016	PKC delta.	3.77e-01
	8	18	6.0	2909	1	Q01758	cDNA sequence of rat	3.77e-01
	9	18	6.0	3095	6	Q38525	Human CENP-C antigen	3.77e-01
c	10	18	6.0	3959	4	Q24856	Sequence of T-cadheri	3.77e-01
	11	18	6.0	21126	2	N60970	TL-DNA region of A.rh	3.77e-01
c	12	17	5.7	830	33	T62455	Lambda 5h-1 clone con	1.74e+00
c	13	17	5.7	2542	39	V15094	Human apolipoprotein(	1.74e+00
c	14	17	5.7	5837	7	Q43966	Thermococcus litorali	1.74e+00
c	15	17	5.7	5837	7	Q43515	T. litoralis DNA poly	1.74e+00
c	16	17	5.7	5837	11	Q66293	Vent DNA polymerase g	1.74e+00
c	17	17	5.7	5837	3	Q14461	Polymerase gene.	1.74e+00
c	18	17	5.7	5837	18	T17546	T. litoralis DNA poly	1.74e+00
c	19	17	5.7	5837	31	T70812	Vent DNA polymerase g	1.74e+00
	20	16	5.4	51	5	Q34071	Microsatellite sequen	7.71e+00
	21	16	5.4	518	10	Q55238	Grapevine ribosomal c	7.71e+00
c	22	16	5.4	1260	30	T72049	Sequence around marke	7.71e+00
c	23	16	5.4	1260	30	T43929	Sequence around marke	7.71e+00
	24	16	5.4	2544	30	T67066	Human alpha3(IX) coll	7.71e+00
c	25	16	5.4	2563	8	Q50008	Aspergillus niger Sul	7.71e+00
	26	16	5.4	3224	38	T97393	Aspergillus oryzae pr	7.71e+00
c	27	16	5.4	19124	32	T72882	Plasmodium var-7 gene	7.71e+00
	28	15	5.0	118	39	V13059	Interleukin 6 non-ris	3.23e+01
	29	15	5.0	152	23	T26837	Human gene signature	3.23e+01
	30	15	5.0	154	39	V13060	Interleukin 6 non-ris	3.23e+01
c	31	15	5.0	175	23	T26811	Human gene signature	3.23e+01
c	32	15	5.0	226	21	T24352	Human gene signature	3.23e+01
	33	15	5.0	246	10	Q77415	Human genome fragment	3.23e+01
	34	15	5.0	327	11	Q77289	Human genome fragment	3.23e+01
	35	15	5.0	337	20	T23229	Human gene signature	3.23e+01
	36	15	5.0	342	8	Q60579	Human brain Expressed	3.23e+01
	37	15	5.0	354	8	Q59745	Human brain Expressed	3.23e+01
	38	15	5.0	396	33	T84999	Type II topoisomerase	3.23e+01
	39	15	5.0	850	8	Q48592	Autonomously replicat	3.23e+01
c	40	15	5.0	900	17	T01542	Retinoblastoma protei	3.23e+01
	41	15	5.0	1014	38	V00413	Internal fragment of	3.23e+01
	42	15	5.0	1014	38	T91294	Human G52-24 secreted	3.23e+01

	43	15	5.0	1478	23	T37066	Tomato green fruit-sp	3.23e+01
	44	15	5.0	1560	38	T91295	Human G52-24 secreted	3.23e+01
	45	15	5.0	1633	17	T15608	FGF receptor ligand F	3.23e+01
c	46	15	5.0	1697	17	T01541	Retinoblastoma protei	3.23e+01
c	47	15	5.0	1698	29	T62390	Antisense sequence to	3.23e+01
c	48	15	5.0	1802	29	T62391	Direct tandem repeat	3.23e+01
	49	15	5.0	2037	13	Q78149	Large subunit of ADP-	3.23e+01
c	50	15	5.0	2125	16	Q91762	Fission yeast Rad51 g	3.23e+01
	51	15	5.0	2293	35	T89279	TPRP-F1 gene promoter	3.23e+01
	52	15	5.0	2437	17	Q97381	Human type II steroid	3.23e+01
	53	15	5.0	2437	37	T99802	Human genital tissue	3.23e+01
c	54	15	5.0	2634	29	T62394	Construct pMK10/AS14.	3.23e+01
c	55	15	5.0	2975	39	V15529	Human lymphoid-specif	3.23e+01
c	56	15	5.0	3240	39	V15530	Human lymphoid-specif	3.23e+01
c	57	15	5.0	3265	1	N92188	TA29 gene insert in p	3.23e+01
c	58	15	5.0	3499	29	T62393	Construct pMK10/AS3'U	3.23e+01
c	59	15	5.0	3504	31	T62498	Nematocidal toxin gen	3.23e+01
c	60	15	5.0	3504	28	T61018	Bacillus thuringiensi	3.23e+01
c	61	15	5.0	3504	37	T89185	Nematode toxin 167P g	3.23e+01
c	62	15	5.0	3603	29	T62396	Construct pMK10/AS14X	3.23e+01
	63	15	5.0	3754	37	V01891	Human OTK18 gene.	3.23e+01
c	64	15	5.0	3771	31	T68924	Drosophila deltex cDN	3.23e+01
c	65	15	5.0	3771	17	Q92980	Deltex cDNA.	3.23e+01
	66	15	5.0	4320	35	T89278	TPRP-F1 gene, includi	3.23e+01
	67	15	5.0	4656	1	N80747	Sequence of 2A11 geno	3.23e+01
c	68	15	5.0	4656	1	N80747	Sequence of 2A11 geno	3.23e+01
	69	15	5.0	4747	20	T09019	Arabidopsis thaliana	3.23e+01
c	70	15	5.0	5063	17	Q92979	Deltex locus composit	3.23e+01
	71	15	5.0	5125	5	Q29703	IRS-1.	3.23e+01
c	72	15	5.0	5309	2	Q10280	Complete sequence of	3.23e+01
	73	15	5.0	6042	20	T09018	Arabidopsis thaliana	3.23e+01
	74	15	5.0	6727	37	T88014	Murine IL-5 cDNA geno	3.23e+01
	75	15	5.0	7920	20	T05899	Human IL-4 gene.	3.23e+01
	76	15	5.0	7997	34	T79737	Rat cholesterol 7-alp	3.23e+01
	77	15	5.0	7997	15	Q89553	Rat cholesterol 7 alp	3.23e+01
	78	15	5.0	8010	18	T13413	Human papillomavirus	3.23e+01
	79	15	5.0	10614	15	Q89555	Hamster cholesterol 7	3.23e+01
	80	15	5.0	19012	6	Q36952	HSA gene.	3.23e+01
	81	15	5.0	20303	31	T71699	Human deoxycytidylate	3.23e+01
	82	15	5.0	26764	31	T71696	Human deoxycytidylate	3.23e+01
	83	14	4.7	34	29	T63579	PCR primer used in CD	1.28e+02
	84	14	4.7	73	20	T23480	Human gene signature	1.28e+02
c	85	14	4.7	85	5	Q33624	Microsatellite sequen	1.28e+02
	86	14	4.7	92	38	T96836	Intron 6 of human pRb	1.28e+02
	87	14	4.7	97	14	Q81902	Interferon-gamma bind	1.28e+02
	88	14	4.7	132	34	T67321	Human alpha-signalin	1.28e+02
	89	14	4.7	180	18	T19280	Human gene signature	1.28e+02
	90	14	4.7	193	19	T21954	Human gene signature	1.28e+02
	91	14	4.7	196	19	T20188	Human gene signature	1.28e+02
	92	14	4.7	210	19	T20162	Human gene signature	1.28e+02
c	93	14	4.7	231	19	T21931	Human gene signature	1.28e+02

	94	14	4.7	261	21	T26276	Human gene signature	1.28e+02
c	95	14	4.7	310	8	Q61110	Human brain Expressed	1.28e+02
c	96	14	4.7	318	38	T88839	Alpha light chain nuc	1.28e+02
	97	14	4.7	342	7	Q40891	T-BAM gene sequence f	1.28e+02
c	98	14	4.7	374	40	V09128	5' nucleotide sequenc	1.28e+02
	99	14	4.7	396	16	Q85528	DNA probe 34 detects	1.28e+02
c	100	14	4.7	405	8	Q59912	Human brain Expressed	1.28e+02
c	101	14	4.7	449	38	T91255	Human papilloma virus	1.28e+02
	102	14	4.7	507	18	T19486	Human gene signature	1.28e+02
	103	14	4.7	530	1	N90314	Hepatitis C virus (HC	1.28e+02
	104	14	4.7	531	1	N92084	Sequence of the hepat	1.28e+02
c	105	14	4.7	560	1	Q06235	Bovine male-specific	1.28e+02
c	106	14	4.7	560	2	N60553	Fragment of plasmid p	1.28e+02
	107	14	4.7	626	18	T19066	Human gene signature	1.28e+02
	108	14	4.7	718	3	N92332	Sequence of a ribosom	1.28e+02
c	109	14	4.7	750	31	T43445	ATM gene exons 4-5.	1.28e+02
	110	14	4.7	793	3	N92333	Sequence of a ribosom	1.28e+02
	111	14	4.7	796	3	N92334	Sequence of a ribosom	1.28e+02
	112	14	4.7	871	3	N92336	Sequence of a ribosom	1.28e+02
	113	14	4.7	916	6	Q40689	Insertion sequence IS	1.28e+02
	114	14	4.7	1021	3	N92335	Sequence of a ribosom	1.28e+02
	115	14	4.7	1023	8	Q48575	HPV E6/7 Genomic regi	1.28e+02
c	116	14	4.7	1216	38	T93320	Tomato-infecting gemi	1.28e+02
	117	14	4.7	1228	36	T94471	Human Fchd605 gene di	1.28e+02
c	118	14	4.7	1234	12	Q70415	Full length pre-pro-v	1.28e+02
c	119	14	4.7	1293	8	Q46244	B18R gene.	1.28e+02
c	120	14	4.7	1293	4	Q25097	Vaccinia virus B18R g	1.28e+02
	121	14	4.7	1365	32	T61098	DNA encoding porcine	1.28e+02
c	122	14	4.7	1389	24	T40120	Human papillomavirus	1.28e+02
	123	14	4.7	1414	4	Q27190	Tumour necrosis facto	1.28e+02
c	124	14	4.7	1487	9	Q51544	Lipopolysaccharide in	1.28e+02
	125	14	4.7	1491	22	T30153	Trabecular meshwork i	1.28e+02
	126	14	4.7	1512	2	N60962	Sequence of human pre	1.28e+02
	127	14	4.7	1560	14	Q83201	Precursor of novel am	1.28e+02
c	128	14	4.7	1566	1	Q06689	5' UTR of ELAM1 from	1.28e+02
	129	14	4.7	1584	24	T35525	Human interferon-beta	1.28e+02
c	130	14	4.7	1605	24	T37310	Aromatic acyl transfe	1.28e+02
	131	14	4.7	1631	30	T63603	Dehiscence zone-selec	1.28e+02
	132	14	4.7	1648	33	T62311	Rat amelin-2 gene.	1.28e+02
	133	14	4.7	1657	24	T33994	Oilseed rape seed pod	1.28e+02
	134	14	4.7	1691	2	Q11684	Clone encoding human	1.28e+02
	135	14	4.7	1692	40	V04268	Lectin-like antigen c	1.28e+02
	136	14	4.7	1692	29	T47749	Hodgkin's disease ant	1.28e+02
c	137	14	4.7	1696	13	Q75819	Non-A Non-B hepatitis	1.28e+02
c	138	14	4.7	1811	22	T28820	Stem cell marker cDNA	1.28e+02
	139	14	4.7	1825	26	T45951	Rat 25-hydroxyvitamin	1.28e+02
	140	14	4.7	1836	4	N20057	Interferon-beta gene.	1.28e+02
c	141	14	4.7	1850	22	T28819	Stem cell marker cDNA	1.28e+02
	142	14	4.7	1939	33	T62310	Rat amelin-1 gene.	1.28e+02
c	143	14	4.7	1968	19	T17375	Bcl-Y cDNA.	1.28e+02
	144	14	4.7	1969	22	T30152	Trabecular meshwork i	1.28e+02

c	145	14	4.7	1969	6	Q34553	Herbicide sensitive A 1.28e+02
	146	14	4.7	1976	30	T47438	Penicillium chrysogen 1.28e+02
c	147	14	4.7	2064	28	T37087	ElAF matrix metallopr 1.28e+02
c	148	14	4.7	2070	14	Q82873	Human ST30 sialyltran 1.28e+02
c	149	14	4.7	2072	15	Q95492	Human Cdn-1 cDNA. 1.28e+02
c	150	14	4.7	2073	9	Q55149	Adenovirus E1A-F gene 1.28e+02
	151	14	4.7	2090	1	N90368	Fragment of tobacco P 1.28e+02
	152	14	4.7	2092	1	Q04091	2.1 kb XbaI-BamHI fra 1.28e+02
c	153	14	4.7	2094	25	T42138	Bak gene. 1.28e+02
	154	14	4.7	2126	7	Q41229	Clone GP14 encoding m 1.28e+02
	155	14	4.7	2190	7	Q41225	Clone PC14 encoding m 1.28e+02
	156	14	4.7	2207	39	V00500	Human alpha-(2) subun 1.28e+02
	157	14	4.7	2299	17	T11640	Murine osteogenic pro 1.28e+02
	158	14	4.7	2368	15	Q90970	Alpha-beta dimer subu 1.28e+02
c	159	14	4.7	2368	4	Q23585	Anther specific promo 1.28e+02
c	160	14	4.7	2368	1	Q04444	Sequence of wun1-85 g 1.28e+02
c	161	14	4.7	2368	2	Q13372	Potato wun1 gene with 1.28e+02
	162	14	4.7	2438	1	N80408	Plasmid pRAJ320 inser 1.28e+02
	163	14	4.7	2457	1	N81451	Sequence encoding fus 1.28e+02
	164	14	4.7	2467	2	N60307	Equine IFN-beta from 1.28e+02
c	165	14	4.7	2484	3	Q15360	NF-YA. 1.28e+02
	166	14	4.7	2511	1	N92113	Polynucleotide encodi 1.28e+02
c	167	14	4.7	2539	26	T46075	Human activated leuko 1.28e+02
c	168	14	4.7	2571	1	N81168	Fragment from clone e 1.28e+02
	169	14	4.7	2579	1	N90315	Hepatitis cDNA virus 1.28e+02
	170	14	4.7	2579	1	N92085	Open reading frame of 1.28e+02
	171	14	4.7	2673	1	N81449	Sequence encoding fus 1.28e+02
c	172	14	4.7	2676	13	Q76371	Mouse growth differen 1.28e+02
c	173	14	4.7	2753	5	Q29229	The abfB gene of A. n 1.28e+02
	174	14	4.7	2762	35	T91038	Yeast checkpoint cont 1.28e+02
c	175	14	4.7	2973	9	Q53501	Sequence of bean mali 1.28e+02
	176	14	4.7	3169	24	T38397	pTrc 99 plasmid fragm 1.28e+02
c	177	14	4.7	3231	36	T86486	Eya type I cDNA encod 1.28e+02
c	178	14	4.7	3312	2	N71034	pMx34 cDNA insert. 1.28e+02
	179	14	4.7	3410	1	N80617	Genomic clone encodin 1.28e+02
	180	14	4.7	3411	2	N80644	Sequence of human pul 1.28e+02
c	181	14	4.7	3458	3	Q14256	51.4 and 41.9 KD toxi 1.28e+02
c	182	14	4.7	3458	8	Q48714	Insecticidal protein 1.28e+02
	183	14	4.7	3505	17	Q94350	Full length Bacillus 1.28e+02
	184	14	4.7	3505	1	Q06320	Sequence encoding cel 1.28e+02
c	185	14	4.7	3675	21	T17045	B.thuringiensis insec 1.28e+02
	186	14	4.7	3704	29	T59938	Phage resistance gene 1.28e+02
c	187	14	4.7	3861	9	Q51543	Lipopolysaccharide in 1.28e+02
c	188	14	4.7	4000	33	T86166	Nucleotide sequence o 1.28e+02
	189	14	4.7	4146	20	T09026	Arabidopsis thaliana 1.28e+02
	190	14	4.7	4284	12	Q74259	Stress-induced tobacc 1.28e+02
c	191	14	4.7	4286	20	T18079	Heml gene. 1.28e+02
c	192	14	4.7	4300	4	Q25812	Clone 45-A. 1.28e+02
	193	14	4.7	4577	33	T72813	Mouse SPT6 cDNA. 1.28e+02
c	194	14	4.7	4796	38	T76897	Cauliflower floral me 1.28e+02
	195	14	4.7	4982	33	T85267	HMG-CoA reductase deg 1.28e+02

	196	14	4.7	5183	39	V04063	Schizosaccharomyces p 1.28e+02
c	197	14	4.7	5236	4	Q25815	Clone 45-A; subtype 1 1.28e+02
	198	14	4.7	5300	1	N92097	Combined open reading 1.28e+02
	199	14	4.7	5360	1	N90327	Hepatitis C virus com 1.28e+02
	200	14	4.7	5361	32	T78868	P. falciparum liver s 1.28e+02
c	201	14	4.7	5408	15	Q95494	Human Cdn-3 DNA. 1.28e+02
	202	14	4.7	5433	7	Q40945	Human DNA polymerase 1.28e+02
c	203	14	4.7	5438	5	Q29263	Human calcium channel 1.28e+02
c	204	14	4.7	5496	12	Q73395	CwLJI coding sequence 1.28e+02
	205	14	4.7	5534	25	T43137	pUMIGIT sequence incl 1.28e+02
	206	14	4.7	5534	6	Q35988	Tomato hsp80 genomic 1.28e+02
c	207	14	4.7	5537	15	Q90330	Exons I, II & III and 1.28e+02
c	208	14	4.7	5537	15	Q89554	Human cholesterol 7 a 1.28e+02
c	209	14	4.7	5537	34	T79736	Human cholesterol 7-a 1.28e+02
	210	14	4.7	5642	8	Q51193	pDE4; plasmid DNA rep 1.28e+02
c	211	14	4.7	6075	4	Q27451	Type A human platelet 1.28e+02
	212	14	4.7	6152	32	T78867	P. falciparum liver s 1.28e+02
	213	14	4.7	6350	23	T35198	Plasmid pSC19-SFG. 1.28e+02
	214	14	4.7	6412	1	Q05989	TR4 cDNA clone relate 1.28e+02
c	215	14	4.7	6511	15	Q95493	Human Cdn-2 DNA. 1.28e+02
	216	14	4.7	6854	15	Q99852	Yeast TIH2 gene. 1.28e+02
	217	14	4.7	6905	1	N92103	Combined open reading 1.28e+02
	218	14	4.7	7050	7	Q40419	Sequence of pTE4. 1.28e+02
	219	14	4.7	7310	20	Q98221	Hepatitis C virus clo 1.28e+02
	220	14	4.7	7310	1	N92106	Combined open reading 1.28e+02
	221	14	4.7	7310	1	N90336	Composite hepatitis C 1.28e+02
c	222	14	4.7	7499	1	N91312	pCHL1 plasmid. 1.28e+02
c	223	14	4.7	7502	4	Q27429	pCTD. 1.28e+02
c	224	14	4.7	7620	6	Q39286	Glucocerebrosidase ge 1.28e+02
	225	14	4.7	8239	29	T61981	Human ataxia and rad 1.28e+02
c	226	14	4.7	8298	9	Q50168	p190 protein gene. 1.28e+02
	227	14	4.7	8316	1	Q05955	Hepatitis C virus ope 1.28e+02
	228	14	4.7	8388	13	Q78191	Vector M48 used for g 1.28e+02
	229	14	4.7	9185	2	Q10566	Hepatitis C virus str 1.28e+02
	230	14	4.7	9185	1	Q05956	Sense strand of the c 1.28e+02
	231	14	4.7	9379	6	Q36209	Composite cDNA for He 1.28e+02
	232	14	4.7	9400	3	Q21744	Compiled HCV cDNA. 1.28e+02
	233	14	4.7	9401	36	T99981	HCV polyprotein codin 1.28e+02
	234	14	4.7	9401	18	T12710	Hepatitis C virus pol 1.28e+02
	235	14	4.7	9401	39	V09989	HCV polyprotein codin 1.28e+02
c	236	14	4.7	14176	34	T84564	Swinepox virus HindII 1.28e+02
c	237	14	4.7	14237	39	V05384	Human cathepsin K gen 1.28e+02
c	238	14	4.7	18177	1	N90490	DNA of human retinobl 1.28e+02
c	239	14	4.7	24025	21	T17515	Mutated BRCA1 genomic 1.28e+02
c	240	14	4.7	24025	21	T17455	Mutated BRCA1 genomic 1.28e+02
c	241	14	4.7	24026	21	T17513	Mutated BRCA1 genomic 1.28e+02
c	242	14	4.7	24026	21	T17512	Mutated BRCA1 genomic 1.28e+02
c	243	14	4.7	24026	21	T17528	Mutated BRCA1 genomic 1.28e+02
c	244	14	4.7	24026	21	T17522	Mutated BRCA1 genomic 1.28e+02
c	245	14	4.7	24026	23	T32612	BRCA1, human breast a 1.28e+02
c	246	14	4.7	24026	21	T17516	Mutated BRCA1 genomic 1.28e+02

c	247	14	4.7	24026	21	T17527	Mutated BRCA1 genomic 1.28e+02
c	248	14	4.7	24026	21	T17524	Mutated BRCA1 genomic 1.28e+02
c	249	14	4.7	24026	21	T17519	Mutated BRCA1 genomic 1.28e+02
c	250	14	4.7	24026	21	T17514	Mutated BRCA1 genomic 1.28e+02
c	251	14	4.7	24026	18	T18325	BRCA1, human breast a 1.28e+02
c	252	14	4.7	24026	21	T17518	Mutated BRCA1 genomic 1.28e+02
c	253	14	4.7	24026	21	T17521	Mutated BRCA1 genomic 1.28e+02
c	254	14	4.7	24026	21	T17530	BRCA1 genomic sequenc 1.28e+02
c	255	14	4.7	24026	21	T17523	Mutated BRCA1 genomic 1.28e+02
c	256	14	4.7	24026	21	T17526	Mutated BRCA1 genomic 1.28e+02
c	257	14	4.7	24026	21	T17517	Mutated BRCA1 genomic 1.28e+02
c	258	14	4.7	24026	21	T17529	Mutated BRCA1 genomic 1.28e+02
c	259	14	4.7	24029	21	T17520	Mutated BRCA1 genomic 1.28e+02
c	260	14	4.7	24031	21	T17525	Mutated BRCA1 genomic 1.28e+02
	261	14	4.7	26698	20	T08126	Mouse syndecan-1 gene 1.28e+02
c	262	14	4.7	26698	20	T08126	Mouse syndecan-1 gene 1.28e+02
	263	14	4.7	26700	40	V15946	Mouse syndecan gene s 1.28e+02
	264	14	4.7	26700	11	Q67902	Syndecan gene. 1.28e+02
c	265	14	4.7	26700	40	V15946	Mouse syndecan gene s 1.28e+02
c	266	14	4.7	26700	11	Q67902	Syndecan gene. 1.28e+02
c	267	14	4.7	32026	36	T60559	Recombinant adenoviru 1.28e+02
c	268	14	4.7	35000	36	T60557	Recombinant cis-actin 1.28e+02
c	269	14	4.7	35408	31	T59272	Recombinant adenoviru 1.28e+02
c	270	14	4.7	36538	36	T60558	Recombinant trans-act 1.28e+02
	271	14	4.7	37895	24	T34137	Gene for biosynthetic 1.28e+02
	272	14	4.7	37913	10	Q51128	Eicosapentaenoic acid 1.28e+02
	273	13	4.4	15	39	V09756	Mutant N-ras RNA frag 4.72e+02
	274	13	4.4	16	5	Q29776	5' amidated probe N-R 4.72e+02
	275	13	4.4	16	2	Q13900	Probe YZ1 to N-ras co 4.72e+02
	276	13	4.4	18	9	Q50788	HBV target sequence 1 4.72e+02
	277	13	4.4	20	9	Q53120	Gene detection sequen 4.72e+02
c	278	13	4.4	20	16	Q95416	Primer B (Group 2, Se 4.72e+02
c	279	13	4.4	20	3	N92509	Sequence of portion o 4.72e+02
c	280	13	4.4	20	32	T48679	Probe for detecting N 4.72e+02
c	281	13	4.4	20	32	T48680	Probe for detecting N 4.72e+02
c	282	13	4.4	20	32	T48678	Probe for detecting N 4.72e+02
c	283	13	4.4	20	36	T94337	Human DPC4 sequence t 4.72e+02
	284	13	4.4	21	17	Q98500	3' primer for human N 4.72e+02
	285	13	4.4	22	14	Q82309	Chromosome 11 (locus 4.72e+02
	286	13	4.4	23	1	Q04793	3'-5' primer used for 4.72e+02
	287	13	4.4	23	13	Q75037	S. aureus toxic shock 4.72e+02
	288	13	4.4	24	31	T72775	Primer #18 for Neisse 4.72e+02
	289	13	4.4	24	31	T72769	Primer #13 for Neisse 4.72e+02
c	290	13	4.4	24	2	N60025	Sequence of synthetic 4.72e+02
c	291	13	4.4	24	31	T72756	Primer #1 for Neisser 4.72e+02
c	292	13	4.4	24	31	T72758	Primer #3 for Neisser 4.72e+02
c	293	13	4.4	27	39	V09753	N-ras activated oncog 4.72e+02
c	294	13	4.4	30	10	Q55850	Primer 1 to amplify n 4.72e+02
c	295	13	4.4	30	14	Q86882	NlallIR gene PCR prim 4.72e+02
c	296	13	4.4	31	18	T00599	IL-3 amplimer IL-3-2. 4.72e+02
	297	13	4.4	32	18	T00600	IL-3 amplimer IL-3-3. 4.72e+02

	298	13	4.4	33	18	T00596	GM-CSF amplimer GM-CS 4.72e+02
c	299	13	4.4	34	18	T00595	GM-CSF amplimer GM-CS 4.72e+02
c	300	13	4.4	40	5	Q29780	Probe RAS wt for Ras 4.72e+02
	301	13	4.4	45	19	T11584	Oligonucleotide U6 fo 4.72e+02
c	302	13	4.4	48	19	T11592	Oligonucleotide L7 fo 4.72e+02
	303	13	4.4	62	32	T78324	Plasmid vector pvA(1) 4.72e+02
c	304	13	4.4	62	32	T78325	Plasmid vector pvA(1) 4.72e+02
	305	13	4.4	68	5	Q34088	Sequence of a microsa 4.72e+02
	306	13	4.4	86	11	Q62498	Staphylococcus aureus 4.72e+02
	307	13	4.4	98	2	Q10457	SP promoter expressib 4.72e+02
	308	13	4.4	109	19	T19771	Human gene signature 4.72e+02
c	309	13	4.4	111	19	T19769	Human gene signature 4.72e+02
	310	13	4.4	115	2	N91696	Sequence of 5' end of 4.72e+02
	311	13	4.4	120	35	T58614	N-ras gene exon 1. 4.72e+02
c	312	13	4.4	124	22	T25935	Human gene signature 4.72e+02
c	313	13	4.4	128	19	T20703	Human gene signature 4.72e+02
c	314	13	4.4	137	21	T24483	Human gene signature 4.72e+02
c	315	13	4.4	142	36	T91730	Mouse SmLIM/CRP2 gene 4.72e+02
c	316	13	4.4	146	5	Q29946	Packaging sequence re 4.72e+02
	317	13	4.4	151	22	T25044	Human gene signature 4.72e+02
	318	13	4.4	156	21	T25714	Human gene signature 4.72e+02
	319	13	4.4	157	21	T24323	Human gene signature 4.72e+02
c	320	13	4.4	159	26	T40825	Serine protease nfSP7 4.72e+02
	321	13	4.4	174	24	T18646	Corynebacterium sp. J 4.72e+02
	322	13	4.4	178	20	T23559	Human gene signature 4.72e+02
	323	13	4.4	190	19	T20870	Human gene signature 4.72e+02
c	324	13	4.4	190	3	Q23346	Fragment 3 of the res 4.72e+02
c	325	13	4.4	194	34	T90364	Phospholipase inhibit 4.72e+02
c	326	13	4.4	205	20	T22510	Human gene signature 4.72e+02
c	327	13	4.4	212	20	T22243	Human gene signature 4.72e+02
c	328	13	4.4	214	8	Q60908	Human brain Expressed 4.72e+02
	329	13	4.4	231	19	T20847	Human gene signature 4.72e+02
	330	13	4.4	236	20	T23780	Human gene signature 4.72e+02
	331	13	4.4	239	22	T25463	Human gene signature 4.72e+02
c	332	13	4.4	239	10	Q76804	Human genome fragment 4.72e+02
c	333	13	4.4	249	24	T35169	Human aprotinin-like 4.72e+02
	334	13	4.4	251	5	Q28276	PCR prod. contg. part 4.72e+02
c	335	13	4.4	265	20	T23743	Human gene signature 4.72e+02
c	336	13	4.4	281	8	Q59716	Human brain Expressed 4.72e+02
	337	13	4.4	282	8	Q61019	Human brain Expressed 4.72e+02
c	338	13	4.4	282	29	T44914	Partial sequence of C 4.72e+02
c	339	13	4.4	284	21	T15504	Streptococcus saproph 4.72e+02
	340	13	4.4	285	9	Q50695	Monellin single-chain 4.72e+02
	341	13	4.4	285	19	T09154	Single chain form mon 4.72e+02
	342	13	4.4	285	2	N93187	Sequence encoding sin 4.72e+02
	343	13	4.4	285	19	T11578	Single-chain monellin 4.72e+02
	344	13	4.4	285	40	V17095	Single chain monellin 4.72e+02
	345	13	4.4	285	35	T93527	DNA encoding single-c 4.72e+02
	346	13	4.4	285	38	T89115	Synthetic gene encodi 4.72e+02
c	347	13	4.4	285	37	V02630	S. xylosus 16S-23S rR 4.72e+02
	348	13	4.4	286	9	Q51022	Human FceRI beta gene 4.72e+02

	349	13	4.4	292	8	Q48036	"Fused monellin" gene 4.72e+02
	350	13	4.4	296	3	Q21454	Sequence of a synthet 4.72e+02
c	351	13	4.4	296	6	Q39703	Expressed Sequence Ta 4.72e+02
c	352	13	4.4	296	8	Q59115	Human brain Expressed 4.72e+02
	353	13	4.4	297	22	T25974	Human gene signature 4.72e+02
	354	13	4.4	299	4	Q24504	Cytochrome-b gene #4. 4.72e+02
	355	13	4.4	299	4	Q24501	Cytochrome-b gene #1. 4.72e+02
	356	13	4.4	299	4	Q24505	Cytochrome-b gene #5. 4.72e+02
	357	13	4.4	299	4	Q24503	Cytochrome-b gene #3. 4.72e+02
	358	13	4.4	299	4	Q24502	Cytochrome-b gene #2. 4.72e+02
	359	13	4.4	303	20	T21667	Human gene signature 4.72e+02
c	360	13	4.4	304	10	Q77422	Human genome fragment 4.72e+02
	361	13	4.4	305	19	T21318	Human gene signature 4.72e+02
c	362	13	4.4	312	2	Q10978	Human bone specific p 4.72e+02
	363	13	4.4	318	29	T65653	Polymorphic repeat se 4.72e+02
c	364	13	4.4	320	20	T22161	Human gene signature 4.72e+02
c	365	13	4.4	321	28	T61380	Hantavirus nuclear pr 4.72e+02
c	366	13	4.4	324	34	T90357	Phospholipase inhibit 4.72e+02
	367	13	4.4	328	20	T22636	Human gene signature 4.72e+02
c	368	13	4.4	332	21	T25413	Human gene signature 4.72e+02
	369	13	4.4	333	19	T20634	Human gene signature 4.72e+02
c	370	13	4.4	335	22	T25317	Human gene signature 4.72e+02
	371	13	4.4	353	8	Q60258	Human brain Expressed 4.72e+02
c	372	13	4.4	367	8	Q59479	Human brain Expressed 4.72e+02
	373	13	4.4	368	35	T69309	Murine metastatic nuc 4.72e+02
c	374	13	4.4	369	20	T23689	Human gene signature 4.72e+02
c	375	13	4.4	388	6	Q39664	Expressed Sequence Ta 4.72e+02
c	376	13	4.4	388	8	Q59076	Human brain Expressed 4.72e+02
c	377	13	4.4	389	8	Q61215	Human brain Expressed 4.72e+02
c	378	13	4.4	395	21	T24481	Human gene signature 4.72e+02
c	379	13	4.4	396	9	Q51042	Human glucokinase gen 4.72e+02
c	380	13	4.4	396	9	Q51043	Human glucokinase gen 4.72e+02
c	381	13	4.4	396	9	Q51041	Human glucokinase gen 4.72e+02
c	382	13	4.4	396	9	Q51044	Human glucokinase gen 4.72e+02
c	383	13	4.4	398	8	Q45087	Human glucokinase gen 4.72e+02
	384	13	4.4	402	8	Q59315	Human brain Expressed 4.72e+02
	385	13	4.4	402	6	Q39903	Expressed Sequence Ta 4.72e+02
c	386	13	4.4	407	2	Q13110	HuVhlys. 4.72e+02
	387	13	4.4	416	21	T24348	Human gene signature 4.72e+02
c	388	13	4.4	417	1	N80438	Plasmid pERT87-4 cont 4.72e+02
c	389	13	4.4	417	3	N90337	Sequence of part of m 4.72e+02
c	390	13	4.4	420	33	T80743	Type II topoisomerase 4.72e+02
	391	13	4.4	431	19	T19912	Human gene signature 4.72e+02
	392	13	4.4	458	18	T19051	Human gene signature 4.72e+02
c	393	13	4.4	480	1	Q03631	Partial gene sequence 4.72e+02
	394	13	4.4	501	40	V24711	H. pylori ORF 06cp203 4.72e+02
	395	13	4.4	502	34	T87971	pCAL module M7-II(ss- 4.72e+02
	396	13	4.4	520	34	T87962	pCAL module.M7-II. 4.72e+02
c	397	13	4.4	525	31	T47150	Dendroides canadensis 4.72e+02
c	398	13	4.4	531	12	Q73901	Borrelia KL10 antigen 4.72e+02
	399	13	4.4	534	17	T05543	Human hepatitis B vir 4.72e+02

	400	13	4.4	549	13	Q73205	Hepatitis B Virus cor 4.72e+02
	401	13	4.4	552	14	Q81561	Hepatitis B virus pol 4.72e+02
	402	13	4.4	552	14	Q81562	Hepatitis B virus pol 4.72e+02
	403	13	4.4	582	8	Q47735	Hepatitis B core gene 4.72e+02
c	404	13	4.4	582	12	Q73876	Borrelia KL10 antigen 4.72e+02
	405	13	4.4	585	15	Q87589	DNA encoding Leucocyt 4.72e+02
	406	13	4.4	591	1	N91782	DNA probe complementa 4.72e+02
c	407	13	4.4	603	3	Q22242	Yeast Phosphoglycerat 4.72e+02
	408	13	4.4	612	13	Q73664	SLN-1 gene exon-2 fro 4.72e+02
	409	13	4.4	612	8	Q47736	Hepatitis B core gene 4.72e+02
c	410	13	4.4	612	10	Q54687	Anther specific cDNA 4.72e+02
	411	13	4.4	615	8	Q47737	Hepatitis B core gene 4.72e+02
	412	13	4.4	630	15	Q91984	Phage surface express 4.72e+02
	413	13	4.4	630	17	T04165	Delta-cpIII coding se 4.72e+02
c	414	13	4.4	632	1	Q06806	Clone encoding interl 4.72e+02
	415	13	4.4	645	8	Q47738	Hepatitis B core gene 4.72e+02
	416	13	4.4	646	18	T05205	Murine retinoic acid 4.72e+02
c	417	13	4.4	652	5	Q34564	Shrgrp. 4.72e+02
c	418	13	4.4	653	2	Q13676	Encodes A.niger Polyg 4.72e+02
	419	13	4.4	661	1	Q04646	Encodes deleted form 4.72e+02
	420	13	4.4	661	3	Q14735	Encodes truncated bet 4.72e+02
	421	13	4.4	666	10	Q44655	Sequence encoding mem 4.72e+02
	422	13	4.4	666	13	Q78812	Bacteriophage coat pr 4.72e+02
	423	13	4.4	666	11	Q64222	Gene coding for membr 4.72e+02
c	424	13	4.4	669	14	Q75975	HSP27 gene upstream e 4.72e+02
	425	13	4.4	687	1	Q04000	Sequence complementar 4.72e+02
	426	13	4.4	701	30	T70243	Plasmid Library 16-19 4.72e+02
	427	13	4.4	704	30	T70242	Plasmid pDW1 #14 enco 4.72e+02
	428	13	4.4	704	30	T70244	DNA fragment encoding 4.72e+02
c	429	13	4.4	705	30	T72057	Sequence flanking mar 4.72e+02
c	430	13	4.4	705	30	T43937	Sequence flanking mar 4.72e+02
c	431	13	4.4	706	5	Q31360	pUC-RVh-PM1f. 4.72e+02
	432	13	4.4	708	10	Q44657	Sequence encoding mem 4.72e+02
	433	13	4.4	708	11	Q64228	PCR amplified M13 coa 4.72e+02
	434	13	4.4	708	5	Q30712	Sequence of PCR modif 4.72e+02
	435	13	4.4	720	9	Q53372	Sequence encoding glu 4.72e+02
c	436	13	4.4	720	7	Q45918	Phage P1 ref gene. 4.72e+02
c	437	13	4.4	731	4	Q23575	CD4VHKOL-Thr113 cDNA 4.72e+02
c	438	13	4.4	731	4	Q23573	Reshaped CD4 antibody 4.72e+02
	439	13	4.4	773	13	Q80307	Synthetic uPA1-48 gen 4.72e+02
c	440	13	4.4	774	35	T79673	BRCA2 cancer suscepti 4.72e+02
	441	13	4.4	776	22	T36777	Prostate-specific mem 4.72e+02
	442	13	4.4	779	13	Q80305	M1Flag uPA1-48/pIII f 4.72e+02
c	443	13	4.4	783	1	Q04064	Fowl Pox Virus promot 4.72e+02
c	444	13	4.4	789	2	Q10792	Bovine vascular endot 4.72e+02
c	445	13	4.4	789	7	Q44255	Partial bovine VEGF-1 4.72e+02
c	446	13	4.4	804	40	V24742	H. pylori ORF 07ap806 4.72e+02
c	447	13	4.4	817	4	Q23572	Reshaped CD4 antibody 4.72e+02
c	448	13	4.4	817	4	Q23569	Rat CD4 antibody heav 4.72e+02
c	449	13	4.4	817	4	Q23574	Reshaped CD4 antibody 4.72e+02
c	450	13	4.4	818	1	Q04461	M13 VHPCR1 with Ig he 4.72e+02

c	451	13	4.4	826	10	Q56689	Genomic sequence of t 4.72e+02
c	452	13	4.4	829	2	Q10796	Bovine vascular endot 4.72e+02
	453	13	4.4	837	38	T96841	Intron 11 of human pR 4.72e+02
c	454	13	4.4	839	34	T90355	Novel phospholipase i 4.72e+02
c	455	13	4.4	850	2	N60477	N-terminal region of 4.72e+02
c	456	13	4.4	850	1	Q03624	19kD maize cz19C2 pre 4.72e+02
c	457	13	4.4	851	4	N30060	Sequence of the PGF (4.72e+02
c	458	13	4.4	851	1	Q04372	Sequence of clone cZ1 4.72e+02
	459	13	4.4	854	40	V17598	Human truncated cycli 4.72e+02
c	460	13	4.4	877	13	Q78945	Human immunoglobulin 4.72e+02
	461	13	4.4	892	33	T61423	Human cytidine deamin 4.72e+02
c	462	13	4.4	894	26	T40849	Serine protease nfSP7 4.72e+02
	463	13	4.4	903	1	Q03878	Merozoite surface ant 4.72e+02
	464	13	4.4	912	35	T78779	Pyrobaculum aerophil 4.72e+02
c	465	13	4.4	915	38	V06126	Viral infection gene 4.72e+02
	466	13	4.4	923	16	Q74054	Human interleukin-3-a 4.72e+02
c	467	13	4.4	927	1	N92874	Expression plasmid pU 4.72e+02
c	468	13	4.4	943	14	Q89539	Wheat cold tolerance 4.72e+02
	469	13	4.4	953	13	Q80299	Plasmid insert compri 4.72e+02
c	470	13	4.4	961	7	Q44259	Bovine VEGF-164 codin 4.72e+02
c	471	13	4.4	961	2	Q10791	Bovine vascular endot 4.72e+02
c	472	13	4.4	962	4	Q23521	Encodes resistance to 4.72e+02
c	473	13	4.4	963	16	T02360	TIMP-3 DNA clone Timp 4.72e+02
	474	13	4.4	996	15	Q87588	DNA encoding Leucocyt 4.72e+02
	475	13	4.4	1006	40	V17599	Human cyclin C cDNA. 4.72e+02
c	476	13	4.4	1015	37	Q01887	Human cytokeleton-as 4.72e+02
	477	13	4.4	1016	12	Q70049	Dirofilaria immitis p 4.72e+02
c	478	13	4.4	1020	1	N92873	Expression plasmid pU 4.72e+02
c	479	13	4.4	1029	6	Q38933	deg-1 gene. 4.72e+02
	480	13	4.4	1045	30	T43938	Sequence flanking mar 4.72e+02
	481	13	4.4	1045	30	T72058	Sequence flanking mar 4.72e+02
	482	13	4.4	1080	31	T68789	Melanocortin-3 recept 4.72e+02
	483	13	4.4	1080	38	V06399	Human melanocortin-3 4.72e+02
c	484	13	4.4	1080	2	Q11802	Clone pTOM36. 4.72e+02
c	485	13	4.4	1080	8	Q46682	pTOM36 cDNA clone. 4.72e+02
	486	13	4.4	1082	39	V01534	Human acylcoenzyme A: 4.72e+02
c	487	13	4.4	1107	11	Q68716	GDI D4 cDNA. 4.72e+02
c	488	13	4.4	1114	29	T51784	Human colon specific 4.72e+02
c	489	13	4.4	1118	3	Q20263	Deg-1 gene. 4.72e+02
	490	13	4.4	1119	39	V15099	N-methylamino acid ox 4.72e+02
c	491	13	4.4	1120	2	N70643	Cycloheximide resista 4.72e+02
c	492	13	4.4	1155	23	T34864	Human GP2 pancreatic 4.72e+02
	493	13	4.4	1161	2	N70828	DNA encoding human pr 4.72e+02
c	494	13	4.4	1171	2	Q13429	Calcium channel gamma 4.72e+02
	495	13	4.4	1181	15	Q85933	cDNA encoding Dac gV 4.72e+02
	496	13	4.4	1196	2	N91627	Sequence of gamma-3 i 4.72e+02
c	497	13	4.4	1203	15	Q87971	Factor-XA cDNA. 4.72e+02
	498	13	4.4	1205	2	N70509	Sequence of prepro Rh 4.72e+02
	499	13	4.4	1205	4	Q25436	Sequence of prepro 11 4.72e+02
	500	13	4.4	1244	31	T67289	Human aurora-1 cDNA. 4.72e+02
	501	13	4.4	1254	1	N92937	cDNA from Plasmodium 4.72e+02

c	502	13	4.4	1260	38	T99408	Human integrin regula 4.72e+02
c	503	13	4.4	1286	28	T46177	Cytotoxic T lymphocyt 4.72e+02
c	504	13	4.4	1287	37	T75506	P. americanus antifr 4.72e+02
c	505	13	4.4	1287	29	T66420	Hantaan virus S genom 4.72e+02
c	506	13	4.4	1295	1	N92737	Insert of clone pHDS4 4.72e+02
	507	13	4.4	1302	5	Q28835	M13 major coat protei 4.72e+02
c	508	13	4.4	1309	1	Q04914	pING123CVS sequence c 4.72e+02
c	509	13	4.4	1309	1	Q04913	pING152CVS sequence c 4.72e+02
c	510	13	4.4	1310	22	T36778	Prostate-specific mem 4.72e+02
	511	13	4.4	1320	1	N92341	Single stranded DNA s 4.72e+02
c	512	13	4.4	1320	5	Q30998	Serrate gene. 4.72e+02
c	513	13	4.4	1347	4	Q25180	Enterococcus faecium 4.72e+02
	514	13	4.4	1381	3	Q14472	E1-beta subunit of hu 4.72e+02
c	515	13	4.4	1390	2	Q13751	Peptidylglycine alpha 4.72e+02
	516	13	4.4	1432	2	Q11685	Clone encoding human 4.72e+02
	517	13	4.4	1434	38	V04900	Nucleotide sequence o 4.72e+02
c	518	13	4.4	1438	1	N93001	Rhoptry membrane anti 4.72e+02
	519	13	4.4	1440	26	T28571	Bacterial antibiotic 4.72e+02
	520	13	4.4	1443	3	Q14943	Xylose isomerase gene 4.72e+02
	521	13	4.4	1455	5	Q28836	Signal:BPTI:mature II 4.72e+02
	522	13	4.4	1480	31	T73288	A. thaliana origin of 4.72e+02
	523	13	4.4	1480	30	T62361	Arabidopsis thaliana 4.72e+02
	524	13	4.4	1501	3	Q21726	Parainfluenza type 4A 4.72e+02
	525	13	4.4	1510	18	T15760	Human interleukin-15 4.72e+02
	526	13	4.4	1514	1	Q05981	Insert from pTPRLI en 4.72e+02
c	527	13	4.4	1518	1	N92872	Expression plasmid pU 4.72e+02
	528	13	4.4	1537	20	T26895	Bacteriophage resista 4.72e+02
c	529	13	4.4	1543	2	Q06600	Clone lambda.vegf.6 c 4.72e+02
	530	13	4.4	1547	40	V15419	Human macrophage/dend 4.72e+02
	531	13	4.4	1555	23	T29142	rRNA gene (Stp.aureus 4.72e+02
	532	13	4.4	1566	24	T18645	16S ribosomal RNA gen 4.72e+02
	533	13	4.4	1599	1	N90733	Genomic DNA from Plas 4.72e+02
c	534	13	4.4	1619	12	Q71364	Expression cassette c 4.72e+02
c	535	13	4.4	1665	1	N92871	Expression plasmid pU 4.72e+02
	536	13	4.4	1666	24	T45050	Beta-globin gene for 4.72e+02
	537	13	4.4	1676	31	T73289	C. elegans origin of 4.72e+02
	538	13	4.4	1676	30	T62362	Caenorhabditis elegan 4.72e+02
	539	13	4.4	1685	9	Q55345	Sequence of gene K1.P 4.72e+02
	540	13	4.4	1686	15	Q87587	DNA encoding Leucocyt 4.72e+02
c	541	13	4.4	1694	10	Q58735	Hantaan virus cDNA. 4.72e+02
	542	13	4.4	1717	15	Q92299	Mango class II thioes 4.72e+02
	543	13	4.4	1717	40	V15233	Class II acyl-ACP thi 4.72e+02
c	544	13	4.4	1722	15	Q94598	ShET2 enterotoxin loc 4.72e+02
	545	13	4.4	1747	15	Q89585	SnaB1 restriction fra 4.72e+02
	546	13	4.4	1748	37	T85206	Squalene epoxidase nu 4.72e+02
	547	13	4.4	1751	1	N71021	Sequence encoding D-a 4.72e+02
c	548	13	4.4	1754	5	Q30045	Canine parvovirus VP2 4.72e+02
c	549	13	4.4	1754	12	Q71365	Expression cassette c 4.72e+02
c	550	13	4.4	1759	31	T71218	Type I, p80 IL-1-rece 4.72e+02
c	551	13	4.4	1768	21	T07326	Mouse PH30 beta chain 4.72e+02
	552	13	4.4	1770	18	T14228	Human neural cell pro 4.72e+02

	553	13	4.4	1779	35	T78777	Aquifex glucosamine:f 4.72e+02
c	554	13	4.4	1783	15	Q76198	cDNA of human adenosi 4.72e+02
c	555	13	4.4	1786	2	N60939	Sequence encoding soy 4.72e+02
	556	13	4.4	1810	4	Q25593	Sequence encoding hum 4.72e+02
	557	13	4.4	1813	13	Q85583	Human bactericidal-pe 4.72e+02
	558	13	4.4	1813	17	T00951	Bactericidal/permeabi 4.72e+02
	559	13	4.4	1813	16	Q97607	Bactericidal/permeabi 4.72e+02
	560	13	4.4	1813	13	Q81445	Holoprotein rBPI50 ge 4.72e+02
	561	13	4.4	1813	38	V06404	Human bactericidal/pe 4.72e+02
	562	13	4.4	1813	12	Q72017	cDNA encoding recombi 4.72e+02
	563	13	4.4	1813	39	T99564	Human bactericidal/pe 4.72e+02
	564	13	4.4	1813	9	Q52488	pING4512 coding regio 4.72e+02
	565	13	4.4	1813	24	T39396	Recombinant bacterici 4.72e+02
	566	13	4.4	1813	25	T31853	Bactericidal permeabi 4.72e+02
	567	13	4.4	1813	34	T67140	Bactericidal/permeabi 4.72e+02
	568	13	4.4	1813	16	Q99816	Recombinant bacterici 4.72e+02
	569	13	4.4	1813	16	Q86285	Gene encoding rBPI fo 4.72e+02
	570	13	4.4	1813	9	Q52268	Human bactericidal/pe 4.72e+02
	571	13	4.4	1813	34	T67139	Human bactericidal/pe 4.72e+02
	572	13	4.4	1813	39	V14996	Human bactericidal/pe 4.72e+02
	573	13	4.4	1813	39	V13961	Human BPI protein enc 4.72e+02
	574	13	4.4	1813	39	V10337	Human recombinant BPI 4.72e+02
	575	13	4.4	1813	14	Q80827	Bactericidal/permeabi 4.72e+02
	576	13	4.4	1813	15	Q97483	Recombinant bacterici 4.72e+02
	577	13	4.4	1813	16	Q92640	DNA encoding recombin 4.72e+02
	578	13	4.4	1813	1	N92766	cDNA encoding human b 4.72e+02
	579	13	4.4	1813	12	Q67270	Bactericide BPI. 4.72e+02
	580	13	4.4	1813	12	Q73382	Recombinant bacterici 4.72e+02
	581	13	4.4	1830	9	Q50572	Asparaginylendopeptid 4.72e+02
c	582	13	4.4	1869	5	Q29189	cDNA encoding Plasmod 4.72e+02
	583	13	4.4	1882	36	T72942	Phaffia carotenoid bi 4.72e+02
c	584	13	4.4	1882	13	Q80910	Plasmodium falciparum 4.72e+02
c	585	13	4.4	1882	13	Q80913	Plasmodium falciparum 4.72e+02
	586	13	4.4	1893	20	T11560	Rat kynurenine aminot 4.72e+02
c	587	13	4.4	1896	23	T36788	Prostate-specific mem 4.72e+02
c	588	13	4.4	1906	1	Q04531	Human prolidase gene. 4.72e+02
	589	13	4.4	1920	10	Q61791	Sequence of murine pr 4.72e+02
	590	13	4.4	1920	28	T59634	Mouse prostaglandin H 4.72e+02
c	591	13	4.4	1934	2	N60201	Interferon-omega-1. 4.72e+02
c	592	13	4.4	1939	18	T00520	H6 promoted canine di 4.72e+02
	593	13	4.4	1942	15	T00952	Bovine phospholipase 4.72e+02
	594	13	4.4	1952	17	T04164	Plasmid pM13Fv fragme 4.72e+02
	595	13	4.4	1952	15	Q91987	Fv restriction fragme 4.72e+02
c	596	13	4.4	1969	39	V05125	Haemagglutinin gene s 4.72e+02
c	597	13	4.4	1969	6	Q34551	Herbicide resistant A 4.72e+02
c	598	13	4.4	1969	6	Q34552	Herbicide sensitive A 4.72e+02
c	599	13	4.4	1974	38	T49506	Transferrin receptor 4.72e+02
c	600	13	4.4	1974	21	Q94448	Bacterial transferrin 4.72e+02
c	601	13	4.4	1979	35	T74203	Human LYST2 cDNA. 4.72e+02
	602	13	4.4	1982	1	N90225	Malaria-specific Plas 4.72e+02
	603	13	4.4	1992	8	Q48591	Alcohol oxidase gene. 4.72e+02

c	604	13	4.4	2008	15	Q94599	EIET enterotoxin enco 4.72e+02
c	605	13	4.4	2018	17	T00755	Recombinant C.albican 4.72e+02
	606	13	4.4	2040	23	T29774	Bacillus thuringiensi 4.72e+02
c	607	13	4.4	2042	29	T59913	Yeast transcription r 4.72e+02
c	608	13	4.4	2076	13	Q75355	Potassium KGA channel 4.72e+02
c	609	13	4.4	2079	1	N94527	Expression plasmid pU 4.72e+02
c	610	13	4.4	2089	4	Q25380	Sequence encoding ace 4.72e+02
c	611	13	4.4	2097	9	Q52316	Coriolus hirsutus ARS 4.72e+02
	612	13	4.4	2100	1	N90712	cryD gene. 4.72e+02
	613	13	4.4	2101	20	T11743	Rat kynurenine aminot 4.72e+02
c	614	13	4.4	2121	35	T89348	Bacillus popilliae cr 4.72e+02
	615	13	4.4	2125	23	T36787	Prostate-specific mem 4.72e+02
c	616	13	4.4	2143	30	T66910	Human neuropeptide Y- 4.72e+02
	617	13	4.4	2156	32	T77305	cDNA encoding acetola 4.72e+02
	618	13	4.4	2156	32	T77306	cDNA encoding herbici 4.72e+02
	619	13	4.4	2171	1	N81668	Sequence of part of t 4.72e+02
c	620	13	4.4	2177	1	N82402	B subunit of human Fa 4.72e+02
	621	13	4.4	2178	4	Q25592	Encodes 4D5 Fab IgG a 4.72e+02
	622	13	4.4	2190	6	Q37237	gdhA gene of C.melass 4.72e+02
	623	13	4.4	2202	9	Q52325	Sequence of yeast fru 4.72e+02
	624	13	4.4	2202	2	N70939	DNA sequence encoding 4.72e+02
c	625	13	4.4	2220	20	T27536	Attachment-invasion-l 4.72e+02
c	626	13	4.4	2226	4	Q25381	Sequence encoding ace 4.72e+02
	627	13	4.4	2237	24	T38267	200 gene differential 4.72e+02
c	628	13	4.4	2254	6	Q35161	Feline parvovirus cap 4.72e+02
c	629	13	4.4	2294	28	T60957	Maize tapetum-specifi 4.72e+02
	630	13	4.4	2295	23	T31994	Nonsense-mediated mRN 4.72e+02
	631	13	4.4	2304	20	T11742	Rat kynurenine aminot 4.72e+02
	632	13	4.4	2315	24	T38935	Transposon Tan. 4.72e+02
	633	13	4.4	2330	23	T28191	Feline herpesvirus gl 4.72e+02
c	634	13	4.4	2344	40	V20655	Open reading frame RH 4.72e+02
c	635	13	4.4	2345	19	T07170	Human GRB-7 genomic D 4.72e+02
c	636	13	4.4	2400	9	Q49353	deg-3. 4.72e+02
	637	13	4.4	2401	9	Q50146	Restriction enzyme. 4.72e+02
c	638	13	4.4	2414	17	T04895	Rat SCF genomic DNA e 4.72e+02
	639	13	4.4	2419	7	Q42843	cdc25A cDNA. 4.72e+02
	640	13	4.4	2420	13	Q79743	Human cdc25A gene. 4.72e+02
	641	13	4.4	2420	22	T30082	Human CDC25A phosphat 4.72e+02
	642	13	4.4	2425	21	T10283	Gene for RNA componen 4.72e+02
	643	13	4.4	2425	18	T11027	DNA encoding the huma 4.72e+02
	644	13	4.4	2436	40	V20438	Human c-N-ras oncogen 4.72e+02
c	645	13	4.4	2467	15	Q90271	Saccharomyces sp. rec 4.72e+02
c	646	13	4.4	2467	2	N60307	Equine IFN-beta from 4.72e+02
	647	13	4.4	2489	14	Q79868	Lactobacillus curvatu 4.72e+02
c	648	13	4.4	2529	35	T87468	Hamster Ubiquitin/S27 4.72e+02
c	649	13	4.4	2546	1	Q03661	Maize C3 sequence enc 4.72e+02
c	650	13	4.4	2553	21	T07327	Mouse PH30 beta chain 4.72e+02
	651	13	4.4	2594	34	T67317	Xenopus alpha-signali 4.72e+02
c	652	13	4.4	2599	1	Q05688	Islets of Langerhans 4.72e+02
c	653	13	4.4	2610	1	N93060	cDNA of plasmid pXA45 4.72e+02
c	654	13	4.4	2617	5	Q29753	DNA sequence encoding 4.72e+02

c	655	13	4.4	2625	15	Q87970	Peptidyl C-terminal a 4.72e+02
c	656	13	4.4	2662	20	T08977	Actual shortened B-pe 4.72e+02
c	657	13	4.4	2662	1	N90703	Rhoptry membrane anti 4.72e+02
	658	13	4.4	2707	10	Q57014	PKC epsilon. 4.72e+02
	659	13	4.4	2734	1	N80666	Residues 4457 to 7190 4.72e+02
	660	13	4.4	2748	1	N80665	Residues 4457 to 7204 4.72e+02
	661	13	4.4	2749	6	Q32649	S receptor kinase gen 4.72e+02
c	662	13	4.4	2750	6	Q38808	G6PD coding sequence. 4.72e+02
	663	13	4.4	2755	34	T87967	pCAL4 modular vector. 4.72e+02
	664	13	4.4	2761	15	Q76197	cDNA of rat liver ade 4.72e+02
c	665	13	4.4	2765	2	Q12556	Encodes Plasminogen m 4.72e+02
c	666	13	4.4	2771	2	Q12555	Encodes Plasminogen m 4.72e+02
c	667	13	4.4	2771	2	Q12551	Encodes Plasminogen m 4.72e+02
	668	13	4.4	2802	3	Q14809	Btm PG14 72kDa Cry in 4.72e+02
c	669	13	4.4	2832	11	Q62186	Cyanobacteria replica 4.72e+02
c	670	13	4.4	2834	34	T88222	Castor bean phospholi 4.72e+02
c	671	13	4.4	2835	32	T72666	Cell cycle checkpoint 4.72e+02
	672	13	4.4	2859	23	T28960	Human 26S proteasome 4.72e+02
c	673	13	4.4	2875	29	T47713	Mouse bone morphogene 4.72e+02
	674	13	4.4	2901	3	Q14810	Btm PG14 72kDa Cry in 4.72e+02
c	675	13	4.4	2919	30	T51124	Human homeoprotein re 4.72e+02
c	676	13	4.4	2967	1	Q03659	Maize C1 sequence enc 4.72e+02
c	677	13	4.4	3012	3	Q21188	Sequence encoding a 3 4.72e+02
	678	13	4.4	3025	6	Q39966	PKC-gamma cDNA. 4.72e+02
	679	13	4.4	3031	37	V00642	cDNA encoding murine 4.72e+02
c	680	13	4.4	3031	2	Q11272	Prepro-polygalacturon 4.72e+02
	681	13	4.4	3089	14	Q86268	N-acetylheparosan fra 4.72e+02
	682	13	4.4	3105	17	T07308	Receptor tyrosine kin 4.72e+02
	683	13	4.4	3105	34	T84528	Mouse Nuk tyrosine ki 4.72e+02
	684	13	4.4	3114	1	N80336	Sequence encoding pro 4.72e+02
	685	13	4.4	3142	14	Q82847	P-selectin gene. 4.72e+02
c	686	13	4.4	3186	34	T79099	H. armigera granulovi 4.72e+02
	687	13	4.4	3188	4	N40244	Hepatitis virus adr-B 4.72e+02
c	688	13	4.4	3192	17	T08222	X region antisense he 4.72e+02
c	689	13	4.4	3207	17	T08223	X region antisense he 4.72e+02
	690	13	4.4	3213	14	Q79869	Lactobacillus curvatu 4.72e+02
	691	13	4.4	3213	14	Q79870	Lactobacillus curvatu 4.72e+02
	692	13	4.4	3214	2	N60714	Sequence encoding adr 4.72e+02
	693	13	4.4	3214	1	Q05377	Fragment of plasmid p 4.72e+02
	694	13	4.4	3233	12	Q72690	Retinoblastoma 94kD t 4.72e+02
c	695	13	4.4	3243	2	N60306	Equine IFN-alpha-1 fr 4.72e+02
	696	13	4.4	3254	15	Q90659	Eph-related PTK Cek4. 4.72e+02
	697	13	4.4	3268	38	T96640	Human TUB Form 1 cDNA 4.72e+02
	698	13	4.4	3294	8	Q48939	Steroid hormone recep 4.72e+02
	699	13	4.4	3312	13	Q79326	Human ARD 1. 4.72e+02
c	700	13	4.4	3315	1	N90791	cDNA of plasmid pXA79 4.72e+02
	701	13	4.4	3332	2	Q12498	GAP-DH gene. 4.72e+02
	702	13	4.4	3336	3	N50149	Sequence of the cDNA 4.72e+02
c	703	13	4.4	3337	23	T34620	P. vivax ESP-1 blood 4.72e+02
	704	13	4.4	3360	2	Q05602	SSC1 gene. 4.72e+02
c	705	13	4.4	3383	3	Q20269	Gene encoding AE-III 4.72e+02

	706	13	4.4	3387	40	V16533	cDNA encoding mammali 4.72e+02
c	707	13	4.4	3425	9	Q57072	AGE-modified DNA INS- 4.72e+02
	708	13	4.4	3435	14	Q79746	Mycoplasma pirum adhe 4.72e+02
	709	13	4.4	3438	1	Q05759	GALP3 Promoter. 4.72e+02
c	710	13	4.4	3500	22	T18500	Yeast calcineurin-int 4.72e+02
	711	13	4.4	3509	20	T18813	NF-X1 cDNA. 4.72e+02
	712	13	4.4	3534	29	T67410	H. pylori cytoplasmic 4.72e+02
c	713	13	4.4	3542	15	Q94624	Brush-1 cDNA probe. 4.72e+02
c	714	13	4.4	3545	1	Q03817	Marek's disease virus 4.72e+02
	715	13	4.4	3549	30	T68137	H. pylori cytoplasmic 4.72e+02
c	716	13	4.4	3560	15	Q87444	Drosophila semaphorin 4.72e+02
	717	13	4.4	3615	14	Q55752	Escherichia coli geno 4.72e+02
c	718	13	4.4	3650	39	T88341	Aspergillus lysophosp 4.72e+02
	719	13	4.4	3681	32	T72154	Drosophila virilis Ks 4.72e+02
c	720	13	4.4	3681	32	T72154	Drosophila virilis Ks 4.72e+02
	721	13	4.4	3683	15	Q85999	Thermostable pullulan 4.72e+02
c	722	13	4.4	3730	18	T03195	Group B Streptococcal 4.72e+02
c	723	13	4.4	3746	1	N90693	Bovine peptidyl-glyci 4.72e+02
	724	13	4.4	3835	1	Q05378	Sequence of hepatitis 4.72e+02
c	725	13	4.4	3840	3	N92365	Sequence of maize ubi 4.72e+02
	726	13	4.4	3901	25	T42224	Human TATA-binding pr 4.72e+02
	727	13	4.4	3901	32	T79598	TATA-binding protein 4.72e+02
	728	13	4.4	3902	12	Q70738	TATA-binding protein- 4.72e+02
c	729	13	4.4	3915	20	T08974	Presumed shortened B- 4.72e+02
c	730	13	4.4	3932	11	Q64112	cryET5 gene. 4.72e+02
c	731	13	4.4	3934	30	T68434	CryET5 gene. 4.72e+02
c	732	13	4.4	3934	35	T95051	DNA encoding Bacillus 4.72e+02
	733	13	4.4	3959	4	Q24856	Sequence of T-cadheri 4.72e+02
	734	13	4.4	4000	38	T91902	Mannose-1-phosphate t 4.72e+02
c	735	13	4.4	4004	33	T51260	Human AD4 gene genomi 4.72e+02
	736	13	4.4	4080	23	T31993	Nonsense-mediated mRN 4.72e+02
	737	13	4.4	4108	32	T70123	DNA encoding beta-xyl 4.72e+02
c	738	13	4.4	4138	9	Q55179	Human orphan receptor 4.72e+02
c	739	13	4.4	4138	13	Q77791	Cobra partial CVF2 co 4.72e+02
c	740	13	4.4	4200	18	T03190	Group B Streptococcal 4.72e+02
c	741	13	4.4	4234	15	Q87819	Agmenellum quadruplic 4.72e+02
c	742	13	4.4	4253	28	T58361	Tobacco epi-5-aristol 4.72e+02
c	743	13	4.4	4319	18	T17215	Adhesion and penetrat 4.72e+02
c	744	13	4.4	4343	18	T00522	pMM126 coding sequenc 4.72e+02
	745	13	4.4	4346	24	T45897	Serine kinase SRPK1 c 4.72e+02
c	746	13	4.4	4359	15	Q94625	Brush-1 cDNA. 4.72e+02
	747	13	4.4	4378	1	Q04522	Multidrug Resistance 4.72e+02
c	748	13	4.4	4382	2	Q12759	P40 genomic DNA. 4.72e+02
c	749	13	4.4	4434	34	T93389	Bloom's syndrome BLM 4.72e+02
c	750	13	4.4	4437	34	T67013	Bloom syndrome active 4.72e+02
c	751	13	4.4	4437	34	T93390	Bloom's syndrome BLM 4.72e+02
c	752	13	4.4	4437	34	T93395	Bloom's syndrome BLM 4.72e+02
c	753	13	4.4	4437	34	T93394	Bloom's syndrome BLM 4.72e+02
c	754	13	4.4	4437	34	T93392	Bloom's syndrome BLM 4.72e+02
c	755	13	4.4	4438	34	T93393	Bloom's syndrome BLM 4.72e+02
c	756	13	4.4	4438	34	T93391	Bloom's syndrome BLM 4.72e+02

c	757	13	4.4	4527	2	Q12031	Herpes Virus of turke	4.72e+02
	758	13	4.4	4553	4	Q25591	Lipoxygenase structur	4.72e+02
	759	13	4.4	4555	17	Q94311	Tyrosine phosphatase	4.72e+02
	760	13	4.4	4579	1	N81261	Probe for retinoblast	4.72e+02
c	761	13	4.4	4590	2	N60472	Sequence encoding the	4.72e+02
	762	13	4.4	4597	12	Q70536	Human retinoblastoma	4.72e+02
	763	13	4.4	4597	1	N81369	Human retinoblastoma	4.72e+02
c	764	13	4.4	4604	18	T00523	pC5LCDVHAP1 coding se	4.72e+02
	765	13	4.4	4624	17	T04318	Expression vector pUC	4.72e+02
	766	13	4.4	4646	12	Q72872	Human multidrug resis	4.72e+02
c	767	13	4.4	4665	38	T97955	Human epididymis-spec	4.72e+02
	768	13	4.4	4669	2	N70752	Sequence of human mul	4.72e+02
	769	13	4.4	4691	16	Q92546	pComb3 expression vec	4.72e+02
c	770	13	4.4	4696	19	T13330	Rat type I insulin-li	4.72e+02
c	771	13	4.4	4698	35	T68849	Photorhabdus luminesc	4.72e+02
c	772	13	4.4	4706	17	T10157	Vector plasmid pOnMTG	4.72e+02
c	773	13	4.4	4744	17	T10155	Sockeye salmon growth	4.72e+02
	774	13	4.4	4796	38	T76897	Cauliflower floral me	4.72e+02
	775	13	4.4	4809	11	Q62185	pAQ1 plasmid fragment	4.72e+02
	776	13	4.4	4839	2	Q06827	Alpha amylase pullula	4.72e+02
c	777	13	4.4	4854	2	Q11538	Rat Stem Cell Factor	4.72e+02
	778	13	4.4	4978	18	T02986	Bovine phospholipase	4.72e+02
	779	13	4.4	4978	12	Q74281	PLA2-I receptor entir	4.72e+02
c	780	13	4.4	4989	15	Q85962	Mouse ST2L gene.	4.72e+02
c	781	13	4.4	5000	40	V20609	Human kynurenine-3-hy	4.72e+02
	782	13	4.4	5009	21	Q94444	Bacterial transferrin	4.72e+02
	783	13	4.4	5018	3	N50150	Sequence of the cDNA	4.72e+02
	784	13	4.4	5018	2	N70646	Clone 23B6p102 encodi	4.72e+02
	785	13	4.4	5040	2	N80916	Sequence encoding the	4.72e+02
c	786	13	4.4	5049	39	T88324	Attenuated canine par	4.72e+02
c	787	13	4.4	5049	39	T88320	Canine parvovirus 39	4.72e+02
c	788	13	4.4	5049	39	T88321	Attenuated canine par	4.72e+02
c	789	13	4.4	5049	22	T15311	Non-attenuated canine	4.72e+02
c	790	13	4.4	5049	22	T15312	Attenuated canine par	4.72e+02
	791	13	4.4	5056	1	N90489	cDNA of human retinob	4.72e+02
	792	13	4.4	5100	6	Q37760	Glucosyltransferease	4.72e+02
c	793	13	4.4	5159	9	Q51559	Keratin K6 gene.	4.72e+02
	794	13	4.4	5227	36	T79537	Plasmid pTERMsc2Hl0my	4.72e+02
	795	13	4.4	5253	17	T05114	mSOS2 gene.	4.72e+02
	796	13	4.4	5254	17	T16231	Human NMDA receptor R	4.72e+02
	797	13	4.4	5278	9	Q58624	Mammalian son of seve	4.72e+02
	798	13	4.4	5327	36	T99971	Zfx gene.	4.72e+02
	799	13	4.4	5376	1	N90732	Sequence of plasmid p	4.72e+02
c	800	13	4.4	5394	17	T00872	Murine mC26 structura	4.72e+02
	801	13	4.4	5418	39	T94191	Promoter region of pi	4.72e+02
	802	13	4.4	5510	2	Q12707	Rice actin gene RAc1.	4.72e+02
c	803	13	4.4	5524	1	N90123	DNA encoding human co	4.72e+02
	804	13	4.4	5538	13	Q79377	Human N-methyl-D-aspa	4.72e+02
c	805	13	4.4	5561	7	Q43910	Sequence encoding a s	4.72e+02
	806	13	4.4	5643	33	T80055	Rice actin 1 gene.	4.72e+02
	807	13	4.4	5858	2	Q10378	Encodes human 160kD m	4.72e+02

	808	13	4.4	5886	15	Q89838	Human death associate	4.72e+02
	809	13	4.4	5893	35	T74200	Murine Lyst1 short is	4.72e+02
	810	13	4.4	5983	22	Q99962	Human excitatory amin	4.72e+02
c	811	13	4.4	6060	28	T48454	Human beta-globin gen	4.72e+02
	812	13	4.4	6124	32	T71265	Neuronal apoptosis in	4.72e+02
	813	13	4.4	6133	32	T71264	Neuronal apoptosis in	4.72e+02
c	814	13	4.4	6136	14	Q90478	Chinook salmon growth	4.72e+02
	815	13	4.4	6208	13	Q81210	Feline herpesvirus-1	4.72e+02
	816	13	4.4	6228	32	T71266	Neuronal apoptosis in	4.72e+02
	817	13	4.4	6505	19	T13394	Hybrid vector pSF-MDR	4.72e+02
	818	13	4.4	6511	15	Q95493	Human Cdn-2 DNA.	4.72e+02
	819	13	4.4	6760	16	T02364	Arabidopsis thaliana	4.72e+02
	820	13	4.4	6901	13	Q73703	Rice sucrose phosphat	4.72e+02
	821	13	4.4	6924	12	Q67283	Human RPTP-beta.	4.72e+02
c	822	13	4.4	6971	24	T38740	Moraxella outer membr	4.72e+02
c	823	13	4.4	7152	1	N80153	Human Rhinovirus Stra	4.72e+02
c	824	13	4.4	7152	2	N81390	Sequence correspondin	4.72e+02
	825	13	4.4	7294	11	Q66626	Right-half vector M13	4.72e+02
	826	13	4.4	7294	11	Q66630	Right-half vector M13	4.72e+02
	827	13	4.4	7294	4	Q24174	M13IX421.	4.72e+02
	828	13	4.4	7294	4	Q24170	M13IX42.	4.72e+02
	829	13	4.4	7317	37	V03651	M13-based vector M13I	4.72e+02
	830	13	4.4	7320	11	Q66627	Left-half vector M13I	4.72e+02
	831	13	4.4	7320	4	Q24171	M13IX22.	4.72e+02
	832	13	4.4	7394	4	Q24175	M13IED04.	4.72e+02
	833	13	4.4	7394	11	Q66631	Left-half vector M13E	4.72e+02
	834	13	4.4	7409	4	Q24173	M13IXD03.	4.72e+02
	835	13	4.4	7409	11	Q66629	Left-half vector M13E	4.72e+02
	836	13	4.4	7440	3	N50375	DNA sequence encoding	4.72e+02
	837	13	4.4	7445	4	Q24172	M13IX30.	4.72e+02
	838	13	4.4	7445	37	V03650	M13-based vector M13I	4.72e+02
	839	13	4.4	7445	11	Q66628	Vector M13IX30 for ra	4.72e+02
c	840	13	4.4	7493	40	V12115	Mus musculus factor V	4.72e+02
c	841	13	4.4	7502	32	T72665	Cell cycle checkpoint	4.72e+02
	842	13	4.4	7557	37	V03653	M13-based vector M13I	4.72e+02
c	843	13	4.4	7624	32	T72664	Cell cycle checkpoint	4.72e+02
	844	13	4.4	7652	7	Q45682	Sequencing vector JAN	4.72e+02
	845	13	4.4	7729	37	V03652	M13-based vector M13I	4.72e+02
c	846	13	4.4	7859	7	Q44265	pSW6 for expression o	4.72e+02
c	847	13	4.4	7859	2	Q12154	Shuttle vector pSW6.	4.72e+02
	848	13	4.4	7904	35	T94724	Human papillomavirus	4.72e+02
c	849	13	4.4	7984	4	Q25185	pSW6 expression vecto	4.72e+02
	850	13	4.4	8041	18	T10105	Adherence conferring	4.72e+02
	851	13	4.4	8045	17	T12237	Retroviral vector MFG	4.72e+02
	852	13	4.4	8113	37	V03654	M13-based vector M13I	4.72e+02
c	853	13	4.4	8239	29	T61981	Human ataxia and rad	4.72e+02
c	854	13	4.4	8371	30	T66909	Human neuropeptide Y-	4.72e+02
	855	13	4.4	8378	4	Q25811	Drosophila SLIT prote	4.72e+02
	856	13	4.4	8395	1	Q05056	BIV 106 viral genome.	4.72e+02
	857	13	4.4	8482	1	Q03642	BIV 127 viral genome.	4.72e+02
	858	13	4.4	8791	23	T37090	Human T cell inositol	4.72e+02

c	859	13	4.4	8806	17	T09349	Rat IP3 receptor gene	4.72e+02
	860	13	4.4	9048	25	T43225	Brassica napus FCA ge	4.72e+02
c	861	13	4.4	9060	3	Q15132	Zucchini ACC synthase	4.72e+02
c	862	13	4.4	9060	40	V15702	Zucchini ACC synthase	4.72e+02
	863	13	4.4	9202	2	N70608	VISNA sheep lentiviru	4.72e+02
c	864	13	4.4	9202	2	N70608	VISNA sheep lentiviru	4.72e+02
c	865	13	4.4	9739	3	Q14751	HIV-1(MN) env protein	4.72e+02
c	866	13	4.4	9871	2	Q13593	Inositol-3-phosphate	4.72e+02
	867	13	4.4	9955	1	Q05877	mGP1-2 plasmid sequen	4.72e+02
	868	13	4.4	9955	5	Q31625	mGP1-2 containing T7	4.72e+02
	869	13	4.4	9955	1	N81542	Sequence of mGP1-2.	4.72e+02
c	870	13	4.4	11236	12	Q70447	Rat nestin gene - its	4.72e+02
	871	13	4.4	11601	2	Q13608	ACV synthetase gene,	4.72e+02
	872	13	4.4	11800	11	Q67676	Acetyl CoA carboxylas	4.72e+02
	873	13	4.4	13058	8	Q48231	Vector containing ACV	4.72e+02
c	874	13	4.4	13417	24	T09224	Partial sequence of v	4.72e+02
	875	13	4.4	13585	17	T11549	Tumour rejection anti	4.72e+02
	876	13	4.4	14311	24	T38744	Plasmid pVSVFL(+) inc	4.72e+02
c	877	13	4.4	14507	32	T73568	Expression augmenting	4.72e+02
	878	13	4.4	15144	29	T47715	Mouse bone morphogene	4.72e+02
c	879	13	4.4	15144	29	T47715	Mouse bone morphogene	4.72e+02
	880	13	4.4	18177	1	N90490	DNA of human retinobl	4.72e+02
c	881	13	4.4	19012	6	Q36952	HSA gene.	4.72e+02
c	882	13	4.4	22481	24	T11658	PEDF full length sequ	4.72e+02
c	883	13	4.4	28598	22	T06769	Sorangium cellulosum	4.72e+02
c	884	13	4.4	28958	36	T89956	Sorangium cellulosum	4.72e+02
	885	13	4.4	28994	39	V15826	Genomic DNA for inter	4.72e+02
	886	13	4.4	29392	40	V15422	Mouse poly Ig recepto	4.72e+02
c	887	13	4.4	40875	33	T80043	Insert from cosmid 10	4.72e+02
c	888	13	4.4	49377	39	V05287	The soraphen biosynth	4.72e+02
	889	13	4.4	580073	27	T58840	Mycoplasma genitalium	4.72e+02
c	890	13	4.4	580073	27	T58840	Mycoplasma genitalium	4.72e+02
c	891	12	4.0	14	25	T50772	Probe P37 for target	1.62e+03
	892	12	4.0	16	2	Q13908	Probe YE50 to N-ras c	1.62e+03
c	893	12	4.0	16	38	V14176	Probe HBPr31 for geno	1.62e+03
c	894	12	4.0	19	38	V14392	Probe HBPr63 for geno	1.62e+03
c	895	12	4.0	29	40	V25347	Primer for H.pylori m	1.62e+03
c	896	12	4.0	41	39	V13063	Interleukin 6 variabl	1.62e+03
c	897	12	4.0	41	39	V13064	Interleukin 6 variabl	1.62e+03
	898	12	4.0	85	19	T21078	Human gene signature	1.62e+03
c	899	12	4.0	128	39	V15527	Polymorphic region of	1.62e+03
c	900	12	4.0	168	10	Q76936	Human genome fragment	1.62e+03
	901	12	4.0	184	6	Q37517	S. dysenteriae ompA f	1.62e+03
	902	12	4.0	207	13	Q77681	Bacterial middle wall	1.62e+03
c	903	12	4.0	226	19	T20051	Human gene signature	1.62e+03
c	904	12	4.0	318	10	Q77440	Human genome fragment	1.62e+03
c	905	12	4.0	336	40	V10325	Human Mab AB17.1.41 v	1.62e+03
c	906	12	4.0	361	39	V13262	N. tabacum strain Sam	1.62e+03
c	907	12	4.0	379	40	V09103	3' nucleotide sequenc	1.62e+03
	908	12	4.0	461	32	T78138	DNA sequence of Ll am	1.62e+03
c	909	12	4.0	542	40	V09124	Nucleotide sequence e	1.62e+03

c	910	12	4.0	654	40	V24611	H. pylori ORF 01cp117	1.62e+03
c	911	12	4.0	690	40	V24807	H. pylori ORF hp3p108	1.62e+03
c	912	12	4.0	803	3	N40014	DNA encoding fusion p	1.62e+03
c	913	12	4.0	815	40	V19702	Human cyclin C cDNA.	1.62e+03
	914	12	4.0	917	32	T77785	cDNA clone 60 encodin	1.62e+03
c	915	12	4.0	951	2	N60290	Sequence encoding hum	1.62e+03
c	916	12	4.0	990	3	N50284	Sequence encoding T-c	1.62e+03
c	917	12	4.0	996	38	V06400	Human melanocortin-4	1.62e+03
c	918	12	4.0	999	40	V03252	Homo sapiens DNA enco	1.62e+03
c	919	12	4.0	999	40	V03251	Homo sapiens DNA enco	1.62e+03
c	920	12	4.0	999	40	V03253	Homo sapiens DNA enco	1.62e+03
c	921	12	4.0	1057	40	T86628	APETALA1 gene from Ar	1.62e+03
c	922	12	4.0	1057	39	V02760	Arabidopsis floral me	1.62e+03
c	923	12	4.0	1057	39	V06018	Arabidopsis floral me	1.62e+03
c	924	12	4.0	1077	40	V02987	Human AX56_8 full len	1.62e+03
c	925	12	4.0	1077	40	V16469	Nucleotide sequence o	1.62e+03
c	926	12	4.0	1095	40	V24815	H. pylori ORF 14ee211	1.62e+03
	927	12	4.0	1121	17	T02574	Human fibroblast grow	1.62e+03
c	928	12	4.0	1182	11	Q67910	Alcaligenes eutrophus	1.62e+03
	929	12	4.0	1396	1	Q03184	Entire sequence of co	1.62e+03
c	930	12	4.0	1428	9	Q52544	Alpha-antichymotrypsi	1.62e+03
c	931	12	4.0	1435	40	V15249	Bay thioesterase clon	1.62e+03
c	932	12	4.0	1444	38	V09623	Human B cell SSRP1 cD	1.62e+03
	933	12	4.0	1536	19	T18765	Pseudomonas testoster	1.62e+03
c	934	12	4.0	1560	38	V12754	Biological clock cont	1.62e+03
	935	12	4.0	1576	6	Q40415	cDNA clone of pTOMUC8	1.62e+03
	936	12	4.0	1648	7	Q40502	Human ALT gene.	1.62e+03
c	937	12	4.0	1759	38	V09307	Human brain P2X-1 rec	1.62e+03
	938	12	4.0	1775	12	Q73303	Sequence describing M	1.62e+03
c	939	12	4.0	1807	38	V09308	Human brain P2X-2 rec	1.62e+03
c	940	12	4.0	1898	38	V09622	Human B cell SSRP1 cD	1.62e+03
c	941	12	4.0	1899	40	V18735	HRSV glycoprotein F c	1.62e+03
c	942	12	4.0	1974	39	V09703	Porcine retrovirus Ra	1.62e+03
	943	12	4.0	1977	17	T16307	Rice HY4 cDNA partial	1.62e+03
c	944	12	4.0	2004	39	V16167	DNA encoding an 80 kD	1.62e+03
c	945	12	4.0	2115	40	V15037	Promoter from B.flavu	1.62e+03
c	946	12	4.0	2122	39	V03995	Chicken matrix metall	1.62e+03
c	947	12	4.0	2135	12	Q67062	Rabphilin-3A cDNA.	1.62e+03
c	948	12	4.0	2176	40	V17698	Recombinant avipox vi	1.62e+03
	949	12	4.0	2309	13	Q73786	Partial PCR fragment	1.62e+03
c	950	12	4.0	2347	26	T48031	Bovine lactoferrin ge	1.62e+03
c	951	12	4.0	2352	40	V20575	Human pro-tumour necr	1.62e+03
c	952	12	4.0	2462	39	V05102	cDNA of cbpA, a choli	1.62e+03
	953	12	4.0	2478	1	N91246	Sequence of B. thurin	1.62e+03
c	954	12	4.0	2556	40	V19873	Drosophila p70S6K gen	1.62e+03
	955	12	4.0	2560	5	Q30910	pMDR1002 insert; pre-	1.62e+03
c	956	12	4.0	2655	39	V16166	DNA encoding a 100 kD	1.62e+03
c	957	12	4.0	2798	39	V04062	Yeast RNA-binding pro	1.62e+03
c	958	12	4.0	2837	38	V09621	Human SSRP1 cDNA.	1.62e+03
c	959	12	4.0	2899	40	V17699	Recombinant pox virus	1.62e+03
c	960	12	4.0	3102	40	V17260	Human protein p164 co	1.62e+03

	961	12	4.0	3131	11	Q66687	EHEC eae gene associa 1.62e+03
c	962	12	4.0	3180	29	T61223	Human cellular apopto 1.62e+03
c	963	12	4.0	3230	39	V09641	Human membrane protei 1.62e+03
c	964	12	4.0	3300	40	V17259	Rat protein p138 codi 1.62e+03
c	965	12	4.0	3479	38	V12284	Saccharomyces cerevis 1.62e+03
c	966	12	4.0	3541	39	V14522	CF-5 pathogen resista 1.62e+03
	967	12	4.0	3588	6	Q40057	Bacterial delta-6-des 1.62e+03
c	968	12	4.0	3596	9	Q52699	Corticosterone-bindin 1.62e+03
c	969	12	4.0	3935	13	Q79388	Human NMDA receptor s 1.62e+03
c	970	12	4.0	3979	39	V14518	CF-5 pathogen resista 1.62e+03
c	971	12	4.0	3979	39	V14519	CF-5 pathogen resista 1.62e+03
c	972	12	4.0	4041	39	V16172	DNA sequence encoding 1.62e+03
c	973	12	4.0	4123	39	V14523	CF-5 pathogen resista 1.62e+03
c	974	12	4.0	4212	39	V01098	Human propyl-4-hydrol 1.62e+03
c	975	12	4.0	4325	40	V15266	Brassica campestris n 1.62e+03
	976	12	4.0	4382	1	N81546	Bio A, Bio B and Bio 1.62e+03
c	977	12	4.0	4399	40	V19788	Mouse Rho target prot 1.62e+03
c	978	12	4.0	4752	38	V04868	Nucleotide sequence o 1.62e+03
c	979	12	4.0	4799	38	V04887	Nucleotide sequence o 1.62e+03
c	980	12	4.0	4815	38	V04886	Nucleotide sequence o 1.62e+03
c	981	12	4.0	4943	38	V12755	Biological clock cont 1.62e+03
	982	12	4.0	5110	14	Q83828	Plasmid pCM-P1160. 1.62e+03
c	983	12	4.0	5244	28	T48869	IamH expression vecto 1.62e+03
	984	12	4.0	5300	30	T62938	3F4 human IgG4 expres 1.62e+03
	985	12	4.0	5620	9	Q53875	Plasmid pVE108 used i 1.62e+03
c	986	12	4.0	5720	25	T42117	Lats gene encoding la 1.62e+03
c	987	12	4.0	6049	39	V16165	DNA encoding vegetati 1.62e+03
	988	12	4.0	6395	17	Q95155	Tobacco mosaic virus. 1.62e+03
c	989	12	4.0	6453	40	V20436	Human c-H-ras oncogen 1.62e+03
c	990	12	4.0	7101	19	T29247	C. difficile toxin B 1.62e+03
c	991	12	4.0	7167	40	V03228	Plasmid CD80Fcalphatp 1.62e+03
c	992	12	4.0	7244	40	V15705	Tomato ACC synthase L 1.62e+03
c	993	12	4.0	7587	40	V15703	Zucchini ACC synthase 1.62e+03
c	994	12	4.0	8196	38	V09699	Porcine retrovirus DN 1.62e+03
c	995	12	4.0	8209	39	V09700	Porcine retrovirus DN 1.62e+03
c	996	12	4.0	8392	9	Q53478	MLL gene 8.3 kb BamHI 1.62e+03
c	997	12	4.0	9100	27	T28521	H. influenzae detecti 1.62e+03
c	998	12	4.0	9997	40	V05839	Mouse huntingtin codi 1.62e+03
c	999	12	4.0	17410	40	V15205	Human osteogenic prot 1.62e+03
c	1000	12	4.0	35100	40	V20441	Human c-fms oncogene. 1.62e+03

ALIGNMENTS

```
RESULT    1
ID   T20017 standard; cDNA to mRNA; 304 BP.
AC   T20017;
DT   17-JUL-1996  (first entry)
DE   Human gene signature HUMGS01155.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 538; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   304 BP;   84 A;   41 C;   67 G;   106 T;

  Query Match           100.0%;  Score 298;  DB 19;  Length 304;
  Best Local Similarity 100.0%;  Pred. No. 5.87e-269;
  Matches   304;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db       1 gatctaaagcaggtngtgttgtttacatgtttctacacatttcatcctttaaaaagttgt 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy       1 GATCTAAAGCAGGTNGTGTTGTTTACATGTTTCTACACATTTCATCCTTTAAAAAGTTGT 60

Db      61 tgagagaggttgtatttaccttcccaaggttggaaagcaggggaatttcccagtgtccta 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      61 TGAGAGAGGTTGTATTTACCTTCCCAAGGTTGGAAAGCAGGGGAATTTCCCAGTGTCCTA 120

Db     121 gttttccaccagaggaatatgtgtaagtagcaaagtatttgctgcttacatatagtgtgt 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     121 GTTTTCCACCAGAGGAATATGTGTAAGTAGCAAAGTATTTGCTGCTTACATATAGTGTGT 180

Db     181 atgtatgtatatatgtaaatngtgtgttaaagagctgatactgatttttcatatgncaatg 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     181 ATGTATGTATATATGTAAATNGTGTGTTAAAGAGCTGATACTGATTTTCATATGNCAATG 240
```



```
Db     241 ttaaggcaaaggcctccctgcatttgangagcaggtnttcatttatatgtattttnggga 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     241 TTAAGGCAAAGGCCTCCCTGCATTTGANGAGCAGGTNTTCATTTATATGTATTTTNGGGA 300

Db     301 taaa 304
           ||||
Qy     301 TAAA 304


RESULT    2
ID   T65661 standard; DNA; 264 BP.
AC   T65661;
DT   17-JUN-1997  (first entry)
DE   Polymorphic repeat sequence clone Mfd22 from locus D4S171.
KW   Polymorphism; repeat sequence; genetic marker; primer; amplification;
KW   PCR; polymerase chain reaction; paternity; maternity; human; pedigree;
KW   linkage analysis; genetic disease; animal; plant; breeding; locus;
KW   hybridisation; chromosome; ds.
OS   Homo sapiens.
PN   US5582979-A.
PD   10-DEC-1996.
PF   21-APR-1989; 341562.
PR   21-APR-1989; US-341562.
PR   05-SEP-1991; US-754351.
PR   04-APR-1994; US-222177.
PA   (MARS-) MARSHFIELD CLINIC.
PI   Weber JL;
DR   WPI; 97-042299/04.
PT   Detection of polymorphic genetic markers of the form
PT   (dC-dA)n(dG-dT)n - using novel nucleic acid mols. as primers
PS   Claim 1; Column 41-42; 186pp; English.
CC   The invention relates to the isolation of polymorphic repeat sequences
CC   having the sequence (dC-dA)n.(dG-dT)n which can be used as genetic
CC   markers.  Primers based on these sequences can be used to detect these
CC   repeats, especially for use in e.g paternity or maternity testing,
CC   human genetic analysis such as linkage analysis of genetic disease,
CC   commercial animal or plant breeding or pedigree analysis.  The clones
CC   T65652-T65703 are examples of the amplified polymorphic repeat sequences
CC   found at the loci given in the DE line.  The repeat sequences were
CC   isolated by hybridisation of chromosome-specific phage libraries with
CC   a synthetic poly(dC-dA).(dG-dT) probe.  Over 100 repeat blocks were
CC   isolated.  Those clones where the repeat sequence has been determined
CC   are shown in T65704-797.
SQ   Sequence   264 BP;   105 A;   65 C;   48 G;   46 T;

  Query Match            6.7%;  Score 20;  DB 29;  Length 264;
  Best Local Similarity 100.0%;  Pred. No. 1.55e-02;
  Matches   20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     139 atatatacatacatacacac 158
```



```
           ||||||||||||||||||||
Cp     194 ATATATACATACATACACAC 175


RESULT    3
ID   T41853 standard; cDNA; 5511 BP.
AC   T41853;
DT   20-FEB-1997  (first entry)
DE   cDNA encoding Plasmodium falciparum erythrocyte membrane protein.
KW   Plasmodium falciparum; erythrocyte membrane protein; malaria;
KW   detection; identification; treatment; prevention; parasite; ss.
OS   Plasmodium falciparum MC type.
FH   Key             Location/Qualifiers
FT   cds             326..5186
FT                   /*tag= a
FT                   /product= Erythrocyte membrane protein
FT   misc_feature    518..520
FT                   /*tag= b
FT                   /transl_except= GTA encodes Tyrosine
FT   misc_feature    656..658
FT                   /*tag= c
FT                   /transl_except= ATT encodes Leucine
FT   misc_feature    3908..3910
FT                   /*tag= d
FT                   /transl_except= CAA encodes Cysteine
FT   misc_feature    4067..4069
FT                   /*tag= e
FT                   /transl_except= ACT not included in the polypeptide
FT                   translation i.e. no amino acid
FT                   encoded by this codon
FT   misc_feature    5009..5011
FT                   /*tag= f
FT                   /transl_except= TGT encodes Serine
FT   misc_feature    5489..5491
FT                   /*tag= g
FT                   /transl_except= CTG not included in the polypeptide
FT                   translation i.e. no amino acid
FT                   encoded by this codon
PN   WO9633736-A1.
PD   31-OCT-1996.
PF   26-APR-1996; U05798.
PR   27-APR-1995; US-430908.
PA   (AFFY-) AFFYMAX TECHNOLOGIES NV.
PI   Baruch DI,  Howard RJ,  Pasloske BL;
DR   WPI; 96-497376/49.
DR   P-PSDB; W00385.
PT   New Plasmodium falciparum erythrocyte membrane proteins - used to
PT   develop products for the diagnosis, treatment or prevention of
PT   malaria parasite infections
PS   Disclosure; Figure 12; 149pp; English.
CC   A polypeptide comprising a Plasmodium falciparum (Pf) erythrocyte
```

```
CC   membrane protein 1 (PfEMP1) or active fragments or analogues of that
CC   protein can be used in the treatment or prevention of symptoms of a
CC   malaria parasite infection.  The polypeptides can inhibit, block or
CC   reverse the sequestration of erythrocytes in patients suffering from
CC   malaria. Nucleic acids derived from the PfEMP1 gene can be used as
CC   probes and primers to identify a Plasmodium falciparum parasite, the
CC   primers used to generate characteristic amplification patterns from
CC   different P. falciparum strains. Antibodies specifically
CC   immunoreactive with the PfEMP1 polypeptide or its fragments may be
CC   used in diagnosis of malaria infection. This sequence encodes a
CC   truncated version of the PfEMP1 protein of the MC type of Plasmodium
CC   falciparum encoded by a cDNA clone.  A full length genomic clone
CC   coding sequence is described in T41852.
SQ   Sequence   5511 BP;     2150 A;     913 C;     1189 G;     1259 T;

  Query Match           6.7%;  Score 20;  DB 24;  Length 5511;
  Best Local Similarity 100.0%;  Pred. No. 1.55e-02;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      98 tacatatatacatacataca 117
           ||||||||||||||||||||
Cp     197 TACATATATACATACATACA 178


RESULT    4
ID   T41852 standard; DNA; 9789 BP.
AC   T41852;
DT   20-FEB-1997  (first entry)
DE   cDNA encoding Plasmodium falciparum erythrocyte membrane protein.
KW   Plasmodium falciparum; erythrocyte membrane protein; malaria;
KW   detection; identification; treatment; prevention; parasite; ss.
OS   Plasmodium falciparum MC type.
FH   Key             Location/Qualifiers
FT   cds             326..9497
FT                   /*tag=  a
FT                   /product= Erythrocyte membrane protein
FT   misc_feature    518..520
FT                   /*tag=  b
FT                   /transl_except= GTA encodes Tyrosine
FT   misc_feature    656..658
FT                   /*tag=  c
FT                   /transl_except= ATT encodes Leucine
FT   misc_feature    2909..2911
FT                   /*tag=  d
FT                   /transl_except= AAC encodes Aspartic acid
FT   misc_feature    3461..3463
FT                   /*tag=  e
FT                   /transl_except= GAA encodes Glutamine
FT   misc_feature    5546..5548
FT                   /*tag=  f
FT                   /transl_except= CCT encodes Arginine
```



```
FT   misc_feature    6254..6256
FT                   /*tag=  g
FT                   /transl_except= AAT encodes Lysine
FT   misc_feature    6257..6259
FT                   /*tag=  h
FT                   /transl_except= ATA encodes Tyrosine
FT   misc_feature    6263..6265
FT                   /*tag=  i
FT                   /transl_except= AAC encodes Lysine
FT   misc_feature    6269..6271
FT                   /*tag=  j
FT                   /transl_except= TTC encodes Isoleucine
FT   misc_feature    6272..6274
FT                   /*tag=  k
FT                   /transl_except= ATA encodes Histidine
FT   misc_feature    6275..6277
FT                   /*tag=  l
FT                   /transl_except= ATT encodes Asparagine
FT   misc_feature    6278..6280
FT                   /*tag=  m
FT                   /transl_except= GGA encodes Tryptophan
FT   intron          7754..8478
FT                   /*tag=  n
PN   WO9633736-A1.
PD   31-OCT-1996.
PF   26-APR-1996; U05798.
PR   27-APR-1995; US-430908.
PA   (AFFY-) AFFYMAX TECHNOLOGIES NV.
PI   Baruch DI,  Howard RJ,  Pasloske BL;
DR   WPI; 96-497376/49.
DR   P-PSDB; W00384.
PT   New Plasmodium falciparum erythrocyte membrane proteins - used to
PT   develop products for the diagnosis, treatment or prevention of
PT   malaria parasite infections
PS   Disclosure; Figure 12; 149pp; English.
CC   A polypeptide comprising a Plasmodium falciparum (Pf) erythrocyte
CC   membrane protein 1 (PfEMP1) or active fragments or analogues of that
CC   protein can be used in the treatment or prevention of symptoms of a
CC   malaria parasite infection.  The polypeptides can inhibit, block or
CC   reverse the sequestration of erythrocytes in patients suffering from
CC   malaria. Nucleic acids derived from the PfEMP1 gene can be used as
CC   probes and primers to identify a Plasmodium falciparum parasite, the
CC   primers used to generate characteristic amplification patterns from
CC   different P. falciparum strains. Antibodies specifically
CC   immunoreactive with the PfEMP1 polypeptide or its fragments may be
CC   used in diagnosis of malaria infection. This sequence encodes the
CC   PfEMP1 protein of the MC type of Plasmodium falciparum.  An
CC   alternative, truncated version of the coding sequence (a cDNA clone)
CC   is given in T41853.
SQ   Sequence   9789 BP;     4061 A;     1393 C;     1837 G;     2498 T;
```



```
  Query Match           6.7%;  Score 20;  DB 24;  Length 9789;
  Best Local Similarity 100.0%;  Pred. No. 1.55e-02;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      98 tacatatatacatacataca 117
           ||||||||||||||||||||
Cp     197 TACATATATACATACATACA 178


RESULT    5
ID   T23920 standard; cDNA to mRNA; 128 BP.
AC   T23920;
DT   23-AUG-1996  (first entry)
DE   Human gene signature HUMGS05859.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 1486; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   128 BP;     44 A;     23 C;     27 G;     30 T;

  Query Match           6.4%;  Score 19;  DB 20;  Length 128;
  Best Local Similarity 100.0%;  Pred. No. 7.80e-02;
  Matches    19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```



```
Db      29 atatatacatacatacaca 47
           ||||||||||||||||||||
Cp     194 ATATATACATACATACACA 176


RESULT    6
ID   T21516 standard; cDNA to mRNA; 122 BP.
AC   T21516;
DT   29-AUG-1996  (first entry)
DE   Human gene signature HUMGS02889.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 904; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   122 BP;     45 A;     8 C;     17 G;     51 T;

  Query Match           6.0%;  Score 18;  DB 20;  Length 122;
  Best Local Similarity 100.0%;  Pred. No. 3.77e-01;
  Matches    18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      20 tatgtatgtatatatgta 37
           ||||||||||||||||||
```

```
Qy    180 TATGTATGTATATATGTA 197

RESULT    7
ID   Q57016 standard; DNA; 2891 BP.
AC   Q57016;
DT   31-AUG-1994  (first entry)
DE   PKC delta.
KW   110 kD catalytic subunit; phosphatidyl inositol 3-kinase;
KW   transformation; Schizosaccharomyces pombe; nmt promoter; thiamine;
KW   PtdIns 3-kinase; assay; detection; cell growth; regulation; cancer;
KW   blood vessel plaques; ss.
OS   Bos taurus.
PN   WO9403609-A.
PD   17-FEB-1994.
PF   05-AUG-1993; G01651.
PR   05-AUG-1992; GB-016654.
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
PI   Goode NT,  Nurse PM,  Parker PJJ,  Waterfield MD;
DR   WPI; 94-065697/08.
PT   Eukaryotic cells transformed with mammalian phospholipid or
PT   protein kinase DNA - useful in assays for compounds involved in
PT   cell growth regulation and for treating cancers
PS   Disclosure; Page 42-44; 71pp; English.
CC   The sequences given in Q57014-17 encode protein kinase C (PKC)
CC   epsilon, gamma, delta and nu respectively.  These sequences were
CC   transformed into Schizosaccharomyces pombe cells under the regulatory
CC   control of the nmt promoter in an embodiment of the invention.  In the
CC   presence of thiamine the promoter is inactive and the cells carrying
CC   the PKC plasmids grow as the parental strain.  In the absence
CC   of thiamine the nmt promoter functions and the PKC is induced.  PKC
CC   activity is substantially increased under these conditions.  Cells
CC   containing constructs such as this, are useful in assays for detecting
CC   compounds involved in cell growth regulation.  It is also used as the
CC   basis for detecting compounds for treating cancers and the formation
CC   of blood vessel plaques.
SQ   Sequence   2891 BP;    718 A;    770 C;    751 G;    652 T;

  Query Match           6.0%;  Score 18;  DB 10;  Length 2891;
  Best Local Similarity 100.0%;  Pred. No. 3.77e-01;
  Matches    18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2794 tgtatgtatgtatatatg 2811
          ||||||||||||||||||
Qy    178 TGTATGTATGTATATATG 195


RESULT    8
ID   Q01758 standard; cDNA; 2909 BP.
AC   Q01758;
DT   03-AUG-1990  (first entry)
```



```
DE   cDNA sequence of rat C kinase delta
KW   Rat phosphorylating enzyme C delta; reagent; diagnostic chemical;
KW   rat C kinase delta.
OS   Rat.
FH   Key             Location/Qualifiers
FT   cds             CDS                363..2384
FT                   /tag= a
PN   J02000433-A.
PD   05-JAN-1990.
PF   08-OCT-1987; 252506.
PR   08-OCT-1987; JP-252506.
PA   (TAKE) Takeda Chemical Ind Kk.
DR   WPI; 90-047984/07.
DR   P-PSDB; R05228.
PT   Rat C kinase delta enzyme -
PT   expressed in animal cells and used as a reagent or diagnostic
PT   chemical
PS   Disclosure;Fig 1; 23pp; Japanese.
CC   It and its encoded rat phosphorylating enzyme C delta are new.  Also new
CC   are a transformant contg. a vector with it and prepn. of its polypeptide
CC   by culturing the transformant.  Its encoded polypeptide can be used as a
CC   reagent and a diagnostic chemical.
SQ   Sequence   2909 BP;    721 A;    773 C;    757 G;    658 T;

  Query Match           6.0%;  Score 18;  DB 1;  Length 2909;
  Best Local Similarity 100.0%;  Pred. No. 3.77e-01;
  Matches    18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2803 tgtatgtatgtatatatg 2820
          ||||||||||||||||||
Qy    178 TGTATGTATGTATATATG 195


RESULT    9
ID   Q38525 standard; cDNA; 3095 BP.
AC   Q38525;
DT   14-JUL-1993  (first entry)
DE   Human CENP-C antigen gene.
KW   Autoimmune rheumatic diseases; CREST syndrome; classification;
KW   diagnosis; immunoassay; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             3..2942
FT                   /*tag=  a
PN   WO9305151-A.
PD   18-MAR-1993.
PF   11-SEP-1992; U07649.
PR   12-SEP-1991; US-759001.
PA   (UYJO ) UNIV JOHNS HOPKINS.
PI   Earnshaw WC, Saitoh H, Tomkiel JE;
DR   WPI; 93-100976/12.
```



```
DR   P-PSDB; R34783.
PT   DNA encoding human CENP-C and its antibodies - for detecting
PT   anti-centromere antibodies in patients with auto-immune diseases
PS   Claim 1; Page 40; 68pp; English.
CC   The sequence is that of a gene encoding the human CENP-C antigen
CC   which can be used to produce the human CENP-C polypeptide in large
CC   amts. Thus the polypeptide is available to more accurately classify
CC   patients with such autoimmune rheumatic diseases as CREST syndrome,
CC   and commercially useful quantities are available for use in
CC   diagnostic systems.
SQ   Sequence   3095 BP;    1161 A;    540 C;    608 G;    786 T;

  Query Match           6.0%;  Score 18;  DB 6;  Length 3095;
  Best Local Similarity 100.0%;  Pred. No. 3.77e-01;
  Matches    18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2964 tatgtatgtatatatgta 2981
          ||||||||||||||||||
Qy    180 TATGTATGTATATATGTA 197


RESULT   10
ID   Q24856 standard; cDNA; 3959 BP.
AC   Q24856;
DT   22-NOV-1992  (first entry)
DE   Sequence of T-cadherin 1 cDNA in clone 266.
KW   Cell adhesion molecule; nerve regeneration; cadherin; cancer;
KW   therapy; diagnosis; ss.
OS   Gallus domesticus.
FH   Key             Location/Qualifiers
FT   cds             45..110
FT                   /*tag=  a
FT   mat_peptide     111..2183
FT                   /*tag=  b
FT   polya_signal    3550..3555
FT                   /*tag=  c
FT   polya_signal    3687..3692
FT                   /*tag=  d
PN   WO9208731-A.
PD   29-MAY-1992.
PF   30-OCT-1991; U08022.
PR   30-OCT-1990; US-607293.
PA   (LJOL-) LA JOLLA CANCER RES FOUND.
PI   Ranscht B;
DR   WPI; 92-200133/24.
DR   P-PSDB; R27823.
PT   New T-cadherin polypeptide(s) and corresp. nucleic acid sequences
PT   - useful for treatment and diagnosis of cancer, and for repair of
PT   neuronal injury
PS   Disclosure; Fig 2A; 51pp; English.
CC   The inventors claim a purified T-cadherin polypeptide (T-cad)
```



```
CC   (T = truncated) and the nucleic acid sequence encoding it.  T-cad
CC   refers to polypeptides having substantially the amino acid sequence
CC   in R27823 and R27824, and which are cross-reactive with antibodies
CC   reactive with T-cad, but not with N-cadherin, E-cadherin, P-cadherin
CC   and L-CAM.  Two forms of T-cadherin identified in the invention
CC   lack the cytoplasmic portion found in all other members of the
CC   cadherin family.  One form of T-cadherin, referred to as T-cad 1,
CC   appears to be anchored to the membrane through a glycosyl
CC   phosphatidylinositol (GPI) linkage.  The other form, T-cad 2, is
CC   predicted by the cDNA to contain sequences for a hydrophobic domain
CC   followed by 5 cytoplasmic AAs.  From preliminary transfection of this
CC   cDNA into COS-cells, it is likely that this form is also GPI-linked.
CC   The extracellular portion of both forms are identical and contain
CC   structural features characteristic of the cadherin family.  The two
CC   forms differ in their COOH-terminal region in that T-cad 2 cDNA encodes
CC   five additional AAs.  The cDNA was isolated from a cDNA library
CC   generated from embryonic day 13 chicken brain.
SQ   Sequence   3959 BP;    1235 A;    826 C;    814 G;    1084 T;

  Query Match           6.0%;  Score 18;  DB 4;  Length 3959;
  Best Local Similarity 100.0%;  Pred. No. 3.77e-01;
  Matches    18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   3343 atatatacatacatacac 3360
          ||||||||||||||||||
Cp    194 ATATATACATACATACAC 177


RESULT   11
ID   N60970 standard; DNA; 21126 BP.
AC   N60970;
DT   28-OCT-1991  (first entry)
DE   TL-DNA region of A.rhizogene agropine-type plasmid pRiHRI.
KW   Ri plasmid; plant plasmid.
OS   Agrobacterium rhizogenes HRI and A4.
PN   EP-204590-A.
PD   10-DEC-1986.
PF   21-APR-1986; 400855.
PR   22-APR-1985; US-725368.
PA   (INRG ) INST NAT RECH AGRON.
PI   Slightom JL, Tepfer DA;
DR   WPI; 86-327378/50.
PT   Plant cell genetic modification - using pRi T-DNA promoter to
PT   control expression of heterologous foreign structure genes, e.g.
PT   for phaseolin.
PS   Disclosure; Fig 2; 86pp; English.
CC   The TL-DNA sequence is that of an Agrobacterium Ri plasmid, it
CC   contains sixteen open reading frames, bounded by eukaryotic
CC   promoters, ribosome binding and polyadenylation site. The plasmid
CC   sequence may be used for genetic manipulation of plants, conveying
CC   phenotypes such as improoved resistance to environmental stress, disease
```

```
CC   and parasitism.
SQ   Sequence   21126 BP;    5359 A;    4901 C;    5308 G;    5558 T;

  Query Match           6.0%;  Score 18;  DB 2;  Length 21126;
  Best Local Similarity  94.7%;  Pred. No. 3.77e-01;
  Matches   18;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db   6879 caggtcttcatttatatgt 6897
         ||||| ||||||||||||
Qy    272 CAGGTNTTCATTTATATGT 290


RESULT   12
ID   T62455 standard; DNA; 830 BP.
AC   T62455;
DT   20-OCT-1997  (first entry)
DE   Lambda 5h-1 clone containing AAV integration site.
KW   Protein phosphatase 2C-alpha; PP2C-alpha; polymorphism;
KW   phosphorylation; tumour marker; cancer; therapy; diagnosis;
KW   gene therapy; vector; AAV; ss.
OS   Chimaeric adeno associated virus;
OS   Chimaeric bacteriophage lambda.
PN   WO9710796-A2.
PD   27-MAR-1997.
PF   30-AUG-1996; IB1021.
PR   01-SEP-1995; US-003114.
PA   (UYRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.
PI   Lavi S;
DR   WPI; 97-202595/18.
PT   Detection of protein phosphatase 2C activity - as a marker for the
PT   diagnosis and therapy of cancers and diseases due to aberrant
PT   phosphorylation caused by altered PP2C activity
PS   Example 10; Page 77; 101pp; English.
CC   2 Lambda clones containing the adeno associated virus (AAV)
CC   integration site in cells were prepd.  One was derived from the
CC   Chinese hamster cell CO60 designated lambda SL9-1, of which parts
CC   were sequenced (T62453 and T62454).  A further sequence, AN8T7
CC   (T62456), was derived from plasmid pSL9-8.  The second lambda phage
CC   was cloned from cell line DA3J7.  Portions were subcloned to
CC   plasmids and a partial sequenced, designated 5h-1 (T62455), was
CC   shown to be homologous to AN8T7.  The role of integrated AAV in the
CC   modulation of protein phosphatase 2C-alpha expression was
CC   investigated.
SQ   Sequence   830 BP;    287 A;    148 C;    150 G;    231 T;

  Query Match           5.7%;  Score 17;  DB 33;  Length 830;
  Best Local Similarity 100.0%;  Pred. No. 1.74e+00;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    197 atatatacatacataca 213
          ||||||||||||||||||
```



```
Cp    194 ATATATACATACATACA 178


RESULT   13
ID   V15094 standard; DNA; 2542 BP.
AC   V15094;
DT   20-MAY-1998  (first entry)
DE   Human apolipoprotein(a) gene 5'-regulatory region.
KW   Human; apolipoprotein(a) gene 5'-regulatory region; expression;
KW   screening; regulation; genetic engineering; gene therapy;
KW   atherosclerosis; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   misc_signal      170..175
FT                    /*tag= a
FT                    /label= IL-6
FT                    /note= "transcription regulatory element"
FT   misc_signal      303..308
FT                    /*tag= b
FT                    /label= IL-6
FT                    /note= "transcription regulatory element"
FT   CAAT_signal      532..536
FT                    /*tag= c
FT   misc_signal      562..567
FT                    /*tag= d
FT                    /label= IL-6
FT                    /note= "transcription regulatory element"
FT   misc_signal      696..701
FT                    /*tag= e
FT                    /label= IL-6
FT                    /note= "transcription regulatory element"
FT   misc_signal      1163..1171
FT                    /*tag= f
FT                    /label= IL-6
FT                    /note= "transcription regulatory element"
FT   CAAT_signal      1190..1194
FT                    /*tag= g
FT   misc_signal      1220..1228
FT                    /*tag= h
FT                    /label= IL-6RC
FT                    /note= "transcription regulatory element"
FT   misc_signal      1248..1252
FT                    /*tag= i
FT                    /label= CEBP-RC
FT                    /note= "transcription regulatory element"
FT   CAAT_signal      1262
FT                    /*tag= j
FT   TATA_signal      1267..1271
FT                    /*tag= k
FT   exon             1282..1491
FT                    /*tag= l
```



```
FT                    /number= 1
FT   misc_signal      1398..1403
FT                    /*tag= m
FT                    /label= IL-6
FT                    /note= "transcription regulatory element"
PN   US5721138-A.
PD   24-FEB-1998.
PF   15-MAY-1995; 441370.
PR   15-DEC-1992; US-991849.
PR   15-MAY-1995; US-441370.
PA   (STRD ) UNIV STANFORD.
PI   Lawn RM;
DR   WPI; 98-168413/15.
PT   Human apolipoprotein (a) gene promoter - useful in genetic
PT   engineering and gene therapy and in the treatment of atherosclerosis
PS   Claim 1; Column 13-16; 16pp; English.
CC   The present sequence represents human apolipoprotein(a) gene 5'
CC   regulatory region. The present invention also describes: (1) a vector
CC   containing human apolipoprotein(a) gene 5'-regulatory region DNA; (2)
CC   an isolated nucleotide sequence comprising at least 30 consecutive
CC   nucleotides of human apolipoprotein(a) gene 5'-regulatory region DNA
CC   or its complement; (3) a nucleotide sequence of at least 15 nucleotides
CC   capable of forming a DNA triplex with human apolipoprotein(a) gene 5'
CC   regulatory region DNA; and (4) a transfected mammalian cell containing
CC   human apolipoprotein(a) gene 5'-regulatory region DNA where the
CC   heterologous sequence codes for an enzyme. The new promoter sequence is
CC   useful in genetic engineering and gene therapy. The promoter can
CC   specifically be used for regulating the expression of apolipoprotein(a),
CC   which is useful in the treatment of atherosclerosis.
SQ   Sequence   2542 BP;    718 A;    495 C;    515 G;    814 T;

  Query Match           5.7%;  Score 17;  DB 39;  Length 2542;
  Best Local Similarity 100.0%;  Pred. No. 1.74e+00;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2108 tatacatacatacacac 2124
         ||||||||||||||||||
Cp    191 TATACATACATACACAC 175


RESULT   14
ID   Q43966 standard; DNA; 5837 BP.
AC   Q43966;
DT   25-OCT-1993  (first entry)
DE   Thermococcus litoralis thermostable DNA polymerase gene.
KW   Heat stability; DNA polymerisation; DNA amplification; ds.
OS   Thermococcus litoralis.
FH   Key             Location/Qualifiers
FT   misc_feature     1..1274
FT                    /*tag= a
FT                    /note= "region amplified for use as probe"
```



```
FT   misc_feature     1656..2660
FT                    /*tag= b
FT                    /note= "region amplified for use as probe"
FT   misc_feature     3069..3737
FT                    /*tag= c
FT                    /note= "region amplified for use as probe"
PN   EP-547359-A.
PD   23-JUN-1993.
PF   05-NOV-1992; 118965.
PR   18-DEC-1991; US-809822.
PA   (NEWE ) NEW ENGLAND BIOLABS INC.
PI   Comb DG,  Jack WE,  Kong H,  Kucera R,  Perler F;
DR   WPI; 93-198432/25.
PT   Purified thermostable DNA polymerase from Pyrococcus sp. - has
PT   increased 3'-5' proof reading exo-nuclease activity
PS   Example 2; Page 17-20; 26pp; English.
CC   Labelled DNA probes of 3 fragments (see Q43965) of the T.litoralis
CC   DNA polymerase gene (obtainable from ATCC 40795) were prepared using
CC   a random primer kit and used to screen a Pyrococcus genomic library.
CC   A partial sequence (see Q43965) for Pyrococcus sp. DNA polymerase was
CC   isolated which contains an intervening sequence within the conserved
CC   pol-alpha motif region III.
SQ   Sequence   5837 BP;    1947 A;    1030 C;    1332 G;    1528 T;

  Query Match           5.7%;  Score 17;  DB 7;  Length 5837;
  Best Local Similarity  94.4%;  Pred. No. 1.74e+00;
  Matches   17;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db   3061 atataaatgaagacctgc 3078
         |||||||||| ||||||
Cp    288 ATATAAATGAANACCTGC 271


RESULT   15
ID   Q43515 standard; DNA; 5837 BP.
AC   Q43515;
DT   21-JAN-1994  (first entry)
DE   T. litoralis DNA polymerase gene.
KW   DNA polymerase; intron; translation; bacteriophage NEB #619; E. coli;
KW   mutation; cell growth; endonuclease; self splicing; stability;
KW   expression; transformation; expression vector; ss.
OS   Thermococcus litoralis.
FH   Key             Location/Qualifiers
FT   cds              291..5399
FT                    /*tag= a
FT   exon             291..1775
FT                    /*tag= b
FT   intron           1776..3389
FT                    /*tag= c
FT                    /note= "Intervening sequence"
FT   exon             3390..3533
```

```
FT                   /*tag=  d
FT   intron          3534..4703
FT                   /*tag=  e
FT                   /note= "Intervening sequence"
FT   exon            4704..5396
FT                   /*tag=  f
PN   EP-547920-A.
PD   23-JUN-1993.
PF   18-DEC-1992; 311622.
PR   18-DEC-1991; US-811421.
PA   (NEWE ) NEW ENGLAND BIOLABS INC.
PI   Comb DG,  Jack WE,  Kucera R,  Perler F;
DR   WPI; 93-198889/25.
DR   P-PSDB; R38095.
PT   Recombinant thermostable DNA polymerase from archaebacteria -
PT   expressed efficiently by removal of intervening introns
PS   Claim 1; Fig 6; 63pp; English.
CC   This sequence represents the T. litoralis DNA polymerase gene.  The
CC   introns marked are as decribed in the specification, however the
CC   described protein sequence appears to be translated through these
CC   intronic sequences.  This T. litoralis sequence may be derived from
CC   bacteriophage NEB #619 on an approx. 14kb BamHI restriction fragment.
CC   When this gene is transformed in to E. coli and expressed in its
CC   entirety, it is found to be unstable.  There is a high frequency of
CC   mutation in the polymerase gene, cell growth is slow and there is some
CC   degree of cell mortality.  This instability is due in part to the
CC   presence of the two introns.  It is thought that the intervening DNA
CC   in this sequence encodes a separate protein which may be spliced out
CC   at the protein level.  The second, 1170bp, intron has been found to
CC   code for an endonuclease which self splices out in E. coli.  To
CC   improve stability of the DNA when it is expressed in E. coli, the
CC   specification states that it is desirable to delete the introns
CC   before transformation of the expression vector into E. coli.
SQ   Sequence   5837 BP;    1947 A;    1027 C;    1336 G;    1527 T;

  Query Match           5.7%;  Score 17;  DB 7;  Length 5837;
  Best Local Similarity  94.4%;  Pred. No. 1.74e+00;
  Matches    17;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db   3061 atataaatgaagacctgc 3078
           |||||||||||| ||||||
Cp    288 ATATAAATGAANACCTGC 271


RESULT   16
ID   Q66293 standard; DNA; 5837 BP.
AC   Q66293;
DT   03-FEB-1995  (first entry)
DE   Vent DNA polymerase gene.
KW   Target protein; controllable intervening protein sequence;
KW   CIVPS; Vent IVPS; Deep Vent IVPS; Thermococcus litoralis;
```



```
KW   Pyrococcus sp.; external protein sequence; EPS; ds.
OS   Thermococcus litoralis.
FH   Key             Location/Qualifiers
FT   misc_rna        1773..3386
FT                   /*tag=  a
FT                   /label= CIVPS1
FT   misc_rna        3534..4703
FT                   /*tag=  b
FT                   /label= CVIPS2
PN   EP-602899-A.
PD   22-JUN-1994.
PF   09-DEC-1993; 309920.
PR   09-DEC-1992; US-004139.
PR   03-NOV-1993; US-146885.
PA   (NEWE ) NEW ENGLAND BIOLABS INC.
PI   Comb DG,  Hodges RA,  Jack WE,  Noren CJ,  Perler FB;
PI   Xu M;
DR   WPI; 94-193433/24.
PT   Modified proteins - comprising target protein and controllable
PT   intervening protein sequence.
PS   Disclosure; Page 31-34; 73pp; English.
CC   New modified proteins comprise a target protein and a controllable
CC   intervening protein sequence (CIVPS), the CIVPS being capable of
CC   excision or cleavage under predetermined conditions. Pref. CIVPSs
CC   include CIVPS1 and 2 from T.litoralis (=Vent IVPS1 and 2 or IVS1
CC   and 2) and CIVPS 3 from Pyrococcus sp. (=Deep Vent IVPS1 or IVS1).
CC   These CIVPSs are capable of excision, i.e. removal via protein
CC   splicing, from modified proteins upon an increase in temp.
CC   This can be used to inactivate protein activity and also
CC   to restore the activity of an inactivated protein.
SQ   Sequence   5837 BP;    1947 A;    1033 C;    1330 G;    1527 T;

  Query Match           5.7%;  Score 17;  DB 11;  Length 5837;
  Best Local Similarity  94.4%;  Pred. No. 1.74e+00;
  Matches    17;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db   3061 atataaatgaagacctgc 3078
           |||||||||| ||||||
Cp    288 ATATAAATGAANACCTGC 271


RESULT   17
ID   Q14461 standard; DNA; 5837 BP.
AC   Q14461;
DT   21-JAN-1992  (first entry)
DE   Polymerase gene.
KW   Thermostable; ss.
OS   Thermococcus litoralis strain NS-C (DSM 5473).
FH   Key             Location/Qualifiers
FT   cds             291..5397
FT                   /*tag=  a
```



```
FT   intron          1776..3389
FT                   /*tag=  b
FT   misc_feature    1..1274
FT                   /*tag=  c
FT                   /note= "claim 14"
FT   misc_feature    1269..2856
FT                   /*tag=  d
FT                   /note= "claim 19"
FT   misc_feature    2851..4771
FT                   /*tag=  e
FT                   /note= "claim 24"
PN   EP-455430-A.
PD   06-NOV-1991.
PF   26-APR-1991; 303787.
PR   26-APR-1990; US-513994.
PR   11-DEC-1990; US-626057.
PR   17-APR-1991; US-686340.
PA   (NEWE-) NEW ENGLAND BIOLABS.
PI   Comb DG, Perler F, Kucera R, Jack WE;
DR   WPI; 91-327286/45.
PT   New thermostable DNA polymerase from Thermococcus.littoralis -
PT   used for second strand synthesis in DNA cloning, sequencing and
PT   amplification.
PS   Claim 12; Fig 6; 43pp; English.
CC   The sequence was obtd. from a 14 kb BamHI restriction fragment of
CC   bacteriophage NEB6199 ATCC 40795). The sequence contains an intron
CC   or intervening sequence in DNA polymerase consensus region III.
CC   The left side of the T. litoralis homogy island III begins at nt.
CC   1737 and the right side can be picked up at nt. 3384.  The sequence
CC   can be mutated to delete the intron and used to express the poly-
CC   merase.  The enzyme can be used in cloning and PCR.
SQ   Sequence   5837 BP;    1947 A;    1027 C;    1336 G;    1527 T;

  Query Match           5.7%;  Score 17;  DB 3;  Length 5837;
  Best Local Similarity  94.4%;  Pred. No. 1.74e+00;
  Matches    17;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db   3061 atataaatgaagacctgc 3078
           |||||||||||| ||||||
Cp    288 ATATAAATGAANACCTGC 271


RESULT   18
ID   T17546 standard; DNA; 5837 BP.
AC   T17546;
DT   20-MAY-1996  (first entry)
DE   T. litoralis DNA polymerase gene.
KW   DNA polymerase; thermostable enzyme; amplification; sequencing;
KW   archaebacterium; polymerase chain reaction; PCR; ds.
OS   Thermococcus litoralis strain NS-C (DSM 5473).
FH   Key             Location/Qualifiers
```



```
FT   exon            1..1775
FT                   /*tag=  a
FT                   /codon_start= 291..293
FT   intron          1776..3389
FT                   /*tag=  b
FT                   /product= endonuclease
FT   exon            3390..3533
FT                   /*tag=  c
FT   intron          3534..4703
FT                   /*tag=  d
FT                   /product= endonuclease
FT   exon            4704..5837
FT                   /*tag=  e
PN   EP-701000-A2.
PD   13-MAR-1996.
PF   27-JUN-1995; 304504.
PR   06-JUL-1994; US-271364.
PA   (NEWE ) NEW ENGLAND BIOLABS INC.
PI   Perler FB,  Southworth M;
DR   WPI; 96-141024/15.
PT   New isolated thermostable DNA polymerase - obtd. from isolate 9 deg.
PT   N-7, useful for amplifying, detecting and/or cloning nucleic acid
PT   sequences
PS   Example 3; Fig 6; 90pp; English.
CC   A DNA-polymerase gene (T17546) was identified on a 14 kb BamHI
CC   fragment of phage NEB619 (ATCC 40795), obtd. by screening genomic
CC   libraries of Thermococcus litoralis NS-C, an archaebacterium
CC   isolated from a submarine thermal vent.  The gene codes for
CC   a thermostable DNA polymerase having 3' to 5' exonuclease
CC   activity and 3'-5' proofreading ability, and which gave good
CC   results in the PCR amplification of esp. very small amounts (e.g.
CC   2.8 fg) of DNA.  The gene has been cloned into Escherichia coli.
CC   Prodn. of large amounts of recombinant enzyme was complicated by
CC   the presence of 2 introns which both encode an endonuclease that
CC   is spliced out in E. coli.
SQ   Sequence   5837 BP;    1946 A;    1027 C;    1334 G;    1530 T;

  Query Match           5.7%;  Score 17;  DB 18;  Length 5837;
  Best Local Similarity  94.4%;  Pred. No. 1.74e+00;
  Matches    17;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db   3061 atataaatgaagacctgc 3078
           |||||||||| ||||||
Cp    288 ATATAAATGAANACCTGC 271


RESULT   19
ID   T70812 standard; DNA; 5837 BP.
AC   T70812;
DT   02-SEP-1997  (first entry)
DE   Vent DNA polymerase gene containing CIVPS1 and CIVPS2.
```

```
KW   DNA polymerase; Vent; controllable intervening protein sequence;
KW   IVPS; CIVPS1; CIVPS2; ds.
OS   Thermococcus litoralis.
FH   Key             Location/Qualifiers
FT   misc_rna        1773..3366
FT                   /*tag= a
FT                   /label= CIVPS1
FT   misc_rna        3534..4703
FT                   /*tag= b
FT                   /label= CIVPS2
FT   primer_bind     complement (3534..3556)
FT                   /*tag= c
FT                   /note= "Vent IVPS2 forward"
FT   primer_bind     4685..4706
FT                   /*tag= d
FT                   /note= "Vent IVPS2/Thr reverse"
PN   WO9701642-A1.
PD   16-JAN-1997.
PF   19-JUN-1996; U10545.
PR   28-JUN-1995; US-496297.
PR   29-DEC-1995; US-580555.
PA   (NEWE ) NEW ENGLAND BIOLABS INC.
PI   Chong SSC,  Comb DG,  Hodges RA,  Jack WE,  Noren CJ;
PI   Perler FB,  Xu M;
DR   WPI; 97-100213/09.
PT   Modified protein comprising target protein and controllable
PT   intervening protein - that can be removed by excision or cleavage
PT   under specific conditions, useful for prepn. and purificn. of
PT   recombinant proteins that are normally toxic to cells
PS   Disclosure; Page 103-106; 161pp; English.
CC   The Thermococcus litoralis Vent DNA polymerase gene (T70812)
CC   includes 2 controllable intervening protein sequence elements,
CC   CIVPS1 and CIVPS2, deposited in phage NEB 619 (ATCC 40795).  The
CC   Pyrococcus Deep Vent polymerase gene (T70813) contains another such
CC   element, CIVPS3.  A modified protein comprises a target protein
CC   (TP) and a CIVPS that can be removed by excision under specific
CC   conditions, i.e. an increase in temperature.  The CIVPS is inserted
CC   into, or fused to, the TP in such a way as to render the TP
CC   inactive, but activity is restored upon excision of the CIVPS.
CC   This can be used e.g. to control the enzymatic activity of a TP, to
CC   facilitate purification of a TP and to permit prodn. of a TP which
CC   in native form is toxic to the cells expressing it.  The TP may be
CC   an enzyme, toxin, cytokine, glycoprotein, growth factor etc.
SQ   Sequence   5837 BP;    1948 A;    1026 C;    1336 G;    1527 T;

  Query Match           5.7%;  Score 17;  DB 31;  Length 5837;
  Best Local Similarity  94.4%;  Pred. No. 1.74e+00;
  Matches   17;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db   3061 atataaatgaagacctgc 3078
          ||||||||||| ||||||
```



```
Cp    288 ATATAAATGAANACCTGC 271


RESULT   20
ID   Q34071 standard; DNA; 51 BP.
AC   Q34071;
DT   02-FEB-1993  (first entry)
DE   Microsatellite sequence from clone TGLA446.
KW   PCR; selection; primers; OPTIPRIM; breeding; cattle; parentage;
KW   genetic mapping; traits; amplification; ss.
OS   Bos taurus.
PN   WO9213102-A.
PD   06-AUG-1992.
PF   15-JAN-1992; U00340.
PR   15-JAN-1991; US-642342.
PA   (GENM-) GENMARK.
PI   Georges M, Massey JM;
DR   WPI; 92-284684/34.
PT   Polymorphic bovine DNA markers - used in genetic identification,
PT   gene mapping, and selective breeding
PS   Table 7; Page 359; 517pp; English.
CC   The sequence is that of a bovine microsatellite sequence obtd. by
CC   screening a library of bovine MboI DNA fragments of between
CC   250 and 500 bp with an (AC)15 and a (TC)15 oligonucleotide probe.
CC   One out of 50 clones cross-hybridised.  Assuming independent
CC   distribution of microsatellites and MboI sites, the frequency of
CC   (T6)n >9 microsatellites in the bovine genome is estimated at >100,
CC   000.  The sequence information for ca. 230 such bovine microsatellites
CC   is summarised in the specification and indexed herein (see below).
CC   The sequences upstream and downstream of the microsatellite sequence
CC   were used to generate the required PCR primers for in vitro
CC   amplification of the corresp. microsatellite (using the program
CC   OPTIPRIM).  The microsatellites may be used to identify individuals,
CC   for parentage testing, and in the genetic mapping of economic trait
CC   loci, or genes involved the determinism of economically important
CC   traits esp. in cattle, to allow selective breeding.
CC   See also Q33501-34437.
SQ   Sequence   51 BP;    5 A;    0 C;    17 G;    29 T;

  Query Match           5.4%;  Score 16;  DB 5;  Length 51;
  Best Local Similarity 100.0%;  Pred. No. 7.71e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     23 gtgtgtatgtatgtat 38
          ||||||||||||||||
Qy    175 GTGTGTATGTATGTAT 190


RESULT   21
ID   Q55238 standard; DNA; 518 BP.
AC   Q55238;
```



```
DT   14-JUL-1994  (first entry)
DE   Grapevine ribosomal clone IN36GT.
KW   Ribosome; grapevine; Vitis; IGS region; rDNA; polymorphism;
KW   grape cultivar; probe; primer; detection; ds.
OS   Vitis vinifera.
FH   Key             Location/Qualifiers
FT   primer_bind     1..23
FT                   /*tag= a
FT   primer_bind     complement (192..214)
FT                   /*tag= b
PN   WO9401580-A.
PD   20-JAN-1994.
PF   30-JUN-1993; AU0320.
PR   03-JUL-1992; AU-003330.
PA   (CSIR ) COMMONWEALTH SCI & IND RES ORG.
PI   Scott NS,  Thomas MR;
DR   WPI; 94-035083/04.
PT   Novel ribosome DNA probe sequences - for the accurate identification
PT   of grape cultivars
PS   Claim 26; Fig 5h; 55pp; English.
CC   The sequences given in Q55231-50 are derived from ribosomes of the
CC   grapevine genus Vitis.  These sequences represent the IGS region of
CC   the ribosomal (r)DNA repeat and contain polymorphisms.  These
CC   polymorphisms may be used in a method for the identification of
CC   different grape cultivars.  These clones contain simple repeat
CC   sequences and were identified in a genomic library of grapevine DNA
CC   using simple di-, tri- or tetra- nucleotide repeats such as (AT)8,
CC   (GT)10, (CGT)10 and such like as probes.
SQ   Sequence   518 BP;    149 A;    69 C;    99 G;    201 T;

  Query Match           5.4%;  Score 16;  DB 10;  Length 518;
  Best Local Similarity 100.0%;  Pred. No. 7.71e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    148 gtgtgtatgtatgtat 163
          ||||||||||||||||
Qy    175 GTGTGTATGTATGTAT 190


RESULT   22
ID   T72049 standard; DNA; 1260 BP.
AC   T72049;
DT   18-AUG-1997  (first entry)
DE   Sequence around marker 24E2 in HH region of chromosome 6p2.1.
KW   Primer; polymerase chain reaction; amplify; hereditary haemochromatosis;
KW   HH; mutation; HH-associated allele; base-pair polymorphism; HHP-1;
KW   HHP-19; HHP-29; microsatellite repeat allele; genetic marker;
KW   interferon treatment; hepatitis C infection; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   repeat_region   591..622
```



```
FT                   /*tag= a
FT                   /note= "Dinucleotide repeat region"
PN   WO9635802-A1.
PD   14-NOV-1996.
PF   06-MAY-1996; U06352.
PR   08-MAY-1995; US-436074.
PR   15-NOV-1995; US-559302.
PR   09-FEB-1996; US-599252.
PA   (MERC-) MERCATOR GENETICS INC.
PI   Drayna DT,  Feder JN,  Gnirke A,  Kimmel BE,  Thomas WJ;
PI   Wolff RK;
DR   WPI; 96-518690/51.
PT   Determn. of the common hereditary haemochromatosis gene mutation -
PT   using primers based on novel microsatellite repeat flanking
PT   sequences or on base-pair polymorphisms HHP-1, HHP-19 or HHP-29
PS   Claim 24; Fig 1E; 67pp; English.
CC   The sequences given in T72045-67 represent portions of the genome
CC   surrounding several markers of the invention.  The markers were
CC   identified using the series of primer pairs given in T71973-2044
CC   which were used to determine the presence or absence of the common
CC   hereditary haemochromatosis (HH) gene mutation in an individual.  The
CC   method comprised assessing genomic DNA from an individual for the
CC   presence or absence of the HH-associated allele of the single base-pair
CC   polymorphism HHP-1, HHP-19 or HHP-29, and/or at least one non-optional
CC   marker comprising the following microsatellite repeat alleles of group
CC   A and optionally of group B:
CC   Group A: 1D9, 18B4, 1A2, 1E4, 24E2, 2B8, 3321-1, 4073-1, 4440-1, 4440-2,
CC   731-1, 5091-1, 3216-1, 4072-2, 950-1, 950-2, 950-3, 950-4, 950-5, 950-6,
CC   950-8, 63-1, 63-2, 63-3, 65-1, 65-2, 373-8, 373-29, 68-1, 241-6, 241-29;
CC   Group B: D6S464, D6S306, D6S258, D6S265, D6S105 and D6S1001.
CC   The absence of the genotype indicates the likelihood of the presence of
CC   the HH mutation.  Knowledge of the new genetic markers allows the
CC   definition of genotypes characteristic of heterozygous carriers and
CC   homozygotes having a HH mutation in their genomic DNA. The potential for
CC   HH in an individual interferes with the effectiveness of interferon
CC   treatment for hepatitis C infection.  By diagnosing this potential, the
CC   responsiveness of interferon treatment may be evaluated.
CC   Note:  Some of the sequences given in this specification are unclear due
CC   to poor print quality.  This sequence is derived from the related patent,
CC   WO9635803.
SQ   Sequence   1260 BP;    410 A;    288 C;    203 G;    359 T;

  Query Match           5.4%;  Score 16;  DB 30;  Length 1260;
  Best Local Similarity 100.0%;  Pred. No. 7.71e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    152 atacatacatacacac 167
          ||||||||||||||||
Cp    190 ATACATACATACACAC 175
```

```
RESULT   23
ID   T43929 standard; DNA; 1260 BP.
AC   T43929;
DT   18-AUG-1997  (first entry)
DE   Sequence around marker 24E2 in HH region of chromosome 6p2.1.
KW   Primer; polymerase chain reaction; amplify; hereditary haemochromatosis;
KW   HH; mutation; HH-associated allele; base-pair polymorphism; HHP-1;
KW   HHP-19; HHP-29; microsatellite repeat allele; genetic marker;
KW   interferon treatment; hepatitis C infection; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   repeat_region    591..622
FT                    /*tag= a
FT                    /note= "Dinucleotide repeat region"
PN   WO9635803-A1.
PD   14-NOV-1996.
PF   08-MAY-1996; U06583.
PR   08-MAY-1995; US-436074.
PR   15-NOV-1995; US-559302.
PR   09-FEB-1996; US-599252.
PA   (MERC-) MERCATOR GENETICS INC.
PI   Drayna DT,  Feder JN,  Gnirke A,  Kimmel BE,  Thomas WJ;
PI   Wolff RK;
DR   WPI; 96-518691/51.
PT   Diagnosing and genotyping of hereditary haemochromatosis (HH)  -
PT   using primers to detect specific polymorphisms of the HH gene on
PT   chromosome 6p2.1 or novel microsatellite markers
PS   Claim 24; Fig 1E; 67pp; English.
CC   The sequences given in T43925-55 represent portions of the genome
CC   surrounding several markers of the invention.  The markers were
CC   identified using the series of primer pairs given in T71901-72
CC   which were used to determine the presence or absence of the common
CC   hereditary haemochromatosis (HH) gene mutation in an individual.  The
CC   method comprised assessing genomic DNA from an individual for the
CC   presence or absence of the HH-associated allele of the single base-pair
CC   polymorphism HHP-1, HHP-19 or HHP-29, and/or at least one non-optional
CC   marker comprising the following microsatellite repeat alleles of group
CC   A and optionally of group B:
CC   Group A: 19D9, 18B4, 1A2, 1E4, 24E2, 2B8, 3321-1, 4073-1, 4440-1, 4440-2,
CC   731-1, 5091-1, 3216-1, 4072-2, 950-1, 950-2, 950-3, 950-4, 950-5, 950-6,
CC   950-8, 63-1, 63-2, 63-3, 65-1, 65-2, 373-8, 373-29, 68-1, 241-6, 241-29;
CC   Group B: D6S464, D6S306, D6S258, D6S265, D6S105 and D6S1001.
CC   The absence of the genotype indicates the likelihood of the presence of
CC   the HH mutation.  Knowledge of the new genetic markers allows the
CC   definition of genotypes characteristic of heterozygous carriers and
CC   homozygotes having a HH mutation in their genomic DNA. The potential for
CC   HH in an individual interferes with the effectiveness of interferon
CC   treatment for hepatitis C infection.  By diagnosing this potential, the
CC   responsiveness of interferon treatment may be evaluated.
SQ   Sequence   1260 BP;    410 A;    288 C;    203 G;    359 T;
```



```
  Query Match           5.4%;  Score 16;  DB 30;  Length 1260;
  Best Local Similarity 100.0%;  Pred. No. 7.71e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    152 atacatacatacacac 167
          ||||||||||||||||
Cp    190 ATACATACATACACAC 175


RESULT   24
ID   T67066 standard; cDNA; 2544 BP.
AC   T67066;
DT   05-AUG-1997  (first entry)
DE   Human alpha3(IX) collagen cDNA.
KW   Type IX collagen; alpha3(IX) collagen; gene therapy;
KW   autologous cell therapy; rheumatoid arthritis; osteoarthritis;
KW   reactive arthritis; autoimmune hearing loss; cartilage;
KW   inflammation; bursitis; corneal disease; parasitic disease;
KW   ankylosing spondylitis; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             46..2100
FT                   /*tag= a
PN   WO9718227-A1.
PD   22-MAY-1997.
PF   13-NOV-1996; IB1324.
PR   13-NOV-1995; US-555669.
PA   (UABR-) UAB RES FOUND.
PI   Brewton RG,  Mayne R;
DR   WPI; 97-289219/26.
DR   P-PSDB; W15287.
PT   DNA encoding human alpha3(IX) collagen protein - used to screen for
PT   collagen-related disease, also in gene or autologous cell therapy
PS   Claim 2; Fig 5; 54pp; English.
CC   A cDNA clone (T67055) codes for novel human alpha3(IX) collagen
CC   (W15287).  It was obtd. by PCR amplification of human chondrocyte
CC   cDNA and genomic DNA (see also T67067-77).  The cDNA clone can be
CC   expressed in host/vector systems, partic. Pichia pastoris or
CC   Hansenula polymorpha, or in transgenic animals with recovery of
CC   collagen from the milk.  Alpha3(IX) collagen polynucleotides can
CC   also be used diagnostically to detect aberrant gene expression in
CC   disease states, e.g. rheumatoid arthritis, osteoarthritis, reactive
CC   arthritis, autoimmune hearing disease, cartilage inflammation,
CC   parasitic disease; bursitis, corneal diseases and ankylosing
CC   spondylitis.  They can also be used therpeutically, in gene or
CC   autologous cell therapy, or as antisense molecules in cases of
CC   overexpression of alpha3(IX) collagen.
SQ   Sequence   2544 BP;    504 A;    796 C;    911 G;    333 T;

  Query Match           5.4%;  Score 16;  DB 30;  Length 2544;
  Best Local Similarity 100.0%;  Pred. No. 7.71e+00;
```



```
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    391 ggcaaaggcctccctg 406
          ||||||||||||||||
Qy    245 GGCAAAGGCCTCCCTG 260


RESULT   25
ID   Q50008 standard; DNA; 2563 BP.
AC   Q50008;
DT   26-APR-1994  (first entry)
DE   Aspergillus niger Sulphydryl oxidase (SOX) gene.
KW   Genomic sequence; Sulphydryl oxidase; scavenger molecule;
KW   disulphide bond formation; oxidation; free sulphydryl group; ds.
OS   Aspergillus niger.
FH   Key             Location/Qualifiers
FT   promoter        1..476
FT                   /*tag= a
FT   signal_peptide  477..533
FT                   /*tag= b
FT   exon            534..702
FT                   /*tag= c
FT                   /number= 1
FT   intron          703..769
FT                   /*tag= d
FT                   /number= 1
FT   exon            770..1573
FT                   /*tag= e
FT                   /number= 2
FT   intron          1574..1681
FT                   /*tag= f
FT                   /number= 2
FT   exon            1682..1827
FT                   /*tag= g
FT                   /number= 3
FT   cds             477..1830
FT                   /*tag= h
FT   3'utr           1831..2563
FT                   /*tag= i
FT                   /note= "3'-flanking sequences"
PN   EP-565172-A.
PD   13-OCT-1993.
PF   31-MAR-1993; 200922.
PR   10-APR-1992; EP-201027.
PA   (UNIL ) QUEST INT BV.
PI   Maat J,  Musters W,  Schaap PJ,  Stam H,  Van de Vondervoort PJI;
PI   Verbakel JMA,  Visser J;
DR   WPI; 93-322408/41.
DR   P-PSDB; R43074.
PT   DNA encoding sulphydryl oxidase - used for prodn. of enzyme,
PT   useful in baking and for removing off-flavour from milk or beer
```



```
PS   Claim 5; Fig 7; 42pp; English.
CC   A PCR product amplified using primers based on partial peptide
CC   fragments of purified A.niger SOX was used to screen an A.niger
CC   N400 genomic DNA library in lambda-EMBL4. A 2.5kb EcoRI fragment
CC   comprising the entire SOX gene was identified in positive plaques
CC   and was sequenced. Additional upstream sequence information was
CC   obtained from a 6.4kb NsiI fragment also present in positive
CC   plaques.
SQ   Sequence   2563 BP;    637 A;    728 C;    577 G;    621 T;

  Query Match           5.4%;  Score 16;  DB 8;  Length 2563;
  Best Local Similarity 100.0%;  Pred. No. 7.71e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1976 atacatacatacacac 1991
          ||||||||||||||||
Cp    190 ATACATACATACACAC 175


RESULT   26
ID   T97393 standard; DNA; 3224 BP.
AC   T97393;
DT   11-MAY-1998  (first entry)
DE   Aspergillus oryzae protease pepC gene.
KW   PepC gene; serine protease; PepE; AreA regulator; gene replacement;
KW   antisense; gene inactivation; protease-free host cell;
KW   recombinant protein; ss.
OS   Aspergillus oryzae strain IFO 4178.
FH   Key             Location/Qualifiers
FT   CDS             388..1997
FT                   /*tag= a
FT                   /note= "contains 2 introns"
FT   exon            388..756
FT                   /*tag= b
FT                   /number= 1
FT   intron          757..817
FT                   /*tag= c
FT                   /number= 1
FT   exon            818..1753
FT                   /*tag= d
FT                   /number= 2
FT   intron          1754..1814
FT                   /*tag= e
FT                   /number= 2
FT   exon            1815..1997
FT                   /*tag= f
FT                   /number= 3
PN   WO9722705-A1.
PD   26-JUN-1997.
PF   16-DEC-1996; DK0528.
PR   15-DEC-1995; DK-001428.
```

```
PA   (NOVO ) NOVO-NORDISK AS.
PI   Christensen T, Lehmbeck J;
DR   WPI; 97-341696/31.
DR   P-PSDB; W31629.
PT   New fungus lacking functional AreA and extracellular protease
PT   gene(s) - for high yield expression of industrial or therapeutic
PT   proteins, also new protease(s) from Aspergillus oryzae
PS   Example 2; Page 35-38; 80pp; English.
CC   This DNA sequence includes a coding region for the serine
CC   protease PepC (see W31629) of Aspergillus oryzae.  The pepC gene
CC   was cloned by cross-hybridisation with the Aspergillus niger pepC
CC   gene.  The invention relates to a new fungus in which: (a) the areA
CC   gene (see T97394) has been modified by recombinant DNA methods so
CC   that it cannot express a functional AreA activator; and (2) the
CC   extracellular protease pepE (see T97392) and/or pepC genes are
CC   inactivated so that functional enzyme is not produced.  The
CC   fungi have complete or partial deletions of the appropriate genes,
CC   or expression sequences that regulate the genes.  Alternatively,
CC   the genes are inactivated by antisense mechanisms or by gene
CC   replacement.  The new fungi are used for the recombinant production
CC   of industrial or therapeutic peptides and proteins, particularly
CC   enzymes (e.g. proteases, lipase, cutinase, cellulase or chymosin),
CC   insulin, growth hormone, glucagon, somatostatin, interferon,
CC   platelet-derived growth factor, factor VII, factor VIII, urokinase,
CC   tissue plasminogen activator, erythropoietin or thrombopoietin.
CC   The new fungi do not produce protease and so provide higher yields
CC   of recombinant protein because of reduced proteolytic degradation.
CC   The isolated pepE and pepC genes can also be used for production of
CC   recombinant Pep proteases.
SQ   Sequence   3224 BP;    749 A;    846 C;    735 G;    894 T;

  Query Match           5.4%;  Score 16;  DB 38;  Length 3224;
  Best Local Similarity 100.0%;  Pred. No. 7.71e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2923 agtgtgtatgtatgta 2938
          ||||||||||||||||
Qy    174 AGTGTGTATGTATGTA 189


RESULT   27
ID   T72882 standard; cDNA; 19124 BP.
AC   T72882;
DT   12-SEP-1997  (first entry)
DE   Plasmodium var-7 gene.
KW   DBL gene family; DABP; sialic acid binding protein; vaccine; therapy;
KW   Duffy binding like gene; Duffy antigen binding protein; erythrocyte;
KW   DABP; merozoite; malaria; var-1; var-2; var-3; var-7; immune response;
KW   Plasmodium; ss.
OS   Plasmodium vivax.
OS   Plasmodium falciparum.
```



```
FH   Key             Location/Qualifiers
FT   exon            7317..15139
FT                   /*tag= a
FT                   /number= 1
FT   intron          15140..16205
FT                   /*tag= b
FT                   /number= 1
FT   exon            16206..17552
FT                   /*tag= c
FT                   /number= 2
FT                   /note= "no stop codon given"
PN   WO9640766-A2.
PD   19-DEC-1996.
PF   07-JUN-1996; UO9508.
PR   07-JUN-1995; US-487826.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI   Chitnis C,  Miller LH,  Peterson DS,  Sim KL,  Su X;
PI   Wellems TE;
DR   WPI; 97-052231/05.
DR   P-PSDB; W22475.
PT   New malaria vaccines - contains cysteine-rich DBL family protein
PT   binding domains homologous domains of the Duffy and sialic acid
PT   binding proteins
PS   Claim 4; Page 56-61; 96pp; English.
CC   This sequence represents the var-7 gene of Plasmodium. Var-7 belongs to
CC   the Duffy binding like (DBL) family of genes which have homology to the
CC   Duffy antigen binding protein (DABP) and sialic acid binding protein
CC   (SABP) conserved regions (see T72889 and T72888 respectively). The var
CC   family of genes modulate cytoadherence and antigenic variation of
CC   Plasmodium infected erythrocytes. SABP and the Duffy antigen binding
CC   protein (DABP) are soluble proteins that appear in the culture
CC   supernatant after infected erythrocytes release merozoites. DABP and SABP
CC   mediate the binding of merozoites and schizonts to the erythrocyte
CC   surface. These proteins are necessary for erythrocyte invasion by the
CC   parasite. This sequence can be used in the compositions of the invention.
CC   The compositions are for the treatment and prevention of malaria, and
CC   comprise either a nucleotide sequence or encoded polypeptide of the
CC   var-1, var-2, var-3 or var-7 genes of the DBL gene family, a family of
CC   genes having homology with conserved regions of DABP and SABP. The
CC   compositions are used for the treatment and prevention of malaria. They
CC   are also used in the preparation of vaccines for inducing a protective
CC   immune response in a mammal to Plasmodium merozoites (especially
CC   Plasmodium falciparum or Plasmodium vivax).
SQ   Sequence   19124 BP;    7824 A;    2190 C;    2790 G;    6320 T;

  Query Match           5.4%;  Score 16;  DB 32;  Length 19124;
  Best Local Similarity 100.0%;  Pred. No. 7.71e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  16165 tacatatatacataca 16180
          ||||||||||||||||
```



```
Cp    197 TACATATATACATACA 182


RESULT   28
ID   V13059 standard; DNA; 118 BP.
AC   V13059;
DT   18-MAY-1998  (first entry)
DE   Interleukin 6 non-risk genotype region sequence SEQ ID NO:1.
KW   Interleukin 6; IL-6; genotype; osteoporosis; human; polymorphism; ss.
OS   Synthetic.
OS   Homo sapiens.
PN   WO9743446-A2.
PD   20-NOV-1997.
PF   16-MAY-1997; G01337.
PR   16-MAY-1996; GB-010281.
PA   (GEMI-) GEMINI INT HOLDINGS LTD.
PI   Grant SFA, Ralston SH;
DR   WPI; 98-008905/01.
PT   Determination of genotype of an IL-6 gene in an individual - used
PT   for diagnosis of or identification of a predisposition or
PT   susceptibility to a disease, particularly osteoporosis
PS   Claim 14; Page 4; 20pp; English.
CC   The present sequence represents an interleukin 6 non-risk genotype
CC   sequence from the present invention. The present invention describes an
CC   in vitro method of diagnosis of a disease in an animal. The method
CC   comprises determining the genotype of an interleukin-6 (IL-6) gene in
CC   the animal. The invention also describes: (1) a method of identifying an
CC   animal predisposed or susceptible to a disease, comprising determining
CC   the genotype of an IL-6 gene in the animal; (2) a composition for use in
CC   diagnosing a disease in an animal, the composition comprising one or
CC   more primer nucleic acid molecules adapted to amplify a portion of a
CC   3' flanking region of an IL-6 gene in the animal; and (3) a composition
CC   for use in identifying an animal predisposed or susceptible to a
CC   disease, the composition comprising one or more primer nucleic acid
CC   molecules adapted to amplify a portion of a 3' flanking region of an
CC   IL-6 gene in the animal. The method can be used for the diagnosis of
CC   or identifying predisposition or susceptibility to a disease,
CC   particularly osteoporosis.
SQ   Sequence   118 BP;    49 A;    7 C;    10 G;    52 T;

  Query Match           5.0%;  Score 15;  DB 39;  Length 118;
  Best Local Similarity 100.0%;  Pred. No. 3.23e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     36 tatgtatgtatatat 50
          |||||||||||||||
Qy    180 TATGTATGTATATAT 194


RESULT   29
ID   T26837 standard; cDNA to mRNA; 152 BP.
```



```
AC   T26837;
DT   14-NOV-1996  (first entry)
DE   Human gene signature HUMGS09087.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 2184; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   152 BP;    62 A;    13 C;    24 G;    49 T;

  Query Match           5.0%;  Score 15;  DB 23;  Length 152;
  Best Local Similarity 100.0%;  Pred. No. 3.23e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     81 acatatagtgtgtat 95
          |||||||||||||||
Qy    168 ACATATAGTGTGTAT 182


RESULT   30
ID   V13060 standard; DNA; 154 BP.
AC   V13060;
DT   18-MAY-1998  (first entry)
DE   Interleukin 6 non-risk genotype region sequence SEQ ID NO:2.
```

```
KW   Interleukin 6; IL-6; genotype; osteoporosis; human; polymorphism; ss.
OS   Synthetic.
OS   Homo sapiens.
PN   WO9743446-A2.
PD   20-NOV-1997.
PF   16-MAY-1997; G01337.
PR   16-MAY-1996; GB-010281.
PA   (GEMI-) GEMINI INT HOLDINGS LTD.
PI   Grant SFA, Ralston SH;
DR   WPI; 98-008905/01.
PT   Determination of genotype of an IL-6 gene in an individual - used
PT   for diagnosis of or identification of a predisposition or
PT   susceptibility to a disease, particularly osteoporosis
PS   Claim 14; Page 4; 20pp; English.
CC   The present sequence represents an interleukin 6 non-risk genotype
CC   sequence from the present invention. The present invention describes an
CC   in vitro method of diagnosis of a disease in an animal. The method
CC   comprises determining the genotype of an interleukin-6 (IL-6) gene in
CC   the animal. The invention also describes: (1) a method of identifying an
CC   animal predisposed or susceptible to a disease, comprising determining
CC   the genotype of an IL-6 gene in the animal; (2) a composition for use in
CC   diagnosing a disease in an animal, the composition comprising one or
CC   more primer nucleic acid molecules adapted to amplify a portion of a
CC   3' flanking region of an IL-6 gene in the animal; and (3) a composition
CC   for use in identifying an animal predisposed or susceptible to a
CC   disease, the composition comprising one or more primer nucleic acid
CC   molecules adapted to amplify a portion of a 3' flanking region of an
CC   IL-6 gene in the animal. The method can be used for the diagnosis of
CC   or identifying predisposition or susceptibility to a disease,
CC   particularly osteoporosis.
SQ   Sequence   154 BP;    62 A;    8 C;    15 G;    69 T;

  Query Match           5.0%;  Score 15;  DB 39;  Length 154;
  Best Local Similarity 100.0%;  Pred. No. 3.23e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     36 tatgtatgtatatat 50
          ||||||||||||||
Qy    180 TATGTATGTATATAT 194


RESULT   31
ID   T26811 standard; cDNA to mRNA; 175 BP.
AC   T26811;
DT   14-NOV-1996  (first entry)
DE   Human gene signature HUMGS09061.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
```



```
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 2179; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   175 BP;    39 A;    45 C;    32 G;    55 T;

  Query Match           5.0%;  Score 15;  DB 23;  Length 175;
  Best Local Similarity 93.8%;  Pred. No. 3.23e+01;
  Matches   15;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db     27 ggnaattcccctgctt 42
          || ||||||||||||
Cp    110 GGAAATTCCCCTGCTT 95


RESULT   32
ID   T24352 standard; cDNA to mRNA; 226 BP.
AC   T24352;
DT   22-SEP-1996  (first entry)
DE   Human gene signature HUMGS06383.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
```



```
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 1590; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   226 BP;    59 A;    38 C;    77 G;    46 T;

  Query Match           5.0%;  Score 15;  DB 21;  Length 226;
  Best Local Similarity 100.0%;  Pred. No. 3.23e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    209 aaaggatgaaatgtg 223
          ||||||||||||||||
Cp     50 AAAGGATGAAATGTG 36


RESULT   33
ID   Q77415 standard; DNA; 246 BP.
AC   Q77415;
DT   23-SEP-1994  (first entry)
DE   Human genome fragment. (Preferred)
KW   Brain; placenta; bone marrow; genetic analysis; gene mapping;
KW   detection; homology; human; adrenal tissue; ds.
OS   Homo sapiens.
PN   WO9401548-A.
PD   20-JAN-1994.
PF   13-JUL-1993; G01467.
PR   13-JUL-1992; GB-014857.
PA   (MEDI-) MEDICAL RES COUNCIL.
PI   Gross J,  Hadfield KM,  Howells D,  Kelly M,  Shaw D;
PI   Sibson DR, Starkey M;
DR   WPI; 94-035056/04.
```



```
PT   New nucleic acid fragment encoding gene products - can be used
PT   for genetic analysis and mapping
PS   Claim 1; Page 514-515; 616pp; English.
CC   Human nucleic acid fragments, isolated from brain, adrenal tissue,
CC   the placenta or bone marrow comprise any of: (A) a sequence
CC   selected from (Q76401-Q77613), (B) an allelic variation of a
CC   sequence as described in (A), or (C) a sequence complementary
CC   to (A) or (B).
CC   Preferred sequences exhibit no more than 90% homology to a human
CC   sequence known per se.
SQ   Sequence   246 BP;    80 A;    41 C;    36 G;    89 T;

  Query Match           5.0%;  Score 15;  DB 10;  Length 246;
  Best Local Similarity 93.8%;  Pred. No. 3.23e+01;
  Matches   15;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db     50 tatgtaaattgtgtgt 65
          ||||||||| ||||||
Qy    192 TATGTAAATNGTGTGT 207


RESULT   34
ID   Q77289 standard; DNA; 327 BP.
AC   Q77289;
DT   23-SEP-1994  (first entry)
DE   Human genome fragment (Preferred).
KW   Brain; placenta; bone marrow; genetic analysis; gene mapping;
KW   detection; homology; human; adrenal tissue; ds.
OS   Homo sapiens.
PN   WO9401548-A.
PD   20-JAN-1994.
PF   13-JUL-1993; G01467.
PR   13-JUL-1992; GB-014857.
PA   (MEDI-) MEDICAL RES COUNCIL.
PI   Gross J,  Hadfield KM,  Howells D,  Kelly M,  Shaw D;
PI   Sibson DR, Starkey M;
DR   WPI; 94-035056/04.
PT   New nucleic acid fragment encoding gene products - can be used
PT   for genetic analysis and mapping
PS   Claim 1; Page 448-449; 616pp; English.
CC   Human nucleic acid fragments, isolated from brain, adrenal tissue,
CC   the placenta or bone marrow comprise any of: (A) a sequence
CC   selected from (Q76401-Q77613), (B) an allelic variation of a
CC   sequence as described in (A), or (C) a sequence complementary
CC   to (A) or (B).
CC   Preferred sequences exhibit no more than 90% homology to a human
CC   sequence known per se.
SQ   Sequence   327 BP;    122 A;    42 C;    52 G;    111 T;

  Query Match           5.0%;  Score 15;  DB 11;  Length 327;
  Best Local Similarity 100.0%;  Pred. No. 3.23e+01;
```

```
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      59 tgtatgtatgtatat 73
           ||||||||||||||
Qy     178 TGTATGTATGTATAT 192


RESULT   35
ID   T23229 standard; cDNA to mRNA; 337 BP.
AC   T23229;
DT   30-AUG-1996  (first entry)
DE   Human gene signature HUMGS05027.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 1316; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   337 BP;    90 A;    68 C;    97 G;    79 T;

  Query Match           5.0%;  Score 15;  DB 20;  Length 337;
  Best Local Similarity 100.0%;  Pred. No. 3.23e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    162 gttggaaagcagggg 176
```



```
           ||||||||||||||||
Qy     89 GTTGGAAAGCAGGGG 103


RESULT   36
ID   Q60579 standard; cDNA; 342 BP.
AC   Q60579;
DT   16-MAR-1994  (first entry)
DE   Human brain Expressed Sequence Tag EST02589.
KW   Gene transcription product; genetic markers; tagging; in vivo;
KW   transcription; mapping; locations; chromosomes; chromosomal; ss.
OS   Homo sapiens.
PN   WO9316178-A.
PD   19-AUG-1993.
PF   12-FEB-1993; U01294.
PR   12-FEB-1992; US-837195.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
PI   Adams MD, Moreno RF, Venter CJ;
DR   WPI; 93-272882/34.
PT   Enriched oligonucleotides and corresp. sequences - used as
PT   markers for human genes transcribed in-vivo, facilitate tagging
PT   of most human genes
PS   Example 4; Page 357; 500pp; English.
CC   The Expressed Sequence Tag was isolated from a human brain cDNA
CC   library as part of a large set of ESTs which can be used as markers
CC   for human genes transcribed in vivo. They can be used to facilitate
CC   tagging of most human genes, for mapping locations of expressed genes
CC   on chromosomes, for individual or forensic identification, for mapping
CC   locations of disease-associated genes, for identification of tissue
CC   type, and for prepn. of antisense sequences, probes and constructs.
CC   EST02589 has a "poor" coding probability as evaluated using the
CC   coding-region prediction program CRM. See also Q59041-Q61440.
SQ   Sequence   342 BP;    105 A;    44 C;    66 G;    124 T;

  Query Match           5.0%;  Score 15;  DB 8;  Length 342;
  Best Local Similarity 100.0%;  Pred. No. 3.23e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    122 tatgtatgtatatat 136
          ||||||||||||||
Qy    180 TATGTATGTATATAT 194


RESULT   37
ID   Q59745 standard; cDNA; 354 BP.
AC   Q59745;
DT   16-MAR-1994  (first entry)
DE   Human brain Expressed Sequence Tag EST00694.
KW   Gene transcription product; genetic markers; tagging; in vivo;
KW   transcription; mapping; locations; chromosomes; chromosomal; ss.
OS   Homo sapiens.
```



```
PN   WO9316178-A.
PD   19-AUG-1993.
PF   12-FEB-1993; U01294.
PR   12-FEB-1992; US-837195.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
PI   Adams MD, Moreno RF, Venter CJ;
DR   WPI; 93-272882/34.
PT   Enriched oligonucleotides and corresp. sequences - used as
PT   markers for human genes transcribed in-vivo, facilitate tagging
PT   of most human genes
PS   Example 4; Page 224; 500pp; English.
CC   The Expressed Sequence Tag was isolated from a human brain cDNA
CC   library as part of a large set of ESTs which can be used as markers
CC   for human genes transcribed in vivo. They can be used to facilitate
CC   tagging of most human genes, for mapping locations of expressed genes
CC   on chromosomes, for individual or forensic identification, for mapping
CC   locations of disease-associated genes, for identification of tissue
CC   type, and for prepn. of antisense sequences, probes and constructs.
CC   EST00694 has a "poor" coding probability as evaluated using the
CC   coding-region prediction program CRM. See also Q59041-Q61440.
SQ   Sequence   354 BP;    115 A;    64 C;    48 G;    125 T;

  Query Match           5.0%;  Score 15;  DB 8;  Length 354;
  Best Local Similarity  93.8%;  Pred. No. 3.23e+01;
  Matches    15;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db     34 taaatggtgtgttaaa 49
          ||||| ||||||||||
Qy    196 TAAATNGTGTGTTAAA 211


RESULT   38
ID   T84999 standard; DNA; 396 BP.
AC   T84999;
DT   12-NOV-1997  (first entry)
DE   Type II topoisomerase database reference sequence SEQ ID NO:181.
KW   Detection; identification; TopoII; contamination; food; human;
KW   animal; plant; soil; water; PCR; polymerase chain reaction; ds.
OS   Spirocheata aurantia.
PN   US5645994-A.
PD   08-JUL-1997.
PF   05-JUL-1990; 548138.
PR   06-JUN-1995; US-470179.
PR   05-JUL-1990; US-548138.
PR   13-AUG-1993; US-106482.
PA   (UTAH ) UNIV UTAH RES FOUND.
PI   Huang WM;
DR   WPI; 97-362925/33.
PT   Detection and identification of organisms - using DNA primers to
PT   amplify signature segment of organism's type II topoisomerase
PS   Claim 6; Column 177-178; 114pp; English.
```



```
CC   A method has been produced for selectively amplifying DNA segments of
CC   one or more species of organisms in a sample. The method involves: (a)
CC   providing a database containing reference sequences, comprising a
CC   subunit sequence of a signature region of a macromolecule selected from
CC   a type II topoisomerase (TopoII), or a homologue, where each reference
CC   sequence is specific to a different species of a chosen group, and the
CC   macromolecule comprises 1st and 2nd conserved regions adjacently
CC   flanking the signature region; and (b) making an extract of DNA
CC   molecules, and selectively amplifying DNA segments of the signature
CC   region using a universal primer composition, comprising a primer
CC   constructed to bind a DNA encoding the macromolecule, to produce
CC   amplified DNA segments. The present sequence represents a DNA fragment
CC   containing a signature segment which is used in the database as a
CC   reference sequence for Spirocheata aurantia. The method can be
CC   used to identify all of the different organisms present in a single
CC   sample without using multiple probes. It can accurately distinguish
CC   among similar and related species, and can be used with, e.g.
CC   contaminated food products, tissue or fluid samples from diseased
CC   humans, animal or plants, soil samples and water samples from any
CC   source.
SQ   Sequence   396 BP;    82 A;    132 C;    116 G;    66 T;

  Query Match           5.0%;  Score 15;  DB 33;  Length 396;
  Best Local Similarity 100.0%;  Pred. No. 3.23e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    175 aagcaggggaatttc 189
          ||||||||||||||
Qy     95 AAGCAGGGGAATTTC 109


RESULT   39
ID   Q48592 standard; DNA; 850 BP.
AC   Q48592;
DT   16-FEB-1994  (first entry)
DE   Autonomously replicating sequence.
KW   Expression cassette; expression vector; inducible; methanol;
KW   glycerol; yeast; ss.
OS   Methylotrophic yeast.
PN   EP-558024-A.
PD   01-SEP-1993.
PF   26-FEB-1993; 103040.
PR   28-FEB-1992; JP-043361.
PA   (SUNR ) SUNTORY LTD.
PI   Hatanaka H,  Kondo H,  Sakai Y,  Shibano Y,  Tani Y;
DR   WPI; 93-274483/35.
PT   New expression cassettes inducible by methanol and/or glycerol -
PT   having promoter and terminator derived from alcohol oxidase gene
PS   Claim 6; Page 22; 37pp; English.
CC   An expression cassette which contains the alcohol oxidase promoter
CC   is induced when in the presence of methanol/glycerol.  Under such
```

```
CC   conditions any heterologous gene located downstream of the promoter
CC   will be expressed.  The expression cassette can be used to construct
CC   an expression vector which in turn can be used to transform cells.
CC   When transformed cells are cultured in the presence of methanol/
CC   glycerol expression of the heterologous gene is induced such that
CC   large quantities of the desired product is produced.  The
CC   transformed cells are used particularly for the production of
CC   enzymes such as adenylate kinases, cytochrome C and peroxidases.
CC   The autonomously replicating sequence is used in the construction
CC   of the expression vector, conferring the ability of replication
CC   on the expression vector.
SQ   Sequence   850 BP;    351 A;    83 C;    121 G;    295 T;

  Query Match           5.0%;  Score 15;  DB 8;  Length 850;
  Best Local Similarity 100.0%;  Pred. No. 3.23e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    678 catttatatgtattt 692
          |||||||||||||||
Qy    280 CATTTATATGTATTT 294


RESULT   40
ID   T01542 standard; DNA; 900 BP.
AC   T01542;
DT   23-MAY-1996  (first entry)
DE   Retinoblastoma protein synthesis-inhibiting antisense sequence #2.
KW   Antisense; untranslated region; mouse; retinoblastoma; PCR; primer;
KW   amplification; Balb/c 3T3; cytomegalovirus promoter; expression plasmid;
KW   animal; protein synthesis; cancer; leukaemia; ss.
OS   Synthetic.
PN   J07250684-A.
PD   03-OCT-1995.
PF   15-MAR-1994; 068941.
PR   15-MAR-1994; JP-068941.
PA   (CHIK-) ZH CHIKYU KANKYO SANGYO GIJITSU KENKYU.
DR   WPI; 95-370473/48.
PT   Nucleic acid fragment for inhibiting retinoblastoma protein
PT   synthesis - has sequence complementary to 3'-UTR of mouse Rb gene,
PT   useful in anti:sense gene therapy to inhibit cancer and leukaemia
PT   cell growth
PS   Disclosure; Page 9-10; 12pp; Japanese.
CC   Sequences T01541-3 are nucleic acid fragments contg. antisense sequences
CC   to the 3' untranslated region of the mouse retinoblastoma gene. The
CC   sequences were PCR amplified using the primers T01544-9 on RNA isolated
CC   from mouse Balb/c 3T3 cells as a template.  The sequences were inserted
CC   under control of the cytomegalovirus promoter to produce plasmid pRc/CMV
CC   for expression in animal cells.  Vectors contg. the antisense sequences
CC   can be used to inhibit the synthesis of the retinoblastoma protein in an
CC   animal cell, esp. for inhibiting the growth of cancer or leukaemia cells.
SQ   Sequence   900 BP;    298 A;    164 C;    177 G;    261 T;
```



```
  Query Match           5.0%;  Score 15;  DB 17;  Length 900;
  Best Local Similarity 100.0%;  Pred. No. 3.23e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    533 aaaatacatataaat 547
          |||||||||||||||
Cp    295 AAAATACATATAAAT 281


RESULT   41
ID   V00413 standard; cDNA; 1014 BP.
AC   V00413;
DT   12-MAY-1998  (first entry)
DE   Internal fragment of clone G52_24.
KW   Human; secreted protein; molecular weight marker; genetic fingerprinting;
KW   antibody production; nutritional supplement; therapy; clone G52_24; PBMC;
KW   peripheral blood mononucleocyte; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             1..1014
FT                   /*tag=  a
FT                   /transl_except= (pos:34..36, aa:Xaa)
FT                   /transl_except= (pos:124..126, aa:Xaa)
FT                   /transl_except= (pos:202..204, aa:Xaa)
FT                   /transl_except= (pos:259..261, aa:Xaa)
FT                   /transl_except= (pos:319..321, aa:Xaa)
FT                   /transl_except= (pos:373..375, aa:Xaa)
FT                   /transl_except= (pos:391..393, aa:Xaa)
FT                   /transl_except= (pos:394..396, aa:Xaa)
FT                   /transl_except= (pos:445..447, aa:Xaa)
FT                   /transl_except= (pos:463..465, aa:Xaa)
FT                   /transl_except= (pos:475..477, aa:Xaa)
FT                   /transl_except= (pos:598..600, aa:Xaa)
FT                   /transl_except= (pos:667..669, aa:Xaa)
FT                   /transl_except= (pos:703..705, aa:Xaa)
FT                   /transl_except= (pos:709..711, aa:Xaa)
FT                   /transl_except= (pos:751..753, aa:Xaa)
FT                   /transl_except= (pos:778..780, aa:Xaa)
FT                   /transl_except= (pos:835..837, aa:Xaa)
FT                   /transl_except= (pos:847..849, aa:Xaa)
FT                   /transl_except= (pos:856..858, aa:Xaa)
FT                   /transl_except= (pos:859..861, aa:Xaa)
FT                   /transl_except= (pos:865..867, aa:Xaa)
FT                   /transl_except= (pos:868..870, aa:Xaa)
FT                   /transl_except= (pos:892..894, aa:Xaa)
FT                   /transl_except= (pos:901..903, aa:Xaa)
FT                   /transl_except= (pos:916..918, aa:Xaa)
FT                   /transl_except= (pos:919..921, aa:Xaa)
FT                   /transl_except= (pos:943..945, aa:Xaa)
FT                   /transl_except= (pos:958..960, aa:Xaa)
```



```
FT                   /transl_except= (pos:964..966, aa:Xaa)
FT                   /transl_except= (pos:985..987, aa:Xaa)
FT                   /transl_except= (pos:1012..1014, aa:Xaa)
FT                   /note= "Xaa= unknown; no stop codon given"
PN   WO9740069-A2.
PD   30-OCT-1997.
PF   14-APR-1997; U06134.
PR   19-APR-1996; US-635311.
PA   (GEMY ) GENETICS INST INC.
PI   Jacobs K, Lavallie ER, Mccoy JM, Merberg D, Racie LA,
PI   Spaulding V;
DR   WPI; 97-535776/49.
DR   P-PSDB; W36944.
PT   Isolated nucleic acid clones from ATCC 98028 encode novel secreted
PT   proteins - having many potential uses, e.g. as immunomodulators,
PT   cell proliferation or differentiation inhibitors or haematopoiesis
PT   regulators
PS   Claim 14; Page 63-64; 114pp; English.
CC   This sequence is an internal fragment of clone G52_24, which is a
CC   polynucleotide of the invention. This sequence was isolated from a human
CC   peripheral blood mononucleocyte (PBMC) cDNA library. The polynucleotide,
CC   which encodes a secreted protein, can be used, e.g. as a tissue or
CC   molecular weight marker, in genetic fingerprinting, to raise anti-protein
CC   or anti-DNA antibodies and in interaction trap assays. The protein can be
CC   used to assay biological activity, raise antibodies for use in
CC   immunoassays, as a marker, to identify inhibitors of its interactions and
CC   as a nutritional supplement. It may also have a very wide range of
CC   therapeutic and biological activities (no examples are given to support
CC   this), e.g. cytokine or modulator of cell proliferation and
CC   differentiation, immunostimulant or immunosuppressant, haematopoiesis
CC   regulator, bone, cartilage, tendon, ligament and/or nerve tissue growth
CC   stimulator, follicle inhibitor/stimulator, chemotactic/chemokinetic,
CC   haemostatic, thrombolytic or anti-inflammatory agent, antimicrobial,
CC   biorhythm, metabolism or behaviour modifier, anti-depressant or analgesic
CC   or psoriasis treatative.
SQ   Sequence   1014 BP;    292 A;    192 C;    202 G;    292 T;

  Query Match           5.0%;  Score 15;  DB 38;  Length 1014;
  Best Local Similarity 93.8%;  Pred. No. 3.23e+01;
  Matches   15;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db    344 ctaaagcaggttgtgt 359
          |||||||||||| ||||
Qy      4 CTAAAGCAGGTNGTGT 19


RESULT   42
ID   T91294 standard; cDNA; 1014 BP.
AC   T91294;
DT   19-MAY-1998  (first entry)
DE   Human G52-24 secreted protein cDNA internal fragment.
```



```
KW   Secreted protein; anti-inflammatory; immune stimulant; suppressant;
KW   human; cytokine; autoimmune disease; regulator; activin; inhibin;
KW   G52-24; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             1..1014
FT                   /*tag=  a
FT                   /product= G52-24
FT                   /note= "partial coding sequence of secreted protein"
PN   WO9740151-A2.
PD   30-OCT-1997.
PF   14-APR-1997; U06125.
PR   10-JAN-1997; US-781226.
PR   19-APR-1996; US-635311.
PA   (GEMY ) GENETICS INST INC.
PI   Jacobs K, Lavallie ER, McCoy JM, Merberg D, Racie LA,
PI   Spaulding V;
DR   WPI; 97-535835/49.
DR   P-PSDB; W27287.
PT   Nucleic acids encoding novel secreted proteins - useful e.g. as an
PT   anti-inflammatory, immune stimulant or suppressor, etc.
PS   Claim 1; Page 46; 81pp; English.
CC   This sequence is an internal fragment of a novel secreted protein,
CC   G52-24, which is isolated from a clone, G52-24 (ATCC 98028), derived
CC   from a human PBMC cDNA library. The nucleic acid and encoded G52-24
CC   secreted protein can be used for research purposes (as markers for
CC   tissues, molecular weight markers for gels, primers, probes, etc.), for
CC   nutrition (as C, N or carbohydrate source), as a cytokine for cell
CC   proliferation and differentiation activity, as immune stimulants or
CC   suppressors e.g. for viral, bacterial or fungal infections, for
CC   autoimmune diseases such as multiple sclerosis or systemic lupus
CC   erythematosus, to regulate haematopoiesis, for tissue growth, as an
CC   activin or inhibin, or having chemotactic, chemokinetic, haemostatic and
CC   thrombolytic, receptor/ligand, anti-inflammatory or tumour inhibitory
CC   activities.
SQ   Sequence   1014 BP;    292 A;    192 C;    202 G;    292 T;

  Query Match           5.0%;  Score 15;  DB 38;  Length 1014;
  Best Local Similarity 93.8%;  Pred. No. 3.23e+01;
  Matches   15;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db    344 ctaaagcaggttgtgt 359
          |||||||||||| ||||
Qy      4 CTAAAGCAGGTNGTGT 19


RESULT   43
ID   T37066 standard; DNA; 1478 BP.
AC   T37066;
DT   10-NOV-1996  (first entry)
DE   Tomato green fruit-specific promoter TFM7 3' end.
```

```
KW   Sucrose phosphorylase; gtfA gene; transgenic plant; starch;
KW   tomato; Streptococcus mutans; TFM7; ds.
OS   Lycopersicon esculentum.
PN   WO9624679-A1.
PD   15-AUG-1996.
PF   08-FEB-1996; U01959.
PR   10-FEB-1995; US-386860.
PA   (MONS ) MONSANTO CO.
PI   Barry GF, DeWeerd JW, Kishore GM, Weldon ML;
DR   WPI; 96-384451/38.
PT   Transgenic plants expressing sucrose phosphorylase - have increased
PT   solid, esp. starch, content, more uniform solids distribution and
PT   reduced tendency to bruise
PS   Disclosure; Page 31-32; 45pp; English.
CC   Approx. 1.5 kb of the 3' end of the tomato fruit TFM7 promoter is
CC   given in T37066.  The TFM7 and TFM9 (see also T37067) promoters
CC   function during the green fruit stage of tomatoes.  Such promoters
CC   can be used to express heterologous genes in tomato green fruit.
CC   An example is the sucrose phosphorylase gtfA gene (T37063) of
CC   Streptococcus mutans, expression of which will increase solid,
CC   esp. starch, content in the transgenic plant.
SQ   Sequence   1478 BP;    498 A;    222 C;    208 G;    550 T;

  Query Match            5.0%;  Score 15;  DB 23;  Length 1478;
  Best Local Similarity 100.0%;  Pred. No. 3.23e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1299 gtgtgttaaagagct 1313
          |||||||||||||||
Qy    202 GTGTGTTAAAGAGCT 216


RESULT   44
ID   T91295 standard; cDNA; 1560 BP.
AC   T91295;
DT   19-MAY-1998  (first entry)
DE   Human G52-24 secreted protein cDNA.
KW   Secreted protein; anti-inflammatory; immune stimulant; suppressant;
KW   human; cytokine; autoimmune disease; regulator; activin; inhibin;
KW   G52-24; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             712..1277
FT                   /*tag=  a
FT                   /product= G52-24
FT                   /note= "secreted protein"
PN   WO9740151-A2.
PD   30-OCT-1997.
PF   14-APR-1997; U06125.
PR   10-JAN-1997; US-781226.
PR   19-APR-1996; US-635311.
```



```
PA   (GEMY ) GENETICS INST INC.
PI   Jacobs K, Lavallie ER, McCoy JM, Merberg D, Racie LA,
PI   Spaulding V;
DR   WPI; 97-535035/49.
DR   P-PSDB; W27288.
PT   Nucleic acids encoding novel secreted proteins - useful e.g. as an
PT   anti-inflammatory, immune stimulant or suppressor, etc.
PS   Claim 13; Page 58; 81pp; English.
CC   This sequence encodes a novel secreted protein, G52-24, which is isolated
CC   from a clone, G52-24 (ATCC 98028), derived from a human PBMC cDNA
CC   library. The nucleic acid and encoded G52-24 secreted protein can be used
CC   for research purposes (as markers for tissues, molecular weight markers
CC   for gels, primers, probes, etc.), for nutrition (as C, N or carbohydrate
CC   sources), as a cytokine for cell proliferation and differentiation
CC   activity, as immune stimulants or suppressors e.g. for viral, bacterial
CC   or fungal infections, for autoimmune diseases such as multiple sclerosis
CC   or systemic lupus erythematosus, to regulate haematopoiesis, for tissue
CC   growth, as an activin or inhibin, or having chemotactic, chemokinetic,
CC   haemostatic and thrombolytic, receptor/ligand, anti-inflammatory or
CC   tumour inhibitory activities.
SQ   Sequence   1560 BP;    525 A;    279 C;    299 G;    457 T;

  Query Match            5.0%;  Score 15;  DB 38;  Length 1560;
  Best Local Similarity  93.8%;  Pred. No. 3.23e+01;
  Matches    15;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db    754 ctaaagcaggttgtgt 769
          ||||||||||| ||||
Qy      4 CTAAAGCAGGTNGTGT 19


RESULT   45
ID   T15608 standard; cDNA; 1633 BP.
AC   T15608;
DT   23-APR-1996  (first entry)
DE   FGF receptor ligand FRL-1 gene.
KW   FRL-2; FRL-1; XTl; ALP; EG2; CLP; FGF receptor;
KW   transmembrane receptor-mediated signalling;
KW   fibroblast growth factor receptor; receptor protein tyrosine kinase;
KW   neural disorder; tumour; gene therapy; diagnosis; ds.
OS   Xenopus laevis.
FH   Key             Location/Qualifiers
FT   cds             91..663
FT                   /*tag=  a
PN   WO9603499-A1.
PD   08-FEB-1996.
PF   19-JUL-1995; U09172.
PR   22-JUL-1994; US-279217.
PR   15-MAY-1995; US-441629.
PA   (HARD ) HARVARD COLLEGE.
PI   Kinoshita N,  Kirschner MW;
```



```
DR   WPI; 96-117044/12.
DR   P-PSDB; R90768.
PT   Two new ligands, FRL-2 and FRL-1 for the FGF receptor  - used in
PT   treatment of neural disorders, e.g. tumours, in vertebrate cells
PS   Claim 1; Page 41-43; 83pp; English.
CC   2 Genes (T15607 and T15608) encoding novel fibroblast growth factor
CC   (FGF) receptor ligands, FRL-2 (R90767) and FRL-1 (R90768), were
CC   isolated by screening a Xenopus laevis egg cDNA library with
CC   Saccharomyces cerevisiae PSY316 cells expressing FGF receptor,
CC   using a functional expression technique.  FRL-1 and FRL-2 are
CC   involved in transmembrane receptor-mediated signalling.  The cDNAs
CC   may be labeled and used as probes or primers to detect vertebrate
CC   FRL-2 or FRL-1 cDNAs, used to produce the recombinant ligands, or
CC   used in gene therapy as a means of modulating the function of FGF
CC   receptors in vertebrate cells or tissues.
SQ   Sequence   1633 BP;    478 A;    311 C;    324 G;    520 T;

  Query Match            5.0%;  Score 15;  DB 17;  Length 1633;
  Best Local Similarity 100.0%;  Pred. No. 3.23e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    999 tgtatatatgtaaat 1013
          |||||||||||||||
Qy    186 TGTATATATGTAAAT 200


RESULT   46
ID   T01541 standard; DNA; 1697 BP.
AC   T01541;
DT   23-MAY-1996  (first entry)
DE   Retinoblastoma protein synthesis-inhibiting antisense sequence #1.
KW   Antisense; untranslated region; mouse; retinoblastoma; PCR; primer;
KW   amplification; Balb/c 3T3; cytomegalovirus promoter; expression plasmid;
KW   animal; protein synthesis; cancer; leukaemia; ss.
OS   Synthetic.
PN   J07250684-A.
PD   03-OCT-1995.
PF   15-MAR-1994; 068941.
PR   15-MAR-1994; JP-068941.
PA   (CHIK-) ZH CHIKYU KANKYO SANGYO GIJITSU KENKYU.
DR   WPI; 95-370473/48.
PT   Nucleic acid fragment for inhibiting retinoblastoma protein
PT   synthesis - has sequence complementary to 3'-UTR of mouse Rb gene,
PT   useful in anti:sense gene therapy to inhibit cancer and leukaemia
PT   cell growth
PS   Disclosure; Page 9; 12pp; Japanese.
CC   Sequences T01541-3 are nucleic acid fragments contg. antisense aequences
CC   to the 3' untranslated region of the mouse retinoblastoma gene. The
CC   sequences were PCR amplified using the primers T01544-9 on RNA isolated
CC   from mouse Balb/c 3T3 cells as a template.  The sequences were inserted
CC   under control of the cytomegalovirus promoter to produce plasmid pRc/CMV
```



```
CC   for expression in animal cells.  Vectors contg. the antisense sequences
CC   can be used to inhibit the synthesis of the retinoblastoma protein in an
CC   animal cell, esp. for inhibiting the growth of cancer or leukaemia cells.
SQ   Sequence   1697 BP;    592 A;    295 C;    322 G;    488 T;

  Query Match            5.0%;  Score 15;  DB 17;  Length 1697;
  Best Local Similarity 100.0%;  Pred. No. 3.23e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1330 aaaatacatataaat 1344
          |||||||||||||||
Cp    295 AAAATACATATAAAT 281


RESULT   47
ID   T62390 standard; DNA; 1698 BP.
AC   T62390;
DT   01-JUL-1997  (first entry)
DE   Antisense sequence to bases 2878-4577 of mouse Rb gene 3' UTR.
KW   Antisense; untranslated region; mouse; retinoblastoma; construct; cancer;
KW   expression vector; internal ribosomal entry site; IRES; gene therapy;
KW   leukaemia; proliferation; ss.
OS   Mus musculus.
PN   J09023884-A.
PD   28-JAN-1997.
PF   13-JUL-1995; 177741.
PR   13-JUL-1995; JP-177741.
PA   (CHIK-) ZH CHIKYU KANKYO SANGYO GIJITSU KENKYU.
DR   WPI; 97-148586/14.
PT   Anti-sense RNA expression vector - useful for suppressing expression
PT   of a protein, e.g. retinoblastoma protein, which regulates cellular
PT   proliferation, for cancer therapy
PS   Example 2; Page 20-21; 30pp; Japanese.
CC   This sequence represents the antisense sequence to nucleotides 2878-4577
CC   covering the 3' untranslated region (3' UTR) of the mouse retinoblastoma
CC   (Rb) gene.  The antisense DNA can be used to generate a DNA construct
CC   which has a first DNA sequence coding for an antisense RNA suppressing
CC   the expression of a desired protein and a second DNA sequence acting as a
CC   marker gene, in which the 5'-terminal of the second DNA sequence has a
CC   DNA sequence coding for an internal ribosomal entry site (IRES) and the
CC   3'-terminal of the second DNA sequence has a DNA sequence coding for a
CC   poly(A)-addition signal.  The antisense sequence can also be split into
CC   2 regions covering bases 2878-3778 and 3779-4577 for generation of
CC   constructs containing these regions either singly or as tandem repeats
CC   (T62393-7).  The antisense RNA expression vectors can be applied to e.g.
CC   genetic therapy of cancer, leukaemia, etc., by suppressing the expression
CC   of a protein which regulates the proliferation of cancer cells, leukaemic
CC   cells, etc.
SQ   Sequence   1698 BP;    593 A;    295 C;    322 G;    488 T;

  Query Match            5.0%;  Score 15;  DB 29;  Length 1698;
```

```
  Best Local Similarity 100.0%;  Pred. No. 3.23e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1331 aaaatacatataaat 1345
          ||||||||||||||
Cp    295 AAAATACATATAAAT 281


RESULT   48
ID   T62391 standard; DNA; 1802 BP.
AC   T62391;
DT   01-JUL-1997  (first entry)
DE   Direct tandem repeat of mouse Rb gene 3' UTR bases 2878-3778.
KW   Antisense; untranslated region; mouse; retinoblastoma; construct; cancer;
KW   expression vector; internal ribosomal entry site; IRES; gene therapy;
KW   leukaemia; proliferation; ss.
OS   Mus musculus.
PN   J09023884-A.
PD   28-JAN-1997.
PF   13-JUL-1995; 177741.
PR   13-JUL-1995; JP-177741.
PA   (CHIK-) ZH CHIKYU KANKYO SANGYO GIJITSU KENKYU.
DR   WPI; 97-148586/14.
PT   Anti-sense RNA expression vector - useful for suppressing expression
PT   of a protein, e.g. retinoblastoma protein, which regulates cellular
PT   proliferation, for cancer therapy
PS   Claim 3; Page 22-23; 30pp; Japanese.
CC   This sequence constitutes a direct tandem repeat of the antisense
CC   sequence to nucleotides 2878-3778 of the 3' untranslated region (3'
CC   UTR) of the mouse retinoblastoma (Rb) gene.  The antisense DNA can be
CC   used to generate a DNA construct which has a first DNA sequence coding
CC   for an antisense RNA suppressing the expression of a desired protein and
CC   a second DNA sequence acting as a marker gene, in which the 5'-terminal
CC   of the second DNA sequence has a DNA sequence coding for an internal
CC   ribosomal entry site (IRES) and the 3'-terminal of the second DNA
CC   sequence has a DNA sequence coding for a poly(A)-addition signal.  The
CC   antisense sequence can also be split into 2 regions covering bases
CC   2878-3778 and 3779-4577 for generation of constructs containing these
CC   regions either singly or as tandem repeats (T62393-7).  The antisense RNA
CC   expression vectors can be applied to e.g. genetic therapy of cancer,
CC   leukaemia, etc., by suppressing the expression of a protein which
CC   regulates the proliferation of cancer cells, leukaemic cells, etc.
SQ   Sequence   1802 BP;    598 A;    328 C;    354 G;    522 T;

  Query Match           5.0%;  Score 15;  DB 29;  Length 1802;
  Best Local Similarity 100.0%;  Pred. No. 3.23e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    534 aaaatacatataaat 548
          ||||||||||||||
Cp    295 AAAATACATATAAAT 281
```



```
RESULT   49
ID   Q78149 standard; cDNA; 2037 BP.
AC   Q78149;
DT   06-JUN-1995  (first entry)
DE   Large subunit of ADP-glucose pyrophosphorylase (AGP) from barley
DE   seed endosperm (bepl10).
KW   Barley; endosperm; ADP glucose pyrophosphorylase; AGP; starch;
KW   enzyme; transgenic plant; ss.
OS   Hordeum vulgare.
FH   Key             Location/Qualifiers
FT   cds             189..1772
FT                   /*tag= a
PN   WO9424292-A.
PD   27-OCT-1994.
PF   07-APR-1994; E01082.
PR   08-APR-1993; GB-007408.
PA   (DANI-) DANISCO AS.
PI   Kleczkowski L,  Marcussen J,  Okkels F,  Olsen O,  Poulsen P;
PI   Villand P;
DR   WPI; 94-341871/42.
DR   P-PSDB; R63446.
PT   Transgenic starch producing organism - comprises exogenous ADP
PT   glucose pyrophosphorylase enzyme DNA sequence
PS   Claim 6; Fig 14; 88pp; English.
CC   Q78149 is a 1.9 kb cDNA fragment encoding the large subunit ADP-
CC   glucose pyrophosphorylase from barley endosperm (Villand et al. 1992,
CC   Plant Physiol 100: 1617-1618). For the construction of plasmid pPPL1,
CC   a 1.9 kb EcoRI-HindIII cDNA fragment encoding the barley endosperm
CC   AGP large subunit (bepl) was isolated, the restriction ends were
CC   filled in with klenow DNA polymerase, and the blunt ended DNA
CC   fragment was cloned in the SmaI site of plasmid pPATA1. From the
CC   resulting plasmid the 3.2 kb EcoRI fragment contg. the patatin
CC   promoter, the bepl cDNA, and the 35S terminator was isolated and
CC   inserted in the EcoRI site of the plant transformation vector pBKL4
CC   to yield plasmid pPPL1, which is claimed.
SQ   Sequence   2037 BP;    496 A;    557 C;    552 G;    432 T;

  Query Match           5.0%;  Score 15;  DB 13;  Length 2037;
  Best Local Similarity 100.0%;  Pred. No. 3.23e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     87 cctccctgcatttga 101
          ||||||||||||||
Qy    253 CCTCCCTGCATTTGA 267


RESULT   50
ID   Q91762 standard; DNA; 2125 BP.
AC   Q91762;
```



```
DT   15-FEB-1996  (first entry)
DE   Fission yeast Rad51 gene, encodes protein involved in DNA repair.
KW   DNA mismatch repair; DNA recombination; site-specific; unwinding;
KW   right-handed helical nucleoprotein; RAD51; recA; S.pombe;
KW   topoisomerase; ss.
OS   Schizosaccharomyces pombe.
FH   Key             Location/Qualifiers
FT   cds             478..1575
FT                   /*tag= a
FT                   /product= RAD51
PN   J07143890-A.
PD   06-JUN-1995.
PF   28-MAY-1993; 127594.
PR   28-MAY-1993; JP-127594.
PA   (TOYM ) TOYOBO KK.
DR   WPI; 95-236467/31.
DR   P-PSDB; R78182.
PT   Structural gene encoding RAD51, used in production of RAD51 - used
PT   to study drugs against diseases caused by DNA damage, e.g. by UV or
PT   X radiation, and to improve efficiency of gene therapy targetting
PS   Claim 1; Page 12-13; 19pp; Japanese.
CC   The present sequence, isolated from S.pombe DNA (chromosome 1), is a
CC   specific example of DNA coding for a RAD51 protein having the amino
CC   acid sequence in R78182. The RAD51 protein, which contains an ATP-
CC   binding domain, is involved in DNA repair of mismatched base pairs
CC   and site-specific recombination in antibody gene rearrangements. The
CC   protein binds to double-stranded DNA to form a right-handed helical
CC   nucleoprotein that extends (by 1.5 times) the pitch of B-form DNA. As
CC   a result, the helix is unwound. The Rad51 gene and the protein it
CC   encodes will be useful in designing drugs to treat diseases associated
CC   with environmental DNA damage.
SQ   Sequence   2125 BP;    635 A;    378 C;    402 G;    710 T;

  Query Match           5.0%;  Score 15;  DB 16;  Length 2125;
  Best Local Similarity 93.8%;  Pred. No. 3.23e+01;
  Matches   15;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db   1259 ctgctcgtcaaatgca 1274
          |||||| |||||||||
Cp    274 CTGCTCNTCAAATGCA 259


Search completed: Tue Dec 22 20:31:12 1998
Job time : 172 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Tue Dec 22 20:06:22 1998; MasPar time 468.31 Seconds
1430.571 Million cell updates/sec

Tabular output not generated.

Title: >US-08-530-112A-1017
Description: (1-304) from US08530112A.seq
Perfect Score: 298
N.A. Sequence: 1 GATCTAAAGCAGGTNGTGTT.....ATATGTATTTTNGGGATAAA 304
Comp: CTAGATTTCGTCCANCACAA.....TATACATAAAANCCCTATTT

Scoring table: TABLE jmetric
Gap 60

Kmatch STD : Dbase 0; Query 0

Searched: 567134 seqs, 1101898692 bases x 2

Post-processing: Minimum Match 0%
Listing first 1000 summaries

Database: embl55
1:em_ba 2:em_fun 3:em_htg 4:em_hum1 5:em_hum2 6:em_in
7:em_om 8:em_or 9:em_ov 10:em_pat 11:em_ph 12:em_pl
13:em_ro 14:em_vi

Database: genbank107
15:gb_ba1 16:gb_ba2 17:gb_htg 18:gb_in 19:gb_om 20:gb_ov
21:gb_pat 22:gb_ph 23:gb_pl1 24:gb_pl2 25:gb_pr1
26:gb_pr2 27:gb_pr3 28:gb_ro 29:gb_st 30:gb_sts 31:gb_sy
32:gb_un 33:gb_vi

Statistics: Mean 8.250; Variance 2.992; scale 2.757

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
c	1	25	8.4	508	30	HS024XG1	H. sapiens (D16S400) D	3.55e-05
	2	25	8.4	130491	26	HUAC003977	Homo sapiens Chromosom	3.55e-05
	3	23	7.7	291	30	RLU92274	Rana luteiventris repe	1.49e-03
c	4	23	7.7	1020	28	MML41RIBO	M.musculus mRNA for L4	1.49e-03
c	5	23	7.7	78616	28	AC000398	Genomic sequence from	1.49e-03
	6	23	7.7	136569	17	AC005104	*** SEQUENCING IN PROG	1.49e-03
	7	23	7.7	149752	26	AC004616	Homo sapiens Xp22 BAC	1.49e-03
c	8	23	7.7	173139	17	HS1191B2	Human DNA sequence ***	1.49e-03
	9	23	7.7	199553	17	HS526I14	Human DNA sequence ***	1.49e-03
c	10	22	7.4	103268	26	HS67M12	Human DNA sequence fro	9.15e-03
	11	22	7.4	171952	28	AC002324	Mus musculus chromosom	9.15e-03
c	12	21	7.0	312	30	HSA108WE1	H.sapiens (D12S1585) D	5.37e-02
c	13	21	7.0	421	26	HS22QT017	H.sapiens telomeric DN	5.37e-02
c	14	21	7.0	7354	27	HSU50912	Human XIST gene, poly	5.37e-02
c	15	21	7.0	31520	26	HSXIST1	Human Xq13 5' end of P	5.37e-02
	16	21	7.0	75793	26	HS321I15	Human DNA sequence fro	5.37e-02
c	17	21	7.0	203583	17	AC004940	*** SEQUENCING IN PROG	5.37e-02
	18	20	6.7	182	20	CHKMCW82	Gallus domesticus DNA	3.01e-01
	19	20	6.7	225	30	G37914	c3m52 Plasmodium falci	3.01e-01
c	20	20	6.7	264	21	I31098	Sequence 10 from paten	3.01e-01
c	21	20	6.7	264	25	HSMFD22	Human MFD22 dinucleoti	3.01e-01
c	22	20	6.7	265	30	HS240XD10	H.sapiens DNA segment	3.01e-01
c	23	20	6.7	392	30	HS324XH5	H.sapiens DNA segment	3.01e-01
c	24	20	6.7	546	28	MMU31315	Mus musculus ACAT tetr	3.01e-01
c	25	20	6.7	1516	28	MMC35FLA	M.musculus DNA for C3	3.01e-01
	26	20	6.7	1799	23	ATHG3P	Arabidopsis thaliana c	3.01e-01
c	27	20	6.7	2227	28	RATSCG10A	Rattus norvegicus neur	3.01e-01
c	28	20	6.7	5430	18	PFU27339	Plasmodium falciparum	3.01e-01
c	29	20	6.7	6608	26	HSXLRPGN2	H.sapiens RP3 gene (XL	3.01e-01
c	30	20	6.7	7749	23	PNPHTT	P.nudum gene for phyto	3.01e-01
	31	20	6.7	9012	23	FTPPCA	F.trinervia ppcA1 gene	3.01e-01
c	32	20	6.7	10213	18	PFU27338	Plasmodium falciparum	3.01e-01
	33	20	6.7	10824	28	MMU62765	Mus musculus meprin A	3.01e-01
	34	20	6.7	28865	27	HSN104C7	Human DNA sequence fro	3.01e-01
	35	20	6.7	30101	18	PFSC03015	Plasmodium falciparum	3.01e-01
	36	20	6.7	31162	18	PFSC03013	Plasmodium falciparum	3.01e-01
	37	20	6.7	37056	26	AC004700	Homo sapiens chromosom	3.01e-01
c	38	20	6.7	39132	26	HSU75A4	Human DNA sequence fro	3.01e-01
	39	20	6.7	53419	28	MMU35323	Mus musculus H2-M alph	3.01e-01
	40	20	6.7	58663	18	AC004300	Drosophila melanogaste	3.01e-01
	41	20	6.7	62031	28	AF021335	Mus musculus T cell re	3.01e-01
	42	20	6.7	77612	17	AC004726	*** SEQUENCING IN PROG	3.01e-01

c	43	20	6.7	85832	26	HS344F17	Human DNA sequence fro	3.01e-01
c	44	20	6.7	94336	26	AC005250	Homo sapiens BAC clone	3.01e-01
	45	20	6.7	99411	26	AC005207	Homo sapiens chromosom	3.01e-01
c	46	20	6.7	100000	26	AP000030	Homo sapiens genomic D	3.01e-01
c	47	20	6.7	105726	28	AF037352	Mus musculus T cell re	3.01e-01
	48	20	6.7	105754	17	AC005000	*** SEQUENCING IN PROG	3.01e-01
c	49	20	6.7	112098	17	AC004927	*** SEQUENCING IN PROG	3.01e-01
	50	20	6.7	112220	26	AC005373	Homo sapiens chromosom	3.01e-01
	51	20	6.7	112309	26	AC003025	Human Chromosome 11p12	3.01e-01
	52	20	6.7	113916	27	HUAC003007	Human Chromosome 16 BA	3.01e-01
	53	20	6.7	114736	18	PFMAL3P3	Plasmodium falciparum	3.01e-01
	54	20	6.7	122638	26	AB001523	Homo sapiens gene for	3.01e-01
c	55	20	6.7	124181	17	HS285F3	Human DNA sequence ***	3.01e-01
	56	20	6.7	138706	17	AC004946	*** SEQUENCING IN PROG	3.01e-01
c	57	20	6.7	143577	17	AC004846	*** SEQUENCING IN PROG	3.01e-01
	58	20	6.7	144820	17	AC005260	*** SEQUENCING IN PROG	3.01e-01
	59	20	6.7	147394	27	HS353H6	Human DNA sequence fro	3.01e-01
c	60	20	6.7	148204	17	AC004881	*** SEQUENCING IN PROG	3.01e-01
	61	20	6.7	152135	17	HS753D4	Human DNA sequence ***	3.01e-01
	62	20	6.7	154028	26	HSAC002087	Human BAC clone GS113D	3.01e-01
	63	20	6.7	155398	17	AC003020	*** SEQUENCING IN PROG	3.01e-01
c	64	20	6.7	156213	17	HS126B4	Human DNA sequence ***	3.01e-01
c	65	20	6.7	156854	26	HSU82672	Human chromosome X clo	3.01e-01
	66	20	6.7	160048	28	AC002109	Genomic sequence from	3.01e-01
	67	20	6.7	160986	17	AC005505	*** SEQUENCING IN PROG	3.01e-01
c	68	20	6.7	161652	26	AC004774	Homo sapiens BAC clone	3.01e-01
	69	20	6.7	163712	27	AC004065	Homo sapiens chromosom	3.01e-01
c	70	20	6.7	171258	17	HS873P14	Human DNA sequence ***	3.01e-01
c	71	20	6.7	175737	27	HS1033B10	Human DNA sequence ***	3.01e-01
	72	20	6.7	185402	17	AC003066	*** SEQUENCING IN PROG	3.01e-01
	73	20	6.7	186314	28	AC003694	Mus musculus chromosom	3.01e-01
c	74	20	6.7	199082	17	AC005632	*** SEQUENCING IN PROG	3.01e-01
	75	20	6.7	238670	17	AC003059	*** SEQUENCING IN PROG	3.01e-01
c	76	20	6.7	240114	17	HS90L6	Human DNA sequence ***	3.01e-01
	77	19	6.4	217	30	G18430	BM1602 cow Bos taurus	1.61e+00
	78	19	6.4	250	30	G15557	human STS SHGC-17078.	1.61e+00
c	79	19	6.4	258	25	HUM275MFD	Homo sapiens DNA seque	1.61e+00
c	80	19	6.4	322	30	G10664	human STS CHLC.UTR_039	1.61e+00
c	81	19	6.4	330	30	HS294WG1	H. sapiens (D18S481) DN	1.61e+00
c	82	19	6.4	330	30	G34832	human STS SHGC-52302,	1.61e+00
c	83	19	6.4	333	25	HSU60166	Human microsatellite f	1.61e+00
c	84	19	6.4	338	30	HUMUT5627	Human STS UT5627, sequ	1.61e+00
c	85	19	6.4	389	30	HS269XG9	H. sapiens (D10S571) D	1.61e+00
	86	19	6.4	417	30	HUMUT6082B	Human STS UT6082, 3' p	1.61e+00
	87	19	6.4	474	28	MMU31665	Mus musculus chromosom	1.61e+00
c	88	19	6.4	504	30	G32112	STS fD17CA, sequence	1.61e+00
c	89	19	6.4	645	30	HSU27132	Human STS 12-RM76 from	1.61e+00
	90	19	6.4	2040	26	HSDSC2X1	H.sapiens DSC2 gene, p	1.61e+00
c	91	19	6.4	2341	23	SCPMT3	S.cerevisiae PMT3 gene	1.61e+00
c	92	19	6.4	2398	25	HSRTKEPH	H.sapiens mRNA for rec	1.61e+00
c	93	19	6.4	3370	25	HUMTKR	Human tyrosine kinase	1.61e+00

c	94	19	6.4	3502	23	YSCTPD3X	Yeast protein phosphat	1.61e+00
c	95	19	6.4	3511	23	SCYOR321W	S.cerevisiae chromosom	1.61e+00
	96	19	6.4	3706	18	PFAATUBII	P.falciparum alpha-tub	1.61e+00
c	97	19	6.4	3723	24	DDIMYOC	Dictyostelium discoide	1.61e+00
	98	19	6.4	3749	18	TRBG572	Trypanosoma cruzi inse	1.61e+00
c	99	19	6.4	4472	18	PFSC04018	Plasmodium falciparum	1.61e+00
c	100	19	6.4	4519	18	PFSC04089	Plasmodium falciparum	1.61e+00
	101	19	6.4	4791	26	AB010894	Homo sapiens mRNA for	1.61e+00
c	102	19	6.4	8476	28	MMU48804	Mus musculus Zn-finger	1.61e+00
c	103	19	6.4	18280	18	PFSC03029	Plasmodium falciparum	1.61e+00
c	104	19	6.4	22243	18	PFAVAR23A	Plasmodium falciparum	1.61e+00
c	105	19	6.4	29410	23	YSCLTESPO	Saccharomyces cerevisi	1.61e+00
	106	19	6.4	30848	18	AC004307	Drosophila melanogaste	1.61e+00
c	107	19	6.4	34127	18	DMC56A1	Drosophila melanogaste	1.61e+00
c	108	19	6.4	35259	27	AC005603	Homo sapiens subtelome	1.61e+00
c	109	19	6.4	35710	23	SCCHXVCOS	S.cerevisiae DNA from	1.61e+00
	110	19	6.4	36273	25	HUM7501	Homo sapiens chromosom	1.61e+00
c	111	19	6.4	37640	27	AC005605	Homo sapiens subtelome	1.61e+00
	112	19	6.4	38167	26	AC004751	Homo sapiens chromosom	1.61e+00
	113	19	6.4	40881	26	AC004679	Homo sapiens chromosom	1.61e+00
c	114	19	6.4	41804	27	AC005604	Homo sapiens subtelome	1.61e+00
	115	19	6.4	56737	25	HUMHPRTB	Human hypoxanthine pho	1.61e+00
c	116	19	6.4	73371	27	AC004872	Homo sapiens PAC clone	1.61e+00
c	117	19	6.4	80611	26	HS323P24	Human DNA sequence fro	1.61e+00
c	118	19	6.4	85661	24	AB013594	Arabidopsis thaliana g	1.61e+00
	119	19	6.4	102703	26	AC004910	Homo sapiens PAC clone	1.61e+00
	120	19	6.4	106148	26	HS73A14	Human DNA sequence fro	1.61e+00
	121	19	6.4	114918	26	AC005188	Homo sapiens PAC clone	1.61e+00
c	122	19	6.4	122256	17	AC005443	*** SEQUENCING IN PROG	1.61e+00
	123	19	6.4	123211	26	AC005208	Homo sapiens chromosom	1.61e+00
	124	19	6.4	123452	17	AC005079	*** SEQUENCING IN PROG	1.61e+00
	125	19	6.4	124230	28	AC005259	Mouse BAC CitbCJ7 219m	1.61e+00
c	126	19	6.4	124519	28	AC003019	Mus musculus Bac B6, c	1.61e+00
	127	19	6.4	132229	26	HS447N6	Human DNA sequence fro	1.61e+00
c	128	19	6.4	135648	26	AC004069	Homo sapiens chromosom	1.61e+00
	129	19	6.4	136222	27	AC004099	Homo sapiens chromosom	1.61e+00
c	130	19	6.4	137260	17	PFMAL1P4	Plasmodium falciparum	1.61e+00
c	131	19	6.4	138155	17	AC004843	*** SEQUENCING IN PROG	1.61e+00
c	132	19	6.4	138839	26	HUAC002400	Human Chromosome 16 BA	1.61e+00
c	133	19	6.4	150000	17	AC004701	*** SEQUENCING IN PROG	1.61e+00
	134	19	6.4	152383	26	HSF18108	Homo sapiens chromosom	1.61e+00
c	135	19	6.4	154821	17	AC005341	*** SEQUENCING IN PROG	1.61e+00
	136	19	6.4	156461	26	AC004383	Human Chromosome X clo	1.61e+00
c	137	19	6.4	160986	17	AC005505	*** SEQUENCING IN PROG	1.61e+00
c	138	19	6.4	168238	17	AC005014	*** SEQUENCING IN PROG	1.61e+00
c	139	19	6.4	177426	17	HS436K10	Human DNA sequence ***	1.61e+00
	140	19	6.4	185893	26	AC005191	Homo sapiens PAC clone	1.61e+00
c	141	19	6.4	191565	26	HUAC004531	Homo sapiens Chromosom	1.61e+00
	142	19	6.4	197391	17	AC005140	*** SEQUENCING IN PROG	1.61e+00
	143	19	6.4	206056	17	AC004387	*** SEQUENCING IN PROG	1.61e+00
	144	19	6.4	224064	17	AC005055	*** SEQUENCING IN PROG	1.61e+00

	145	19	6.4	230023	17	HS522P13	Human DNA sequence ***	1.61e+00
	146	19	6.4	323932	17	AC005682	*** SEQUENCING IN PROG	1.61e+00
c	147	19	6.4	326569	17	AC004912	*** SEQUENCING IN PROG	1.61e+00
c	148	18	6.0	311	30	HUMUT7098	Human STS UT7098.	8.10e+00
c	149	18	6.0	316	30	HSB304YB9	H.sapiens (D18S1125) D	8.10e+00
	150	18	6.0	400	30	G17196	human STS SHGC-14746 c	8.10e+00
c	151	18	6.0	401	30	HS423N7T	H.sapiens STS from gen	8.10e+00
c	152	18	6.0	411	30	G31858	209m Human chromosome	8.10e+00
	153	18	6.0	464	28	RATMSA	Rattus norvegicus poly	8.10e+00
c	154	18	6.0	467	28	MMROMY1P	Mouse myl pseudogene f	8.10e+00
	155	18	6.0	500	30	G27287	human STS CHLC.GATA181	8.10e+00
c	156	18	6.0	546	15	STPUTPA	S. typhimurium put reg	8.10e+00
	157	18	6.0	723	27	HUMZB82D09	Homo sapiens full leng	8.10e+00
c	158	18	6.0	730	15	ECOPUTC	E.coli putC region enc	8.10e+00
c	159	18	6.0	841	24	PCAF004072	Putoria calabrica rps1	8.10e+00
	160	18	6.0	1394	28	MUSAEG2A	Mouse acidic epididyma	8.10e+00
c	161	18	6.0	1400	18	DMCUBINTS1	Drosophila melanogaste	8.10e+00
	162	18	6.0	1406	28	MUSCRISPB	Mouse cysteine-rich se	8.10e+00
	163	18	6.0	1461	23	YSCPTC1A	Saccharomyces cerevisi	8.10e+00
c	164	18	6.0	1559	28	AF068749	Mus musculus sphingosi	8.10e+00
	165	18	6.0	1653	23	SCYDL006W	S.cerevisiae chromosom	8.10e+00
c	166	18	6.0	1786	25	LMGL01	Lemur albifrons pseudo	8.10e+00
c	167	18	6.0	1815	28	AF068748	Mus musculus sphingosi	8.10e+00
	168	18	6.0	1884	15	STYPUTPC	Salmonella sp. (S2985)	8.10e+00
	169	18	6.0	1884	15	STYPUTPD	Salmonella (POTTEMP) p	8.10e+00
	170	18	6.0	1887	15	STYPUTPF	Salmonella (S2983) pro	8.10e+00
	171	18	6.0	1887	15	STYPUTPE	Salmonella (S2980) pro	8.10e+00
	172	18	6.0	1888	15	STYPUTPP	Salmonella (S3044) pro	8.10e+00
	173	18	6.0	1888	15	ECOPUTPL	Escherichia coli (E851	8.10e+00
	174	18	6.0	1888	15	STYPUTPO	Salmonella (S3041) pro	8.10e+00
	175	18	6.0	1889	15	STYPUTPM	Salmonella (S2978) pro	8.10e+00
	176	18	6.0	1889	15	STYPUTPI	Salmonella (S3015) pro	8.10e+00
	177	18	6.0	1889	15	STYPUTPJ	Salmonella (S3027) pro	8.10e+00
	178	18	6.0	1889	15	STYPUTPN	Salmonella (S2979) pro	8.10e+00
	179	18	6.0	1889	15	STYPUTPK	Salmonella (S3013) pro	8.10e+00
	180	18	6.0	1889	15	RIRPUTPA	Rickettsia typhi (S333	8.10e+00
	181	18	6.0	1889	15	ECOPUTPK	Escherichia coli (EC70	8.10e+00
	182	18	6.0	1889	15	STYPUTPL	Salmonella (S3014) pro	8.10e+00
	183	18	6.0	1889	15	STYPUTPB	Salmonella typhimurium	8.10e+00
	184	18	6.0	1890	15	ECOPUTPH	Escherichia coli (EC58	8.10e+00
	185	18	6.0	1890	15	ECOPUTPI	Escherichia coli (EC64	8.10e+00
	186	18	6.0	1890	15	ECOPUTPB	Escherichia coli (E340	8.10e+00
	187	18	6.0	1890	15	ECOPUTPG	Escherichia coli (EC52	8.10e+00
	188	18	6.0	1890	15	ECOPUTPJ	Escherichia coli (E830	8.10e+00
	189	18	6.0	1890	15	ECOPUTPA	Escherichia coli (E266	8.10e+00
	190	18	6.0	1890	15	STYPUTPG	Salmonella (S2995) pro	8.10e+00
	191	18	6.0	1890	15	ECOPUTPE	Escherichia coli (EC32	8.10e+00
	192	18	6.0	1890	15	ECOPUTPD	Escherichia coli (EC14	8.10e+00
	193	18	6.0	1890	15	STYPUTPH	Salmonella (S3057) pro	8.10e+00
	194	18	6.0	1890	15	ECOPUTPC	Escherichia coli (EC10	8.10e+00
	195	18	6.0	1890	15	ECOPUTPF	Escherichia coli (EC40	8.10e+00

c	196	18	6.0	1910	23	SCYBR211C	S.cerevisiae chromosom	8.10e+00
	197	18	6.0	1944	23	PVRBCS4SS	P.vittata mRNA for rub	8.10e+00
	198	18	6.0	2212	15	ECPUTP	E.coli putP gene for p	8.10e+00
	199	18	6.0	2274	24	SCYDL005C	S.cerevisiae chromosom	8.10e+00
	200	18	6.0	2279	21	I24442	Sequence 8 from patent	8.10e+00
c	201	18	6.0	2611	18	DMU1951	D. melanogaster U1-95.	8.10e+00
c	202	18	6.0	2612	23	MIPPLOOPA	P.pijperi mitochondrio	8.10e+00
	203	18	6.0	2731	23	MIPPLOOPB	P.pijperi mitochondrio	8.10e+00
	204	18	6.0	2891	21	A37237	Sequence 6 from Patent	8.10e+00
	205	18	6.0	2909	21	E02147	Rat protein kinase C(d	8.10e+00
	206	18	6.0	2909	28	RATPKCDA	Rat protein kinase C d	8.10e+00
	207	18	6.0	3132	25	HUMCENPRO	H.sapiens centromere a	8.10e+00
c	208	18	6.0	3263	23	SCYBR210W	S.cerevisiae chromosom	8.10e+00
c	209	18	6.0	3462	23	SCYJR078W	S.cerevisiae chromosom	8.10e+00
c	210	18	6.0	3551	28	RNAALB	R.norvegicus (Sprague	8.10e+00
c	211	18	6.0	3821	28	MUS129SV1	Mouse DNA for tRNA[Ser	8.10e+00
	212	18	6.0	3854	25	HUMMYCL2A	Human MYCL2 gene, comp	8.10e+00
c	213	18	6.0	3960	20	CHKTCADH	G.gallus T-cadherin mR	8.10e+00
c	214	18	6.0	4957	28	AB007136	Mus musculus Psme1 gen	8.10e+00
	215	18	6.0	6350	23	SCSCC1	Saccharomyces cerevisi	8.10e+00
c	216	18	6.0	6883	18	PFSC04010	Plasmodium falciparum	8.10e+00
c	217	18	6.0	7576	18	PFSC04062	Plasmodium falciparum	8.10e+00
c	218	18	6.0	7969	23	ALFASPSYN	Medicago sativa aspara	8.10e+00
c	219	18	6.0	7988	28	MMU05265	Mus musculus BALB/c gp	8.10e+00
c	220	18	6.0	8633	18	PFPFMDR1	P.falciparum pfmdr1 ge	8.10e+00
c	221	18	6.0	10003	18	PFSC04009	Plasmodium falciparum	8.10e+00
	222	18	6.0	10188	18	PFSC04050	Plasmodium falciparum	8.10e+00
	223	18	6.0	10359	15	U67508	Methanococcus jannasch	8.10e+00
c	224	18	6.0	10509	18	AF033037	Plasmodium falciparum	8.10e+00
	225	18	6.0	10718	18	PFSC03079	Plasmodium falciparum	8.10e+00
	226	18	6.0	10751	15	ECAE000203	Escherichia coli K-12	8.10e+00
c	227	18	6.0	13132	18	PFSC03021	Plasmodium falciparum	8.10e+00
c	228	18	6.0	14293	18	PFSC03052	Plasmodium falciparum	8.10e+00
	229	18	6.0	17528	15	D90738	Escherichia coli genom	8.10e+00
c	230	18	6.0	19116	18	DMHDGHOGD	D.melanogaster hedgeho	8.10e+00
	231	18	6.0	21126	21	I04770	Sequence 94 from Paten	8.10e+00
	232	18	6.0	21126	15	RIATL	Integrated Ri plasmid	8.10e+00
	233	18	6.0	21126	21	I24453	Sequence 19 from paten	8.10e+00
c	234	18	6.0	24102	26	AC004649	Homo sapiens chromosom	8.10e+00
c	235	18	6.0	26497	18	CEF40FB	Caenorhabditis elegans	8.10e+00
c	236	18	6.0	30147	18	DMU66884	Drosophila melanogaste	8.10e+00
	237	18	6.0	34127	18	DMC66A1	Drosophila melanogaste	8.10e+00
c	238	18	6.0	34731	25	HSU75931	Human Xp22 cosmid U250	8.10e+00
c	239	18	6.0	34921	18	CELP10D7	Caenorhabditis elegans	8.10e+00
c	240	18	6.0	35882	25	HSU41384	Human small nuclear ri	8.10e+00
c	241	18	6.0	36587	18	AC003700	Drosophila melanogaste	8.10e+00
	242	18	6.0	37404	23	SC8119	S.cerevisiae chromosom	8.10e+00
c	243	18	6.0	37731	27	HSB1D7	Human genomic DNA sequ	8.10e+00
	244	18	6.0	38858	17	PFMAL1P3	Plasmodium falciparum	8.10e+00
c	245	18	6.0	39491	28	AC004101	Mouse Cosmid ma53a016	8.10e+00
c	246	18	6.0	39769	26	AC005174	Homo sapiens clone UWG	8.10e+00

	247	18	6.0	41372	23	SCCHRIV42	S. cerevisiae chromoso	8.10e+00
c	248	18	6.0	41830	28	AC005240	Mus musculus clone UWG	8.10e+00
c	249	18	6.0	43299	27	AC005756	Homo sapiens chromosom	8.10e+00
c	250	18	6.0	44614	18	CEC29E6	Caenorhabditis elegans	8.10e+00
	251	18	6.0	46717	17	HSN102C10	Human DNA sequence ***	8.10e+00
c	252	18	6.0	53942	25	HS506G2A	Human DNA sequence fro	8.10e+00
c	253	18	6.0	57795	26	AC004219	Homo sapiens chromosom	8.10e+00
	254	18	6.0	59383	18	PFSC03066	Plasmodium falciparum	8.10e+00
	255	18	6.0	59782	18	PFSC03093	Plasmodium falciparum	8.10e+00
c	256	18	6.0	61989	23	YSCORFSAA	Saccharomyces cerevisi	8.10e+00
	257	18	6.0	69062	17	AC004133	*** SEQUENCING IN PROG	8.10e+00
c	258	18	6.0	70315	26	AP035397	Homo sapiens cosmids Q	8.10e+00
	259	18	6.0	72050	17	AC002506	*** SEQUENCING IN PROG	8.10e+00
c	260	18	6.0	73470	18	AC005434	Drosophila melanogaste	8.10e+00
	261	18	6.0	76169	26	AC003014	Human PAC clone DJ290B	8.10e+00
c	262	18	6.0	76169	26	AC003014	Human PAC clone DJ290B	8.10e+00
	263	18	6.0	78054	18	AC004296	Drosophila melanogaste	8.10e+00
c	264	18	6.0	78181	23	AB011477	Arabidopsis thaliana g	8.10e+00
c	265	18	6.0	79076	18	AC004423	Drosophila melanogaste	8.10e+00
c	266	18	6.0	79837	23	AB005233	Arabidopsis thaliana g	8.10e+00
	267	18	6.0	81984	25	HSJ30E17	Human DNA sequence fro	8.10e+00
	268	18	6.0	82692	18	AC003924	Drosophila melanogaste	8.10e+00
c	269	18	6.0	82766	23	AB005231	Arabidopsis thaliana g	8.10e+00
	270	18	6.0	83608	17	AC005447	*** SEQUENCING IN PROG	8.10e+00
c	271	18	6.0	83865	23	AB006700	Arabidopsis thaliana g	8.10e+00
	272	18	6.0	85880	17	AC005437	*** SEQUENCING IN PROG	8.10e+00
	273	18	6.0	86139	3	AC004714	*** SEQUENCING IN PROG	8.10e+00
	274	18	6.0	86884	18	AC004366	Drosophila melanogaste	8.10e+00
c	275	18	6.0	91017	27	HS35C21	Human DNA sequence fro	8.10e+00
c	276	18	6.0	92624	23	ATAC003000	Arabidopsis thaliana c	8.10e+00
c	277	18	6.0	95425	17	HSJ506G21	*** SEQUENCING IN PROG	8.10e+00
c	278	18	6.0	98125	24	ATT16L1	Arabidopsis thaliana D	8.10e+00
c	279	18	6.0	98942	18	AC005149	Drosophila melanogaste	8.10e+00
c	280	18	6.0	102798	17	AC000107	*** SEQUENCING IN PROG	8.10e+00
	281	18	6.0	105119	17	AC004856	*** SEQUENCING IN PROG	8.10e+00
	282	18	6.0	106355	17	AC005309	*** SEQUENCING IN PROG	8.10e+00
c	283	18	6.0	106420	26	AC002546	Homo sapiens chromosom	8.10e+00
c	284	18	6.0	108878	26	AC003092	Human BAC clone RG335O	8.10e+00
	285	18	6.0	112621	27	AC004891	Homo sapiens PAC clone	8.10e+00
	286	18	6.0	115602	17	HS1118D24	Human DNA sequence ***	8.10e+00
c	287	18	6.0	119420	26	HS997K18	Human DNA sequence fro	8.10e+00
	288	18	6.0	120156	17	AC005002	*** SEQUENCING IN PROG	8.10e+00
c	289	18	6.0	120156	17	AC005002	*** SEQUENCING IN PROG	8.10e+00
	290	18	6.0	120193	17	AC004966	*** SEQUENCING IN PROG	8.10e+00
	291	18	6.0	121131	26	HS345B16	Human DNA sequence fro	8.10e+00
c	292	18	6.0	122940	17	AC004709	*** SEQUENCING IN PROG	8.10e+00
c	293	18	6.0	123331	26	AC005164	Homo sapiens BAC clone	8.10e+00
c	294	18	6.0	128559	26	AC004677	Homo Sapiens Chromosom	8.10e+00
	295	18	6.0	129281	17	AC004999	*** SEQUENCING IN PROG	8.10e+00
c	296	18	6.0	129354	26	AC004080	Homo sapiens PAC clone	8.10e+00
	297	18	6.0	131929	17	HS239K6	Human DNA sequence ***	8.10e+00

c	298	18	6.0	132470	26	HS203P18	Human DNA sequence fro	8.10e+00
c	299	18	6.0	133925	26	AC004562	Homo sapiens chromosom	8.10e+00
c	300	18	6.0	135007	26	HS423N22	Homo sapiens DNA seque	8.10e+00
c	301	18	6.0	135599	23	CPU30821	Cyanophora paradoxa cy	8.10e+00
	302	18	6.0	140230	17	AC004845	*** SEQUENCING IN PROG	8.10e+00
	303	18	6.0	140679	17	HS246H3	Human DNA sequence ***	8.10e+00
c	304	18	6.0	142807	17	HSAJ02553	*** SEQUENCING IN PROG	8.10e+00
	305	18	6.0	147476	17	AC004157	*** SEQUENCING IN PROG	8.10e+00
c	306	18	6.0	147898	17	HS1111N9	Human DNA sequence ***	8.10e+00
	307	18	6.0	148204	17	AC004881	*** SEQUENCING IN PROG	8.10e+00
	308	18	6.0	149409	26	AC004081	Homo sapiens PAC clone	8.10e+00
c	309	18	6.0	150147	27	AC002543	Homo sapiens BAC clone	8.10e+00
c	310	18	6.0	150158	17	HS1147O16	Human DNA sequence ***	8.10e+00
	311	18	6.0	151018	17	PFMAL1P6	Plasmodium falciparum	8.10e+00
c	312	18	6.0	152383	26	HSF18108	Homo sapiens chromosom	8.10e+00
c	313	18	6.0	153735	26	AC004029	Human BAC clone RG035E	8.10e+00
c	314	18	6.0	155536	26	AC004112	Homo sapiens BAC clone	8.10e+00
	315	18	6.0	156214	26	HSAC002064	Human BAC clone RG016J	8.10e+00
c	316	18	6.0	157653	27	AC005146	Homo sapiens 12q24.2 P	8.10e+00
c	317	18	6.0	160378	26	HS93H18	Human DNA sequence fro	8.10e+00
c	318	18	6.0	160850	17	AC004957	*** SEQUENCING IN PROG	8.10e+00
	319	18	6.0	163889	17	AC004958	*** SEQUENCING IN PROG	8.10e+00
	320	18	6.0	165491	26	AC000385	Human Chromosome 11 pa	8.10e+00
c	321	18	6.0	166219	26	AC003084	Human BAC clone RG084D	8.10e+00
c	322	18	6.0	172008	17	AC005011	*** SEQUENCING IN PROG	8.10e+00
c	323	18	6.0	172151	25	HSAC000114	Human PAC clone DJ527C	8.10e+00
	324	18	6.0	173513	26	HS125H2	Human DNA sequence fro	8.10e+00
c	325	18	6.0	174826	26	HSU91324	Homo sapiens Chromosom	8.10e+00
	326	18	6.0	175839	26	AC004491	Homo sapiens BAC clone	8.10e+00
	327	18	6.0	176781	17	AC002989	*** SEQUENCING IN PROG	8.10e+00
	328	18	6.0	177328	17	HS462D8	Human DNA sequence ***	8.10e+00
c	329	18	6.0	177748	17	CEY64F11	Caenorhabditis elegans	8.10e+00
c	330	18	6.0	177901	27	AC005084	Homo sapiens BAC clone	8.10e+00
c	331	18	6.0	185095	17	AC005495	*** SEQUENCING IN PROG	8.10e+00
	332	18	6.0	187051	17	AC004688	*** SEQUENCING IN PROG	8.10e+00
c	333	18	6.0	188377	17	AC004938	*** SEQUENCING IN PROG	8.10e+00
	334	18	6.0	188730	26	AC004025	Homo sapiens Xp22 BAC	8.10e+00
	335	18	6.0	190000	17	AC005150	*** SEQUENCING IN PROG	8.10e+00
c	336	18	6.0	190000	17	AC004168	*** SEQUENCING IN PROG	8.10e+00
	337	18	6.0	190248	17	AC004988	*** SEQUENCING IN PROG	8.10e+00
c	338	18	6.0	192335	17	AC005042	*** SEQUENCING IN PROG	8.10e+00
c	339	18	6.0	196473	17	HS473B4	Human DNA sequence ***	8.10e+00
c	340	18	6.0	199016	27	HS40G12	Human DNA sequence fro	8.10e+00
	341	18	6.0	199490	17	AC005081	*** SEQUENCING IN PROG	8.10e+00
	342	18	6.0	199927	17	HS101A4	Human DNA sequence ***	8.10e+00
c	343	18	6.0	208609	26	HS389A20	Human DNA sequence fro	8.10e+00
	344	18	6.0	218074	27	HUAC002044	Human Chromosome 16 BA	8.10e+00
	345	18	6.0	236281	26	AC004673	Homo sapiens Xp22 BAC	8.10e+00
c	346	18	6.0	245439	28	MMHC135G15	Mus musculus major his	8.10e+00
	347	18	6.0	261897	17	CEY64G10	Caenorhabditis elegans	8.10e+00
	348	18	6.0	326569	17	AC004912	*** SEQUENCING IN PROG	8.10e+00

	349	17	5.7	151	33	PPHRRA	Human papillomavirus r 3.84e+01
c	350	17	5.7	174	28	RNMINIR54	Rattus norvegicus mini 3.84e+01
	351	17	5.7	200	30	HUMRPT22A	Homo sapiens PCR prime 3.84e+01
c	352	17	5.7	229	30	G01983	human STS SHGC4-1687 c 3.84e+01
c	353	17	5.7	230	25	HS124B8F	H.sapiens CpG island D 3.84e+01
	354	17	5.7	243	25	HS121B8R	H.sapiens CpG island D 3.84e+01
c	355	17	5.7	244	25	HS121B8F	H.sapiens CpG island D 3.84e+01
	356	17	5.7	255	28	MUSCYCP31	Mouse cytochrome c pse 3.84e+01
c	357	17	5.7	267	30	G08427	human STS CHLC.GATA11A 3.84e+01
	358	17	5.7	302	30	G07872	human STS CHLC.GTAT1A7 3.84e+01
c	359	17	5.7	319	30	G41175	Z9192 Zebrafish AB Dan 3.84e+01
	360	17	5.7	330	30	G01306	fruit fly STS Dm2126 c 3.84e+01
c	361	17	5.7	337	30	HSA059ZE1	H.sapiens (D6S1700) DN 3.84e+01
c	362	17	5.7	339	28	USADHA2S02	Mouse alcohol dehydrog 3.84e+01
c	363	17	5.7	343	30	HSB299WB5	H.sapiens DNA segment 3.84e+01
	364	17	5.7	346	18	DM1731L3	Drosophila retrotransp 3.84e+01
	365	17	5.7	351	18	DM1731L5	Drosophila retrotransp 3.84e+01
	366	17	5.7	353	30	HSA301WF1	H.sapiens DNA segment 3.84e+01
	367	17	5.7	356	30	HSA106XG1	H.sapiens (D15S962) DN 3.84e+01
	368	17	5.7	376	30	HSPP14B3	H.sapiens flow-sorted 3.84e+01
	369	17	5.7	397	30	HS289VB5	H.sapiens DNA segment 3.84e+01
c	370	17	5.7	405	28	MMMSAT1	M.musculus (clone D19P 3.84e+01
	371	17	5.7	467	28	AF054043	Rattus norvegicus poly 3.84e+01
c	372	17	5.7	492	18	DDPRODHC	D.discoideum promoter 3.84e+01
c	373	17	5.7	595	30	G39883	Z11250 Zebrafish AB Da 3.84e+01
	374	17	5.7	621	20	XLGREP	X.laevis gastrula repe 3.84e+01
c	375	17	5.7	625	30	HUMUT2341	Human chromosome 1 STS 3.84e+01
	376	17	5.7	630	30	G40597	Z7823 Zebrafish AB Dan 3.84e+01
c	377	17	5.7	810	30	HSSTS1QD	H.sapiens DNA for STS, 3.84e+01
c	378	17	5.7	1116	28	MUSCEREBN	Mus musculus Gcap1 mRN 3.84e+01
	379	17	5.7	1149	23	ATUBQ4	Arabidopsis thaliana U 3.84e+01
c	380	17	5.7	1426	18	DVU26720	Drosophila virilis sod 3.84e+01
c	381	17	5.7	1459	20	RANIDII	Rana catesbeiana iodot 3.84e+01
c	382	17	5.7	1545	26	HSFHIT01	Human diadenosine trip 3.84e+01
	383	17	5.7	1865	23	BNAACCOXI	Brassica napus amino-c 3.84e+01
c	384	17	5.7	1885	16	SSU95360	Sulfolobus solfataricus 3.84e+01
c	385	17	5.7	1903	27	HSF37P6	Human DNA sequence fro 3.84e+01
	386	17	5.7	2155	19	BTU23855	Bos taurus von Willebr 3.84e+01
	387	17	5.7	2215	18	S39048	knob associated histid 3.84e+01
c	388	17	5.7	2310	18	PFAALDO2A	Plasmodium berghei fru 3.84e+01
	389	17	5.7	2507	18	AF010264	Toxoplasma gondii apic 3.84e+01
c	390	17	5.7	2542	21	I89344	Sequence 1 from patent 3.84e+01
c	391	17	5.7	2580	23	AF015922	Arabidopsis thaliana f 3.84e+01
	392	17	5.7	2611	18	DMU1951	D. melanogaster U1-95. 3.84e+01
c	393	17	5.7	2897	28	MUSPEPAA	Mus musculus ferrochel 3.84e+01
	394	17	5.7	2946	23	ATAJ3189	Arabidopsis thaliana P 3.84e+01
	395	17	5.7	3059	23	ATHUBQC	Arabidopsis thaliana p 3.84e+01
	396	17	5.7	3310	25	HSLACTG	Human alpha-lactalbumi 3.84e+01
c	397	17	5.7	3607	18	CELK07D8	C. elegans cosmid K07D 3.84e+01
	398	17	5.7	3820	28	MMLAL1	Mus musculus lysosomal 3.84e+01
c	399	17	5.7	3830	28	MMMIACDRP	M.musculus MIA/CD-RAP 3.84e+01

c	400	17	5.7	3831	28	MMU85612	Mus musculus cartilage 3.84e+01
	401	17	5.7	4013	28	MMAF001794	Mus musculus Treacher 3.84e+01
	402	17	5.7	4044	28	AF047387	Mus musculus telomeras 3.84e+01
	403	17	5.7	4056	23	ATU33014	Arabidopsis thaliana p 3.84e+01
c	404	17	5.7	4325	18	PF38GBP	P.falciparum 3.8 gene 3.84e+01
	405	17	5.7	4428	28	MMUR10310	Mus musculus treacle (3.84e+01
c	406	17	5.7	4447	26	HSMCP3P	H.sapiens MCP-3 promot 3.84e+01
	407	17	5.7	4606	23	CSU82973	Citrus sinensis pectin 3.84e+01
	408	17	5.7	4648	18	DMTN1731	Drosophila melanogaste 3.84e+01
c	409	17	5.7	4804	26	HUMAPOLIAA	Human apolipoprotein (3.84e+01
c	410	17	5.7	5381	23	BNU55032	Brassica napus asparti 3.84e+01
	411	17	5.7	5784	20	CYIMYC2	Cyprinus carpio c-myc 3.84e+01
c	412	17	5.7	5837	21	I18671	Sequence 1 from patent 3.84e+01
c	413	17	5.7	5837	15	THCVDPE	Thermococcus litoralis 3.84e+01
c	414	17	5.7	5837	21	I18459	Sequence 1 from patent 3.84e+01
	415	17	5.7	6305	18	PFSC04060	Plasmodium falciparum 3.84e+01
c	416	17	5.7	6305	18	PFSC04060	Plasmodium falciparum 3.84e+01
	417	17	5.7	6362	29	TRBS3RRBN	Trypanosoma cruzi 5.8S 3.84e+01
	418	17	5.7	6629	24	AB016151	Adiantum capillus-vene 3.84e+01
c	419	17	5.7	6663	18	PFSC03007	Plasmodium falciparum 3.84e+01
	420	17	5.7	6711	30	G31647	sWSS821 Eric D. Green 3.84e+01
	421	17	5.7	6711	25	HSCPH70	Human cyclophilin gene 3.84e+01
	422	17	5.7	7576	18	PFSC04062	Plasmodium falciparum 3.84e+01
	423	17	5.7	7589	18	DRODOUGHA	Drosophila virilis rou 3.84e+01
c	424	17	5.7	8777	18	DMSING1	D. melanogaster singed 3.84e+01
	425	17	5.7	9094	18	PFSC03055	Plasmodium falciparum 3.84e+01
	426	17	5.7	9345	23	LHDEL	Lilium henryi del tran 3.84e+01
	427	17	5.7	9722	28	MUSACHRBB	Mouse nicotinic acetyl 3.84e+01
	428	17	5.7	10141	15	U67489	Methanococcus jannasch 3.84e+01
	429	17	5.7	10172	19	PIGPSHBS	Pig gene for follicle 3.84e+01
c	430	17	5.7	10194	19	GOTGLOBE	Goat beta globin locus 3.84e+01
	431	17	5.7	10213	18	PFU27338	Plasmodium falciparum 3.84e+01
c	432	17	5.7	11395	28	AF059275	Mus musculus heat shoc 3.84e+01
	433	17	5.7	12772	18	CELF53A10	Caenorhabditis elegans 3.84e+01
c	434	17	5.7	13851	18	DMU80064	Drosophila melanogaste 3.84e+01
	435	17	5.7	15446	15	AE000710	Aquifex aeolicus secti 3.84e+01
c	436	17	5.7	16299	28	RATCYP4A1	Rat cytochrome P-450 I 3.84e+01
c	437	17	5.7	16342	26	HSN85E10	Human DNA sequence fro 3.84e+01
	438	17	5.7	17850	26	AF011920	Homo sapiens protein k 3.84e+01
	439	17	5.7	18201	18	DVDNAPTP	D.virilis DNA for trit 3.84e+01
	440	17	5.7	18280	18	PFSC03029	Plasmodium falciparum 3.84e+01
	441	17	5.7	19994	25	HSTITINX5	H.sapiens titin gene e 3.84e+01
c	442	17	5.7	21791	18	PFSC03050	Plasmodium falciparum 3.84e+01
c	443	17	5.7	24054	18	CELK14A12	Caenorhabditis elegans 3.84e+01
c	444	17	5.7	24787	26	AB008656S2	Homo sapiens klotho ge 3.84e+01
	445	17	5.7	25739	17	AC005645	*** SEQUENCING IN PROG 3.84e+01
	446	17	5.7	28768	18	PFSC03020	Plasmodium falciparum 3.84e+01
	447	17	5.7	29548	18	CELF13A2	Caenorhabditis elegans 3.84e+01
	448	17	5.7	30315	18	CEF49H6	Caenorhabditis elegans 3.84e+01
	449	17	5.7	32009	18	CET06D8	Caenorhabditis elegans 3.84e+01
	450	17	5.7	32359	27	HSN120B6B	Human DNA sequence fro 3.84e+01

	451	17	5.7	33037	25	HSU07978	Human T cell receptor 3.84e+01
c	452	17	5.7	33784	18	CELC05D9	Caenorhabditis elegans 3.84e+01
c	453	17	5.7	34719	27	AC005569	Homo sapiens chromosom 3.84e+01
c	454	17	5.7	34831	18	CEC47G2	Caenorhabditis elegans 3.84e+01
	455	17	5.7	35039	25	HSU144A10	Human DNA sequence fro 3.84e+01
	456	17	5.7	35054	26	HS333B10	Human DNA sequence fro 3.84e+01
c	457	17	5.7	35491	26	AP048727	Homo sapiens minisatel 3.84e+01
c	458	17	5.7	36179	18	CELT20F7	Caenorhabditis elegans 3.84e+01
c	459	17	5.7	36238	26	D87015	Human (lambda) DNA for 3.84e+01
c	460	17	5.7	36342	26	AC005364	Homo sapiens chromosom 3.84e+01
	461	17	5.7	36418	26	AC004652	Homo sapiens chromosom 3.84e+01
c	462	17	5.7	37302	26	D87007	Human (lambda) DNA for 3.84e+01
c	463	17	5.7	37322	26	HSU237H1	Human DNA sequence fro 3.84e+01
c	464	17	5.7	37760	26	D86994	Human (lambda) DNA for 3.84e+01
c	465	17	5.7	37821	15	MLCB1770	Mycobacterium leprae c 3.84e+01
c	466	17	5.7	38356	18	CED1081	Caenorhabditis elegans 3.84e+01
c	467	17	5.7	39056	25	HSU96H1	Human DNA sequence fro 3.84e+01
	468	17	5.7	39519	18	CEF32D8	Caenorhabditis elegans 3.84e+01
	469	17	5.7	40016	26	AC003038	Human DNA from chromos 3.84e+01
c	470	17	5.7	40017	26	D86998	Human (lambda) DNA for 3.84e+01
	471	17	5.7	40982	26	AC004203	Homo sapiens clone UWG 3.84e+01
c	472	17	5.7	41468	27	AC005523	Homo sapiens chromosom 3.84e+01
c	473	17	5.7	42872	26	AC000395	Genomic sequence from 3.84e+01
	474	17	5.7	43104	26	HSAC000125	Human Cosmid g1572c037 3.84e+01
c	475	17	5.7	43467	26	AC000085	Homo sapiens Chromosom 3.84e+01
	476	17	5.7	43875	17	AC005432	*** SEQUENCING IN PROG 3.84e+01
c	477	17	5.7	44002	26	HSAP002995	Homo sapiens cosmid, c 3.84e+01
c	478	17	5.7	44042	24	SPCC736	S.pombe chromosome III 3.84e+01
c	479	17	5.7	44072	18	DMC121E7	Drosophila melanogaste 3.84e+01
c	480	17	5.7	45878	24	AB016883	Arabidopsis thaliana g 3.84e+01
c	481	17	5.7	46496	17	AC005646	*** SEQUENCING IN PROG 3.84e+01
	482	17	5.7	47017	26	AC004143	Homo sapiens chromosom 3.84e+01
	483	17	5.7	50135	17	AC005655	*** SEQUENCING IN PROG 3.84e+01
c	484	17	5.7	51167	17	AC005464	*** SEQUENCING IN PROG 3.84e+01
c	485	17	5.7	51476	26	HS5O3F6	Homo sapiens DNA seque 3.84e+01
c	486	17	5.7	56458	27	HS342B11	Human DNA sequence fro 3.84e+01
	487	17	5.7	56913	26	HS67C13	Human DNA sequence fro 3.84e+01
	488	17	5.7	60889	27	HS41P2B	Human DNA sequence fro 3.84e+01
	489	17	5.7	63749	26	AC004694	Homo sapiens BAC clone 3.84e+01
c	490	17	5.7	69037	28	AC003949	Mus musculus chromosom 3.84e+01
	491	17	5.7	70335	26	HS263021	Human DNA sequence fro 3.84e+01
c	492	17	5.7	72408	3	AC005123	*** SEQUENCING IN PROG 3.84e+01
c	493	17	5.7	73303	17	AC005526	*** SEQUENCING IN PROG 3.84e+01
	494	17	5.7	74560	26	HS433B8	Human DNA sequence fro 3.84e+01
c	495	17	5.7	77298	23	AB011483	Arabidopsis thaliana g 3.84e+01
c	496	17	5.7	78631	25	HSAC000057	Human BAC clone RG067M 3.84e+01
	497	17	5.7	79482	26	AP000029	Homo sapiens genomic D 3.84e+01
	498	17	5.7	79574	23	ATAC002340	Arabidopsis thaliana c 3.84e+01
c	499	17	5.7	80531	17	HSV460E1	Human DNA sequence *** 3.84e+01
	500	17	5.7	80679	23	ATM3E9	Arabidopsis thaliana D 3.84e+01
c	501	17	5.7	80770	23	AB005240	Arabidopsis thaliana g 3.84e+01

	502	17	5.7	81231	26	AC003044	Human PAC clone DJ1055 3.84e+01
c	503	17	5.7	81284	24	AB016880	Arabidopsis thaliana g 3.84e+01
c	504	17	5.7	81421	17	CEC53C7	Caenorhabditis elegans 3.84e+01
c	505	17	5.7	81625	18	AC004289	Drosophila melanogaste 3.84e+01
c	506	17	5.7	81741	26	AC005367	Homo sapiens chromosom 3.84e+01
	507	17	5.7	82291	27	AC005376	Homo sapiens PAC clone 3.84e+01
	508	17	5.7	82393	18	AC003923	Drosophila melanogaste 3.84e+01
c	509	17	5.7	82722	17	AC004372	*** SEQUENCING IN PROG 3.84e+01
c	510	17	5.7	83478	23	AB011482	Arabidopsis thaliana g 3.84e+01
	511	17	5.7	84129	24	AB013392	Arabidopsis thaliana g 3.84e+01
c	512	17	5.7	84582	3	AC004730	*** SEQUENCING IN PROG 3.84e+01
	513	17	5.7	85616	17	AC005113	*** SEQUENCING IN PROG 3.84e+01
c	514	17	5.7	85633	26	HSAC002088	Human BAC clone RG013P 3.84e+01
c	515	17	5.7	86567	18	AC004276	Drosophila melanogaste 3.84e+01
	516	17	5.7	87903	26	HS137D17	Human DNA sequence fro 3.84e+01
	517	17	5.7	88987	26	AC004541	Homo sapiens PAC clone 3.84e+01
c	518	17	5.7	89109	17	AC005131	*** SEQUENCING IN PROG 3.84e+01
c	519	17	5.7	91007	17	AC004374	*** SEQUENCING IN PROG 3.84e+01
c	520	17	5.7	91019	18	AC001658	Drosophila melanogaste 3.84e+01
c	521	17	5.7	91638	17	AC002489	*** SEQUENCING IN PROG 3.84e+01
c	522	17	5.7	92128	24	F17O7	Arabidopsis thaliana c 3.84e+01
c	523	17	5.7	92944	17	HSU242F8	Human DNA sequence *** 3.84e+01
	524	17	5.7	94255	26	AC003090	Human BAC clone RG241I 3.84e+01
	525	17	5.7	94665	17	AC004979	*** SEQUENCING IN PROG 3.84e+01
c	526	17	5.7	96675	17	AC004287	*** SEQUENCING IN PROG 3.84e+01
	527	17	5.7	97943	26	HSAC002467	Human BAC clone RG364P 3.84e+01
	528	17	5.7	97979	26	AC004739	Homo sapiens PAC clone 3.84e+01
	529	17	5.7	98274	27	HS714B7	Human DNA sequence fro 3.84e+01
	530	17	5.7	98937	24	ATAC002535	Arabidopsis thaliana c 3.84e+01
c	531	17	5.7	99445	17	HS201D17	Human DNA sequence *** 3.84e+01
	532	17	5.7	100625	26	HS1189K21	Human DNA sequence *** 3.84e+01
	533	17	5.7	101110	18	AC005555	Drosophila melanogaste 3.84e+01
c	534	17	5.7	101370	26	AC003962	Homo sapiens chromosom 3.84e+01
	535	17	5.7	103590	17	HS328E19	Human DNA sequence *** 3.84e+01
c	536	17	5.7	104016	24	ATAC004138	Arabidopsis thaliana c 3.84e+01
	537	17	5.7	104037	26	HSAP002992	Homo sapiens cosmid fr 3.84e+01
	538	17	5.7	104984	26	AC005246	Homo sapiens PAC clone 3.84e+01
	539	17	5.7	105580	25	HS36J3	Human DNA sequence fro 3.84e+01
	540	17	5.7	106377	26	AC004132	Homo sapiens chromosom 3.84e+01
c	541	17	5.7	108390	17	HS312L17	Human DNA sequence *** 3.84e+01
	542	17	5.7	109859	26	HS747E2	Homo sapiens DNA seque 3.84e+01
	543	17	5.7	110858	25	HUM2G3A	Human PAC clone 2G3A f 3.84e+01
c	544	17	5.7	112359	27	HSU60822	Human dystrophin (DMD) 3.84e+01
c	545	17	5.7	112575	17	HS715N11	Human DNA sequence *** 3.84e+01
	546	17	5.7	112690	17	HS19B17	Human DNA sequence *** 3.84e+01
c	547	17	5.7	114058	17	AC005009	*** SEQUENCING IN PROG 3.84e+01
c	548	17	5.7	114231	27	HS390C10	Homo sapiens DNA seque 3.84e+01
c	549	17	5.7	114495	17	HS112K5	Human DNA sequence *** 3.84e+01
c	550	17	5.7	114736	18	PFMAL3P3	Plasmodium falciparum 3.84e+01
	551	17	5.7	115905	26	HS57E3	Homo sapiens DNA seque 3.84e+01
	552	17	5.7	117161	17	CEY43F11	Caenorhabditis elegans 3.84e+01

	553	17	5.7	117721	27	HS113J7	Human DNA sequence fro 3.84e+01
c	554	17	5.7	118001	17	HS566L20	Human DNA sequence *** 3.84e+01
	555	17	5.7	118226	26	HS466I8	Homo sapiens DNA seque 3.84e+01
	556	17	5.7	118308	17	AC005316	*** SEQUENCING IN PROG 3.84e+01
	557	17	5.7	118313	26	AC003991	Human BAC clone RG167B 3.84e+01
c	558	17	5.7	118595	26	AC002379	Human BAC clone GS165I 3.84e+01
c	559	17	5.7	119865	28	AC004407	Mus musculus mbac6, co 3.84e+01
c	560	17	5.7	120376	26	HS271G9	Human DNA sequence fro 3.84e+01
c	561	17	5.7	121236	17	AC005097	*** SEQUENCING IN PROG 3.84e+01
	562	17	5.7	123585	26	AC004707	Homo sapiens chromosom 3.84e+01
	563	17	5.7	124519	28	AC003019	Mus musculus Bac B6, c 3.84e+01
c	564	17	5.7	125179	24	ATT4L20	Arabidopsis thaliana D 3.84e+01
c	565	17	5.7	125930	17	HS124K22	Human DNA sequence *** 3.84e+01
	566	17	5.7	126147	26	HS433M19	Human DNA sequence fro 3.84e+01
c	567	17	5.7	126198	26	AC003089	Human BAC clone RG180F 3.84e+01
c	568	17	5.7	127947	26	AC003683	Homo sapiens Xp22 PAC 3.84e+01
	569	17	5.7	128117	26	AC004000	Human PAC clone DJ404P 3.84e+01
c	570	17	5.7	129098	26	HS46H23	Human DNA sequence fro 3.84e+01
c	571	17	5.7	129169	17	AC005506	*** SEQUENCING IN PROG 3.84e+01
	572	17	5.7	130381	26	HUAC003003	Human Chromosome 16 BA 3.84e+01
	573	17	5.7	133457	26	AC003999	Human PAC clone DJ1139 3.84e+01
c	574	17	5.7	133801	26	AC004585	Homo sapiens chromosom 3.84e+01
	575	17	5.7	134743	26	HUAC002300	Homo sapiens Chromosom 3.84e+01
c	576	17	5.7	135038	25	HUMYWXD703	Homo sapiens ADP/ATP c 3.84e+01
	577	17	5.7	135259	17	AC005018	*** SEQUENCING IN PROG 3.84e+01
c	578	17	5.7	135259	26	AC004590	Homo sapiens chromosom 3.84e+01
c	579	17	5.7	135448	17	AC004900	*** SEQUENCING IN PROG 3.84e+01
	580	17	5.7	135891	17	HS221G9	Human DNA sequence *** 3.84e+01
c	581	17	5.7	136150	27	AC002081	Homo sapiens BAC clone 3.84e+01
c	582	17	5.7	136561	17	HS430A16	Human DNA sequence *** 3.84e+01
	583	17	5.7	136650	17	HS769D20	Human DNA sequence *** 3.84e+01
c	584	17	5.7	137022	26	AC004231	Homo sapiens chromosom 3.84e+01
c	585	17	5.7	138343	17	AC004907	*** SEQUENCING IN PROG 3.84e+01
c	586	17	5.7	138706	17	AC004946	*** SEQUENCING IN PROG 3.84e+01
c	587	17	5.7	139919	26	AC002433	Human BAC clone RG317M 3.84e+01
c	588	17	5.7	140000	17	AC004047	*** SEQUENCING IN PROG 3.84e+01
	589	17	5.7	140210	17	AC002993	*** SEQUENCING IN PROG 3.84e+01
c	590	17	5.7	141779	26	AC002377	Human PAC clone DJ222H 3.84e+01
c	591	17	5.7	142949	17	AC004960	*** SEQUENCING IN PROG 3.84e+01
	592	17	5.7	143092	18	CEY17G7B	Caenorhabditis elegans 3.84e+01
c	593	17	5.7	143701	27	HS29C18	Homo sapiens DNA seque 3.84e+01
	594	17	5.7	144474	26	AF051934	Homo sapiens chromosom 3.84e+01
	595	17	5.7	147416	27	AC005549	Homo sapiens chromosom 3.84e+01
c	596	17	5.7	149526	26	HS411B6	Human DNA sequence fro 3.84e+01
	597	17	5.7	149554	25	HSAC000111	Human BAC clone 068P20 3.84e+01
	598	17	5.7	150000	17	AC004704	*** SEQUENCING IN PROG 3.84e+01
c	599	17	5.7	150228	26	AC003071	Human BAC clone BK085E 3.84e+01
c	600	17	5.7	151111	26	AC002308	Homo sapiens Chromosom 3.84e+01
c	601	17	5.7	151166	17	AC005520	*** SEQUENCING IN PROG 3.84e+01
	602	17	5.7	152077	26	HS130G2	Human DNA sequence fro 3.84e+01
c	603	17	5.7	152184	17	HS593F22	Human DNA sequence *** 3.84e+01

	604	17	5.7	155335	18	AC005734	Drosophila melanogaste 3.84e+01
c	605	17	5.7	155335	18	AC005734	Drosophila melanogaste 3.84e+01
c	606	17	5.7	155529	17	HS738P11	Human DNA sequence *** 3.84e+01
c	607	17	5.7	158427	26	AC002553	Homo sapiens chromosom 3.84e+01
c	608	17	5.7	159640	17	HS28I24	Human DNA sequence *** 3.84e+01
	609	17	5.7	161146	17	HS323A16	Human DNA sequence *** 3.84e+01
	610	17	5.7	162524	17	HS498I24	Human DNA sequence *** 3.84e+01
c	611	17	5.7	164536	17	HS134N8	Human DNA sequence *** 3.84e+01
	612	17	5.7	165363	17	HS459L4	Human DNA sequence *** 3.84e+01
c	613	17	5.7	166531	26	HSAC002123	Human BAC clone GS312P 3.84e+01
c	614	17	5.7	167372	27	AC004945	Homo sapiens PAC clone 3.84e+01
	615	17	5.7	167575	17	HS531H16	Human DNA sequence *** 3.84e+01
c	616	17	5.7	168056	26	AC004458	Homo sapiens BAC clone 3.84e+01
	617	17	5.7	170000	17	AC004066	*** SEQUENCING IN PROG 3.84e+01
c	618	17	5.7	170952	26	HS501N12	Homo sapiens DNA seque 3.84e+01
c	619	17	5.7	171188	17	AC005025	*** SEQUENCING IN PROG 3.84e+01
	620	17	5.7	172533	26	AC002381	Human BAC clone RG020D 3.84e+01
	621	17	5.7	172983	26	AC002454	Human BAC clone GS552A 3.84e+01
c	622	17	5.7	173513	26	HS125H2	Human DNA sequence fro 3.84e+01
c	623	17	5.7	174098	27	AC005737	Homo sapiens chromosom 3.84e+01
c	624	17	5.7	176109	26	AC003992	Human BAC clone RG250D 3.84e+01
	625	17	5.7	176154	17	AC005317	*** SEQUENCING IN PROG 3.84e+01
	626	17	5.7	176511	27	AC003035	Homo sapiens Xp22 bins 3.84e+01
c	627	17	5.7	177311	26	AC004385	Human Chromosome X clo 3.84e+01
c	628	17	5.7	177328	17	HS462D8	Human DNA sequence *** 3.84e+01
	629	17	5.7	180551	26	HUAC004158	Homo sapiens Chromosom 3.84e+01
c	630	17	5.7	186550	17	CEK44F5	Caenorhabditis elegans 3.84e+01
c	631	17	5.7	189134	26	HUAC004682	Homo sapiens Chromosom 3.84e+01
	632	17	5.7	193707	17	AC005035	*** SEQUENCING IN PROG 3.84e+01
	633	17	5.7	195016	17	HS215F16	Human DNA sequence *** 3.84e+01
c	634	17	5.7	197391	17	AC005140	*** SEQUENCING IN PROG 3.84e+01
	635	17	5.7	198776	17	AC005075	*** SEQUENCING IN PROG 3.84e+01
	636	17	5.7	199878	17	AC005027	*** SEQUENCING IN PROG 3.84e+01
c	637	17	5.7	200000	17	AC005510	*** SEQUENCING IN PROG 3.84e+01
	638	17	5.7	202763	26	AC005181	Homo sapiens chromosom 3.84e+01
	639	17	5.7	204877	17	AC005375	*** SEQUENCING IN PROG 3.84e+01
	640	17	5.7	207139	17	CEY17G7	Caenorhabditis elegans 3.84e+01
c	641	17	5.7	210216	17	AC004888	*** SEQUENCING IN PROG 3.84e+01
c	642	17	5.7	210672	17	HS179I15	Human DNA sequence *** 3.84e+01
	643	17	5.7	236157	17	AC005703	*** SEQUENCING IN PROG 3.84e+01
c	644	17	5.7	236157	17	AC005703	*** SEQUENCING IN PROG 3.84e+01
c	645	17	5.7	244649	17	AC005308	*** SEQUENCING IN PROG 3.84e+01
c	646	17	5.7	245265	17	HS581F12	Human DNA sequence *** 3.84e+01
	647	17	5.7	251364	28	MMAE000664	Mus musculus TCR beta 3.84e+01
c	648	17	5.7	251364	28	MMAE000664	Mus musculus TCR beta 3.84e+01
	649	17	5.7	259202	26	AC002366	Human Xp22 BAC GT-285I 3.84e+01
	650	17	5.7	261604	17	AC005531	*** SEQUENCING IN PROG 3.84e+01
	651	17	5.7	267156	26	U66059	Human germline T-cell 3.84e+01
	652	17	5.7	280600	17	HS461P17	Human DNA sequence *** 3.84e+01
c	653	17	5.7	323900	17	HS402G11	Human DNA sequence *** 3.84e+01
	654	17	5.7	348050	26	HS229042	Homo sapiens 959 kb co 3.84e+01

c	655	16	5.4	157	30	HS043XP2	H.sapiens DNA segment 1.70e+02
c	656	16	5.4	220	30	HS074WH11	H.sapiens DNA segment 1.70e+02
	657	16	5.4	235	20	CHKMICBJ	Gallus domesticus micr 1.70e+02
	658	16	5.4	239	30	G13560	human STS SHGC-11171. 1.70e+02
c	659	16	5.4	252	30	DM40A3T	D. melanogaster STS de 1.70e+02
	660	16	5.4	281	19	ECMSASB11	E.caballus microsatell 1.70e+02
c	661	16	5.4	281	30	G31867	137C7 B Human (E.S.Sri 1.70e+02
	662	16	5.4	282	30	AF033527	Hyla regilla repeat re 1.70e+02
	663	16	5.4	292	19	BTMICSA	B.taurus microsatellit 1.70e+02
c	664	16	5.4	292	28	CORMSCRBE	Grey red-backed vole m 1.70e+02
c	665	16	5.4	301	30	HS218XH8	H. sapiens (D4S421) DN 1.70e+02
c	666	16	5.4	301	30	HSB309XE1	H.sapiens (D1S2757) DN 1.70e+02
c	667	16	5.4	307	30	HSB317EG1	H.sapiens (D2S2390) DN 1.70e+02
c	668	16	5.4	313	30	HS206ZF4	H.sapiens (D3S3511) DN 1.70e+02
c	669	16	5.4	314	30	HS320XH1	H. sapiens (D11S1347) 1.70e+02
c	670	16	5.4	316	30	HSB059WG9	H.sapiens (D19S907) DN 1.70e+02
c	671	16	5.4	318	30	HS203EC1	H. sapiens (D8S509) DN 1.70e+02
	672	16	5.4	325	30	G41383	Z1492 Zebrafish AB Dan 1.70e+02
c	673	16	5.4	325	30	HS210XC11	H. sapiens (D13S169) D 1.70e+02
c	674	16	5.4	329	18	DROMDGX2A1	Drosophila melanogaste 1.70e+02
c	675	16	5.4	331	23	CEU93389	Coffea eugenioides chl 1.70e+02
c	676	16	5.4	332	30	HS266WB9	H. sapiens (D4S432) DN 1.70e+02
c	677	16	5.4	336	30	G21844	human STS WI-11103 1.70e+02
c	678	16	5.4	337	30	HSB340YG1	H.sapiens (D8S1785) DN 1.70e+02
	679	16	5.4	339	30	G12402	human STS CHLC.GATA31H 1.70e+02
	680	16	5.4	349	19	SSU06775	Sus scrofa clone sw617 1.70e+02
c	681	16	5.4	349	18	DROMDG231	D.melanogaster mobile 1.70e+02
c	682	16	5.4	360	23	CSU93398	Coffea stenophylla chl 1.70e+02
c	683	16	5.4	360	23	CHU93399	Coffea humilis chlorop 1.70e+02
c	684	16	5.4	360	23	CMU93400	Coffea millotii chloro 1.70e+02
c	685	16	5.4	360	23	CCU93393	Coffea canephora chlor 1.70e+02
c	686	16	5.4	360	23	CAU93387	Coffea arabica chlorop 1.70e+02
c	687	16	5.4	360	23	PEU93404	Psilanthus ebracteolat 1.70e+02
c	688	16	5.4	360	23	CBU93401	Coffea bertrandi chlor 1.70e+02
c	689	16	5.4	363	19	MSU70463	Mesoplodon stejnegeri 1.70e+02
c	690	16	5.4	363	30	G08875	human STS CHLC.GATA4E0 1.70e+02
c	691	16	5.4	365	30	HSA054ZH9	H.sapiens (D6S1695) DN 1.70e+02
c	692	16	5.4	368	30	HS256ZE9	H.sapiens DNA segment 1.70e+02
	693	16	5.4	395	30	AF043206	Equus caballus microsa 1.70e+02
	694	16	5.4	400	30	G07698	human STS SHGC-9430 cl 1.70e+02
c	695	16	5.4	400	30	G17164	human STS SHGC-13645 c 1.70e+02
c	696	16	5.4	404	30	G17462	human STS SHGC-7224 cl 1.70e+02
c	697	16	5.4	406	30	G08054	human STS CHLC.GATA65E 1.70e+02
c	698	16	5.4	416	30	HUMUT5625	Human STS UT5625, sequ 1.70e+02
	699	16	5.4	432	20	SCU85821	Struthio camelus polym 1.70e+02
c	700	16	5.4	445	30	G38415	SHGC-58070 Human Homo 1.70e+02
c	701	16	5.4	480	30	HUMUT7099	Human STS UT7099. 1.70e+02
c	702	16	5.4	484	30	HUMUT2538	Human STS UT2538, sequ 1.70e+02
	703	16	5.4	513	25	HUMCOLREPA	Human variable number 1.70e+02
c	704	16	5.4	536	28	S81886	RT1.E/C: major histoco 1.70e+02
	705	16	5.4	582	30	G39890	Z11320 Zebrafish AB Da 1.70e+02

	706	16	5.4	584	30	G40504	Z7121 Zebrafish AB Dan 1.70e+02
	707	16	5.4	609	25	HUMCOLREPD	Human variable number 1.70e+02
	708	16	5.4	638	25	HUMCOLREPE	Human variable number 1.70e+02
	709	16	5.4	671	25	HUMCOLREPG	Human variable number 1.70e+02
	710	16	5.4	674	25	HUMCOLREPJ	Human variable number 1.70e+02
	711	16	5.4	677	26	HS9QT056	H.sapiens telomeric DN 1.70e+02
	712	16	5.4	677	25	HUMCOLREPH	Human variable number 1.70e+02
	713	16	5.4	686	25	HUMCOLREPL	Human variable number 1.70e+02
	714	16	5.4	699	25	HUMCOLREPP	Human variable number 1.70e+02
	715	16	5.4	739	25	HUMCOLREPR	Human variable number 1.70e+02
	716	16	5.4	767	25	HUMCOLREPT	Human variable number 1.70e+02
c	717	16	5.4	838	23	CCLPROA	Cochliobolus heterostr 1.70e+02
	718	16	5.4	1098	23	D55711	Oryza sativa mRNA for 1.70e+02
	719	16	5.4	1106	24	SCYDL120W	S.cerevisiae chromosom 1.70e+02
	720	16	5.4	1109	23	AB003195	Oryza sativa mRNA for 1.70e+02
c	721	16	5.4	1176	28	MMU17112	Mus musculus casein ki 1.70e+02
c	722	16	5.4	1322	23	ADU21696	Ahnfeltiopsis durvilla 1.70e+02
c	723	16	5.4	1329	18	PFADD2C	Plasmodium falciparum 1.70e+02
c	724	16	5.4	1513	20	CHKCLIIAS	Chicken casein kinase 1.70e+02
	725	16	5.4	1530	28	MUSPRIMP	Mouse putative primord 1.70e+02
	726	16	5.4	1558	28	MUSTFNFE2X	Mouse transcription fa 1.70e+02
c	727	16	5.4	1560	28	MMBGLMAJ	Mouse beta-maj globin 1.70e+02
	728	16	5.4	1665	23	PHU172B	Pelargonium hortorum c 1.70e+02
c	729	16	5.4	1669	28	MUSPRLRA	Mouse pseudo-prolactin 1.70e+02
c	730	16	5.4	1674	20	GGOV01	Gallus gallus fragment 1.70e+02
c	731	16	5.4	1808	23	PSGDCHAA	P.sativum gdcH gene fo 1.70e+02
	732	16	5.4	1932	28	MMPFK2KB	M.musculus mRNA for 1.70e+02
	733	16	5.4	2133	23	SCYDL119C	S.cerevisiae chromosom 1.70e+02
c	734	16	5.4	2194	23	SCU46559	Saccharomyces cerevisi 1.70e+02
c	735	16	5.4	2298	18	DMU33462	Drosophila melanogaste 1.70e+02
	736	16	5.4	2354	33	PPMVP1M	Monkey B-lymphotropic 1.70e+02
c	737	16	5.4	2383	18	PFAALD	P.falciparum aldolase 1.70e+02
c	738	16	5.4	2383	21	A13481	P.falciparum gene for 1.70e+02
	739	16	5.4	2393	25	HUMPSHRE	Human follicle stimula 1.70e+02
	740	16	5.4	2461	25	HSCOLII2	Human alpha 1(II) coll 1.70e+02
	741	16	5.4	2480	27	HUMCOL3IX	Homo sapiens collagen 1.70e+02
c	742	16	5.4	2513	28	RATUGPS	Rat U1 small nuclear R 1.70e+02
c	743	16	5.4	2925	28	RNU32498	Rattus norvegicus rsec 1.70e+02
c	744	16	5.4	2957	23	SCYLR006C	S.cerevisiae chromosom 1.70e+02
	745	16	5.4	2976	23	PSLEGJ	Pisum sativum legJ gen 1.70e+02
	746	16	5.4	2994	27	HSAJ1689	Homo sapiens NKG2D gen 1.70e+02
c	747	16	5.4	3105	26	HSL61724	Homo sapiens (subclone 1.70e+02
	748	16	5.4	3312	20	CHKPLB	Chicken cardiac phosph 1.70e+02
	749	16	5.4	3423	20	CHKPLB2	Chicken phospholamban 1.70e+02
c	750	16	5.4	3454	15	BACBCGIB	Bacillus coagulans res 1.70e+02
c	751	16	5.4	3510	23	SCYDL241W	S.cerevisiae chromosom 1.70e+02
c	752	16	5.4	3690	20	S82457	XmN-cadherin=maternall 1.70e+02
c	753	16	5.4	3915	26	AC002210	Homo sapiens (subclone 1.70e+02
c	754	16	5.4	4128	28	RNU11031	Rattus norvegicus neur 1.70e+02
c	755	16	5.4	4273	23	SCLRG1C	S.cerevisiae (DBY874) 1.70e+02
c	756	16	5.4	4527	20	AF074609	Rattus norvegicus MHC 1.70e+02

	757	16	5.4	4654	28	MMU15139	Mus musculus tyrosine	1.70e+02
c	758	16	5.4	4773	33	PPVPGP	Plum Pox Virus partial	1.70e+02
c	759	16	5.4	5089	33	PPMCG	Monkey B-lymphotropic	1.70e+02
	760	16	5.4	5173	28	MMU83913	Mus musculus prolifera	1.70e+02
	761	16	5.4	5180	25	S78271	SB1.8/DXS423E-mitosis-	1.70e+02
	762	16	5.4	5191	28	MMU28789	Mus musculus p53-assoc	1.70e+02
c	763	16	5.4	5496	18	CTBR3	C.tentans balbiani rin	1.70e+02
c	764	16	5.4	5499	23	LJNIA	L.japonicus NIA gene.	1.70e+02
	765	16	5.4	5803	26	D80000	Human mRNA for KIAA017	1.70e+02
c	766	16	5.4	5821	28	MMMB414A2	Mus musculus X inactiv	1.70e+02
	767	16	5.4	6060	26	D87077	Human mRNA for KIAA024	1.70e+02
	768	16	5.4	6314	28	RNU40790	Rattus norvegicus vasc	1.70e+02
c	769	16	5.4	6532	28	MUSHBBMAJ	Mouse beta-globin majo	1.70e+02
	770	16	5.4	7200	23	STWIN12G	Potato wound-induced g	1.70e+02
c	771	16	5.4	8003	23	FAS99762	Flaveria anomala gdcsP	1.70e+02
c	772	16	5.4	8017	24	ATU76670	Arabidopsis thaliana r	1.70e+02
	773	16	5.4	8054	15	BACAMOKOOO	Bacillus subtilis gene	1.70e+02
c	774	16	5.4	8844	23	SPBC32C12	S.pombe chromosome II	1.70e+02
c	775	16	5.4	8856	23	SOU69142	Spinacia oleracea beta	1.70e+02
	776	16	5.4	9618	27	HUMKALLIST	Homo sapiens kallistat	1.70e+02
	777	16	5.4	9679	18	LEITLE	Leishmania infantum tr	1.70e+02
	778	16	5.4	9782	18	DMU10507	Drosophila melanogaste	1.70e+02
c	779	16	5.4	9840	23	SCYBR140C	S.cerevisiae chromosom	1.70e+02
c	780	16	5.4	10187	23	PFGDCSPB	P.pringlei gdcsPB gene	1.70e+02
	781	16	5.4	12125	18	PFSC04053	Plasmodium falciparum	1.70e+02
c	782	16	5.4	12942	18	CELF58H12	Caenorhabditis elegans	1.70e+02
c	783	16	5.4	13187	28	BTPRMTNP2	R.norvegicus PRM1, PRM	1.70e+02
	784	16	5.4	15016	25	HSATPCP2	H.sapiens gene for mit	1.70e+02
	785	16	5.4	16390	28	MUSSEQAA	Mus musculus (clone R2	1.70e+02
	786	16	5.4	16851	25	HSL69F7B	Human DNA sequence fro	1.70e+02
c	787	16	5.4	17310	22	AF034975	Bacteriophage H-19B es	1.70e+02
	788	16	5.4	18808	19	BTFLUP1	B.taurus lactoferrin g	1.70e+02
c	789	16	5.4	19124	18	PFAVAR1A	Plasmodium falciparum	1.70e+02
	790	16	5.4	20866	25	HSU62292	Human elastin (ELN) ge	1.70e+02
	791	16	5.4	21877	18	DVU95037	Drosophila decapentapl	1.70e+02
	792	16	5.4	23901	23	SCN201952	S.cerevisiae N2019, N2	1.70e+02
	793	16	5.4	25864	17	PFMAL1P5	Plasmodium falciparum	1.70e+02
c	794	16	5.4	27016	18	CEF40E10	Caenorhabditis elegans	1.70e+02
	795	16	5.4	27207	18	PFSC03017	Plasmodium falciparum	1.70e+02
c	796	16	5.4	28762	18	CEZC376	Caenorhabditis elegans	1.70e+02
	797	16	5.4	28992	23	SC9375	S.cerevisiae chromosom	1.70e+02
c	798	16	5.4	29103	25	HSL165D7	Human DNA sequence fro	1.70e+02
	799	16	5.4	30098	18	CEK09E9	Caenorhabditis elegans	1.70e+02
	800	16	5.4	31235	18	CE2K270	Caenorhabditis elegans	1.70e+02
c	801	16	5.4	32195	18	ACC04154	Drosophila melanogaste	1.70e+02
c	802	16	5.4	32911	18	CELT23C6	Caenorhabditis elegans	1.70e+02
c	803	16	5.4	33202	18	CEC55A1	Caenorhabditis elegans	1.70e+02
c	804	16	5.4	33735	18	CELT13A10	Caenorhabditis elegans	1.70e+02
c	805	16	5.4	33796	18	CELF52E4	Caenorhabditis elegans	1.70e+02
	806	16	5.4	34486	24	SPBC23E6	S.pombe chromosome II	1.70e+02
	807	16	5.4	35694	18	CICOS41	Ciona intestinalis DNA	1.70e+02

	808	16	5.4	36393	26	D87013	Human (lambda) DNA for	1.70e+02
c	809	16	5.4	36493	18	CEF19C6	Caenorhabditis elegans	1.70e+02
	810	16	5.4	36772	18	CELT14B4	Caenorhabditis elegans	1.70e+02
c	811	16	5.4	37322	18	CEK10D11	Caenorhabditis elegans	1.70e+02
c	812	16	5.4	37628	26	HSD50P11	Human DNA sequence fro	1.70e+02
	813	16	5.4	38000	26	AF025422	Homo sapiens chromosom	1.70e+02
c	814	16	5.4	39280	26	AC000027	Homo sapiens Chromosom	1.70e+02
	815	16	5.4	39489	25	HSU69570	Human Xp22 cosmid U27H	1.70e+02
c	816	16	5.4	39906	27	HSU71217	Human clone c15H12, 24	1.70e+02
c	817	16	5.4	41169	18	CELE02C12	Caenorhabditis elegans	1.70e+02
	818	16	5.4	42246	26	HSU151E3	Human DNA sequence fro	1.70e+02
c	819	16	5.4	42493	27	HSCB1E7	Human DNA sequence fro	1.70e+02
c	820	16	5.4	42934	23	SPBC947	S.pombe chromosome II	1.70e+02
	821	16	5.4	43576	28	AC003984	Mouse cosmid ma53a271	1.70e+02
	822	16	5.4	43584	18	CELB0207	Caenorhabditis elegans	1.70e+02
	823	16	5.4	43772	26	AC004791	Homo sapiens chromoso	1.70e+02
c	824	16	5.4	44451	18	CELC06E7	Caenorhabditis elegans	1.70e+02
	825	16	5.4	47167	27	HS757P12	Human DNA sequence fro	1.70e+02
c	826	16	5.4	47736	18	AC004353	Drosophila melanogaste	1.70e+02
	827	16	5.4	61267	18	AC004336	Drosophila melanogaste	1.70e+02
c	828	16	5.4	61852	18	AC004445	Drosophila melanogaste	1.70e+02
c	829	16	5.4	65691	18	PFMAL3P1	Plasmodium falciparum	1.70e+02
c	830	16	5.4	66991	18	AC001653	Drosophila melanogaste	1.70e+02
	831	16	5.4	67046	26	HSU63721	Human elastin (ELN) ge	1.70e+02
	832	16	5.4	72546	17	AC005445	*** SEQUENCING IN PROG	1.70e+02
	833	16	5.4	72847	18	AC004266	Drosophila melanogaste	1.70e+02
c	834	16	5.4	72847	18	AC004266	Drosophila melanogaste	1.70e+02
c	835	16	5.4	73239	26	HS373H7	Human DNA sequence fro	1.70e+02
c	836	16	5.4	74507	26	AC004031	Homo sapiens 12q24 PAC	1.70e+02
	837	16	5.4	77051	17	AC004437	*** SEQUENCING IN PROG	1.70e+02
	838	16	5.4	78748	28	AC000400	Genomic sequence from	1.70e+02
c	839	16	5.4	79612	26	AC004073	Human Chromosome X, co	1.70e+02
c	840	16	5.4	79884	18	AC002502	Drosophila melanogaste	1.70e+02
	841	16	5.4	80446	18	AC004288	Drosophila melanogaste	1.70e+02
c	842	16	5.4	80659	27	AC004553	Homo sapiens Xp22 bin	1.70e+02
	843	16	5.4	81284	24	AB016880	Arabidopsis thaliana g	1.70e+02
c	844	16	5.4	81505	17	AC005133	*** SEQUENCING IN PROG	1.70e+02
c	845	16	5.4	81677	18	AC004377	Drosophila melanogaste	1.70e+02
c	846	16	5.4	81888	17	AC004733	*** SEQUENCING IN PROG	1.70e+02
c	847	16	5.4	82109	18	AC003925	Drosophila melanogaste	1.70e+02
	848	16	5.4	82225	25	HS253J14	Human DNA sequence fro	1.70e+02
	849	16	5.4	82251	26	AC004506	Homo sapiens chromosom	1.70e+02
c	850	16	5.4	82692	18	AC003924	Drosophila melanogaste	1.70e+02
c	851	16	5.4	83467	17	AC005470	*** SEQUENCING IN PROG	1.70e+02
c	852	16	5.4	84055	18	AC004439	Drosophila melanogaste	1.70e+02
c	853	16	5.4	85439	18	AC004728	Drosophila melanogaste	1.70e+02
	854	16	5.4	86686	24	ATAC004786	Arabidopsis thaliana c	1.70e+02
	855	16	5.4	87050	26	AC004763	Homo sapiens chromosom	1.70e+02
	856	16	5.4	88605	17	AC005446	*** SEQUENCING IN PROG	1.70e+02
c	857	16	5.4	89743	28	AC003062	Mouse Chromosome 16 Re	1.70e+02
c	858	16	5.4	91017	26	HS162C6	Human DNA sequence fro	1.70e+02

	859	16	5.4	91640	27	HS426F10	Human DNA sequence fro	1.70e+02
	860	16	5.4	92797	26	AF038458	Homo sapiens DNA from	1.70e+02
c	861	16	5.4	93695	23	ATF21G9	Arabidopsis thaliana B	1.70e+02
c	862	16	5.4	93821	26	AC004254	Homo sapiens BAC clone	1.70e+02
c	863	16	5.4	95643	24	F6N15	Arabidopsis thaliana B	1.70e+02
	864	16	5.4	95855	25	HSAC000115	Human BAC clone GS188P	1.70e+02
c	865	16	5.4	96180	17	AC005065	*** SEQUENCING IN PROG	1.70e+02
	866	16	5.4	96281	17	HS343K2	Human DNA sequence ***	1.70e+02
	867	16	5.4	97129	26	HS292H14	Human DNA sequence fro	1.70e+02
c	868	16	5.4	97169	17	AC004924	*** SEQUENCING IN PROG	1.70e+02
c	869	16	5.4	98025	26	AF003530	Homo sapiens homeobox	1.70e+02
	870	16	5.4	98695	18	AC005556	Drosophila melanogaste	1.70e+02
	871	16	5.4	100000	26	AP000018	Homo sapiens genomic D	1.70e+02
c	872	16	5.4	101713	17	AC005518	*** SEQUENCING IN PROG	1.70e+02
	873	16	5.4	101986	17	HS317E23	Human DNA sequence ***	1.70e+02
	874	16	5.4	103576	23	YUP8H12	Arabidopsis thaliana c	1.70e+02
c	875	16	5.4	104793	26	AF007262	Human Xq28 cosmid 3C3	1.70e+02
c	876	16	5.4	104871	27	AC004584	Homo sapiens chromosom	1.70e+02
c	877	16	5.4	106274	17	AC004390	*** SEQUENCING IN PROG	1.70e+02
	878	16	5.4	108393	24	ATAC005315	Arabidopsis thaliana c	1.70e+02
c	879	16	5.4	108534	26	HS124O9	Human DNA sequence fro	1.70e+02
c	880	16	5.4	109201	26	AC000403	Genomic sequence from	1.70e+02
	881	16	5.4	110000	17	CEY113G7	Caenorhabditis elegans	1.70e+02
c	882	16	5.4	110401	26	AC004068	Homo sapiens chromosom	1.70e+02
c	883	16	5.4	110970	28	AC004807	Mus musculus chromosom	1.70e+02
	884	16	5.4	111398	26	HS381E2	Human DNA sequence fro	1.70e+02
	885	16	5.4	113033	26	AC005070	Homo sapiens BAC clone	1.70e+02
c	886	16	5.4	114695	17	AC004918	*** SEQUENCING IN PROG	1.70e+02
c	887	16	5.4	114784	27	HS1189B24	Human DNA sequence fro	1.70e+02
	888	16	5.4	115863	27	HS268D13	Human DNA sequence ***	1.70e+02
	889	16	5.4	118276	27	AC004148	Homo sapiens chromosom	1.70e+02
c	890	16	5.4	118308	17	AC005316	*** SEQUENCING IN PROG	1.70e+02
c	891	16	5.4	118387	17	AC005628	*** SEQUENCING IN PROG	1.70e+02
	892	16	5.4	118552	26	HUACO04514	Homo sapiens Chromosom	1.70e+02
	893	16	5.4	118595	26	AC002379	Human BAC clone GS165I	1.70e+02
	894	16	5.4	118997	27	HS111J24	Human DNA sequence fro	1.70e+02
	895	16	5.4	119566	25	HS274L7	Human DNA sequence fro	1.70e+02
	896	16	5.4	119865	28	AC004407	Mus musculus mbac6, co	1.70e+02
c	897	16	5.4	120387	26	AC003051	Homo sapiens chromosom	1.70e+02
c	898	16	5.4	122532	17	HS232D4	Human DNA sequence ***	1.70e+02
c	899	16	5.4	122638	26	AB001523	Homo sapiens gene for	1.70e+02
c	900	16	5.4	122888	27	HS333H23	Human DNA sequence ***	1.70e+02
c	901	16	5.4	123435	17	CEY61B8	Caenorhabditis elegans	1.70e+02
c	902	16	5.4	123923	17	AC005677	*** SEQUENCING IN PROG	1.70e+02
	903	16	5.4	125301	26	HS388N15	Human DNA sequence fro	1.70e+02
c	904	16	5.4	126295	27	AC004139	Homo sapiens chromosom	1.70e+02
c	905	16	5.4	126474	26	HS181N1	Human DNA sequence fro	1.70e+02
	906	16	5.4	127623	17	AC005378	*** SEQUENCING IN PROG	1.70e+02
	907	16	5.4	128026	28	AC004155	Mus musculus DNA from	1.70e+02
c	908	16	5.4	129964	17	HS688G8	Human DNA sequence ***	1.70e+02
c	909	16	5.4	130626	26	HS1114G22	Human DNA sequence fro	1.70e+02

	910	16	5.4	130753	26	HS510L9	Homo sapiens DNA seque	1.70e+02
c	911	16	5.4	131805	26	HS339A18	Human DNA sequence fro	1.70e+02
	912	16	5.4	132960	26	AC002403	Human Chromosome 11p15	1.70e+02
c	913	16	5.4	133746	26	AP000022	Homo sapiens genomic D	1.70e+02
	914	16	5.4	133925	26	AC004562	Homo sapiens chromosom	1.70e+02
c	915	16	5.4	134450	26	HUAC003049	Homo sapiens Chromosom	1.70e+02
	916	16	5.4	135686	26	HS473J6	Human DNA sequence fro	1.70e+02
	917	16	5.4	136805	17	AC005507	*** SEQUENCING IN PROG	1.70e+02
c	918	16	5.4	138799	26	AC004056	Homo sapiens chromosom	1.70e+02
	919	16	5.4	139097	17	AC004591	*** SEQUENCING IN PROG	1.70e+02
	920	16	5.4	140554	28	AC000399	Genomic sequence from	1.70e+02
	921	16	5.4	140942	26	AC002525	Human PAC clone 257C22	1.70e+02
c	922	16	5.4	141752	17	HS971N18	Human DNA sequence ***	1.70e+02
	923	16	5.4	141779	26	AC002377	Human PAC clone DJ222H	1.70e+02
	924	16	5.4	142167	17	HS198P4	Human DNA sequence ***	1.70e+02
c	925	16	5.4	142228	17	AC004123	*** SEQUENCING IN PROG	1.70e+02
c	926	16	5.4	145414	26	HS78F24	Homo sapiens DNA seque	1.70e+02
	927	16	5.4	145721	17	HS112L6	Human DNA sequence ***	1.70e+02
c	928	16	5.4	148359	17	AC004811	*** SEQUENCING IN PROG	1.70e+02
	929	16	5.4	148600	26	AC002367	Homo sapiens Xp22 PAC	1.70e+02
c	930	16	5.4	148600	26	AC002367	Homo sapiens Xp22 PAC	1.70e+02
	931	16	5.4	151514	26	HS471M13	Human DNA sequence fro	1.70e+02
c	932	16	5.4	151514	26	HS471M13	Human DNA sequence fro	1.70e+02
	933	16	5.4	151770	26	HSAC002066	Human BAC clone RG030H	1.70e+02
c	934	16	5.4	152550	17	HSJ181N11	*** SEQUENCING IN PROG	1.70e+02
c	935	16	5.4	152798	26	AC004057	Homo sapiens chromosom	1.70e+02
c	936	16	5.4	152936	17	AC004229	*** SEQUENCING IN PROG	1.70e+02
c	937	16	5.4	153356	17	AC004995	*** SEQUENCING IN PROG	1.70e+02
	938	16	5.4	153617	17	HS117O3	Human DNA sequence ***	1.70e+02
c	939	16	5.4	153692	17	CEY26D4	Caenorhabditis elegans	1.70e+02
c	940	16	5.4	153892	17	HS1121J18	Human DNA sequence ***	1.70e+02
	941	16	5.4	154166	17	AC005479	*** SEQUENCING IN PROG	1.70e+02
c	942	16	5.4	154407	26	AF042090	Homo sapiens chromosom	1.70e+02
c	943	16	5.4	157897	17	HS171N11	Human DNA sequence ***	1.70e+02
	944	16	5.4	158886	26	AC004674	Homo sapiens Xp22 BAC	1.70e+02
c	945	16	5.4	159245	17	AC004819	*** SEQUENCING IN PROG	1.70e+02
	946	16	5.4	159424	26	AF064865	Homo sapiens chromosom	1.70e+02
	947	16	5.4	160850	17	AC004957	*** SEQUENCING IN PROG	1.70e+02
c	948	16	5.4	160899	17	AC004869	*** SEQUENCING IN PROG	1.70e+02
	949	16	5.4	161824	17	AC005061	*** SEQUENCING IN PROG	1.70e+02
c	950	16	5.4	162075	26	HS127D3	Homo sapiens DNA seque	1.70e+02
c	951	16	5.4	162750	17	AC003061	*** SEQUENCING IN PROG	1.70e+02
c	952	16	5.4	163899	17	AC005517	*** SEQUENCING IN PROG	1.70e+02
	953	16	5.4	164538	27	AC005774	Homo sapiens chromosom	1.70e+02
	954	16	5.4	165015	17	HS1158E12	Human DNA sequence ***	1.70e+02
	955	16	5.4	165507	17	AC004903	*** SEQUENCING IN PROG	1.70e+02
c	956	16	5.4	166126	26	AC004142	Homo sapiens BAC clone	1.70e+02
	957	16	5.4	166978	26	AC004098	Homo sapiens chromosom	1.70e+02
	958	16	5.4	166983	26	AC002541	Human BAC clone RG043K	1.70e+02
c	959	16	5.4	166983	26	AC002541	Human BAC clone RG043K	1.70e+02
c	960	16	5.4	167234	24	AC004708	CIC5B11.1 check: 4870	1.70e+02

```
c  961     16    5.4 171952 28  AC002324    Mus musculus chromosom  1.70e+02
   962     16    5.4 174311 18  AC004758    Drosophila melanogaste  1.70e+02
c  963     16    5.4 175270 17  HS652L8     Human DNA sequence ***  1.70e+02
   964     16    5.4 175488 17  AC004887    *** SEQUENCING IN PROG  1.70e+02
c  965     16    5.4 176781 17  AC002989    *** SEQUENCING IN PROG  1.70e+02
   966     16    5.4 177726 26  AC004083    Homo sapiens BAC clone  1.70e+02
   967     16    5.4 179245 26  HUAC002492  Human Chromosome 16 BA  1.70e+02
   968     16    5.4 179933 17  AC005048    *** SEQUENCING IN PROG  1.70e+02
c  969     16    5.4 180114 17  HS52D1      Human DNA sequence ***  1.70e+02
   970     16    5.4 180388 25  HUMRETBLAS  Human retinoblastoma s  1.70e+02
c  971     16    5.4 181995 17  HS596C15    Human DNA sequence ***  1.70e+02
   972     16    5.4 184982 17  HS327J16    Human DNA sequence ***  1.70e+02
c  973     16    5.4 185371 17  HS625H18    Human DNA sequence ***  1.70e+02
c  974     16    5.4 186510 26  HS451B15    Human DNA sequence fro  1.70e+02
   975     16    5.4 187647 17  CEY16B4     Caenorhabditis elegans  1.70e+02
   976     16    5.4 188205 27  AC003986    Homo sapiens BAC clone  1.70e+02
c  977     16    5.4 188362 27  AC005019    Homo sapiens BAC clone  1.70e+02
   978     16    5.4 189357 17  HS596H12    Human DNA sequence ***  1.70e+02
   979     16    5.4 191191 25  HSAC000119  Human BAC clone RG104I  1.70e+02
c  980     16    5.4 193374 17  AC005139    *** SEQUENCING IN PROG  1.70e+02
c  981     16    5.4 196405 17  AC005053    *** SEQUENCING IN PROG  1.70e+02
c  982     16    5.4 203789 17  AC004032    *** SEQUENCING IN PROG  1.70e+02
   983     16    5.4 205150 26  AC005274    Homo sapiens chromosom  1.70e+02
   984     16    5.4 206606 23  ATPCA1      Arabidopsis thaliana D  1.70e+02
c  985     16    5.4 207957 26  AC004470    Homo sapiens Xp22 BAC   1.70e+02
c  986     16    5.4 212107 17  AC000134    *** SEQUENCING IN PROG  1.70e+02
c  987     16    5.4 213863 27  AC003037    Homo sapiens Xp22 BAC   1.70e+02
c  988     16    5.4 220633 26  HUU91321    Human Chromosome 16 BA  1.70e+02
   989     16    5.4 224645 17  AC004153    *** SEQUENCING IN PROG  1.70e+02
c  990     16    5.4 224645 17  AC004153    *** SEQUENCING IN PROG  1.70e+02
   991     16    5.4 227215 17  HS245G19    Human DNA sequence ***  1.70e+02
   992     16    5.4 227538 28  AC002397    Mouse chromosome 6 BAC  1.70e+02
c  993     16    5.4 227538 28  AC002397    Mouse chromosome 6 BAC  1.70e+02
c  994     16    5.4 228500 17  HS212A2     Human DNA sequence ***  1.70e+02
c  995     16    5.4 234542 27  HUAC002041  Human Chromosome 16 BA  1.70e+02
c  996     16    5.4 235141 26  AC004615    Homo sapiens BAC clone  1.70e+02
c  997     16    5.4 238670 17  AC003059    *** SEQUENCING IN PROG  1.70e+02
   998     16    5.4 259894 26  HUAC002302  Homo sapiens Chromosom  1.70e+02
   999     16    5.4 275138 17  CEY54E5     Caenorhabditis elegans  1.70e+02
c 1000     16    5.4 322746 17  HS15D23     Human DNA sequence ***  1.70e+02
```

ALIGNMENTS

```
RESULT    1
LOCUS       HS024XG1      508 bp    DNA              STS       28-NOV-1994
DEFINITION  H. sapiens (D16S400) DNA segment containing (CA) repeat; clone
            AFM024xg1; single read, sequence tagged site.
ACCESSION   Z16438
NID         g22969
KEYWORDS    STS; CA repeat; dinucleotide repeat; GT repeat; microsatellite DNA;
```



```
            microsatellite marker; repeat polymorphism.
  SOURCE    human.
    ORGANISM  Homo sapiens
              Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
              Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE 1  (bases 1 to 508)
    AUTHORS   Weissenbach,J.
    TITLE     Direct Submission
    JOURNAL   Submitted (05-OCT-1992) Genethon, B.P. 60, 91002 Evry Cedex France.
              E-mail: Jean.Weissenbach@genethon.fr
  REFERENCE 2  (bases 1 to 508)
    AUTHORS   Weissenbach,J., Gyapay,G., Dib,C., Vignal,A., Morissette,J.,
              Millasseau,P., Vaysseix,G. and Lathrop,M.
    TITLE     A second-generation linkage map of the human genome
    JOURNAL   Nature 359 (6398), 794-801 (1992)
    MEDLINE   93063290
  REFERENCE 3  (bases 1 to 508)
    AUTHORS   Gyapay,G., Morissette,J., Vignal,A., Dib,C., Fizames,C.,
              Millasseau,P., Marc,S., Bernardi,G., Lathrop,M. and Weissenbach,J.
    TITLE     The 1993-94 Genethon human genetic linkage map
    JOURNAL   Nature Genet. 7 (2 Spec No), 246-339 (1994)
    MEDLINE   95004593
  COMMENT   cloning vector is M13mp18;
            full automatic.
  FEATURES             Location/Qualifiers
       source          1. .508
                       /organism="Homo sapiens"
                       /db_xref="taxon:9606"
                       /cell_line="CEPH 134702"
                       /clone_lib="genomic DNA"
                       /chromosome="16"
  BASE COUNT      139 a    110 c    122 g    132 t      5 others
  ORIGIN

    Query Match          8.4%;  Score 25;  DB 30;  Length 508;
    Best Local Similarity 100.0%;  Pred. No. 3.55e-05;
    Matches   25;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

  Db    182 TTTACATATATACATACATACACAC 206
            |||||||||||||||||||||||||
  Cp    199 TTTACATATATACATACATACACAC 175


  RESULT    2
  LOCUS       HUAC003977 130491 bp    DNA              PRI       13-JUN-1998
  DEFINITION  Homo sapiens Chromosome 16 BAC clone CIT987SK-A-542B9, complete
              sequence.
  ACCESSION   AC003977
  NID         g3219341
  KEYWORDS    HTG.
  SOURCE      human.
```



```
    ORGANISM  Homo sapiens
              Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
              Primates; Catarrhini; Hominidae; Homo.
  REFERENCE 1  (bases 1 to 130491)
    AUTHORS   Adams,M.D., Loftus,B.J., Zhou,L., Crosby,M., Fuhrmann,J.,
              Mason,T.M., Brandon,R., Kim,U.J., Kerlavage,A.R. and Venter,J.C.
    TITLE     Homo sapiens Chromosome 16 BAC clone CIT987SK-A-542B9
    JOURNAL   Unpublished
  REFERENCE 2  (bases 1 to 130491)
    AUTHORS   Adams,M.D. and Loftus,B.J.
    TITLE     Direct Submission
    JOURNAL   Submitted (05-JAN-1998) The Institute for Genomic Research, 9712
              Medical Center Dr, Rockville, MD 20850, USA, Email:
              bjloftus@tigr.org
  REFERENCE 3  (bases 1 to 130491)
    AUTHORS   Adams,M.D. and Loftus,B.J.
    TITLE     Direct Submission
    JOURNAL   Submitted (13-JUN-1998) The Institute for Genomic Research, 9712
              Medical Center Dr., Rockville, MD 20850, USA
  COMMENT   On Jun 13, 1998 this sequence version replaced gi:3172163.
            Address all correspondence to: Mark Adams The Institute for Genomic
            Research 9712 Medical Center Dr, Rockville, MD 20850, USA e-mail
            address: mdadams@tigr.org.   The orientation of the sequence is
            from SP6 end to T7 end. Genes were identified by a combination of
            five methods including: XGRAIL (available by anonymous ftp from
            arthur.epm.ornl.gov), Genefinder (Phil Green, University of
            Washington), Genscan  (Chris Burge,
            http://gnomic.stanford.Edu/~chris/GENSCANW.html)searches of the
            complete sequence against a peptide database,and the Human gene
            Index database at TIGR (http://www.tigr.org/tdb/hgi/hgi.html).
            Genes without pepetide homolgy having spliced EST hits are termed
            'Unknown gene product'. Genes encoding tRNAs are predicted by
            tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
  FEATURES             Location/Qualifiers
       source          1. .130491
                       /organism="Homo sapiens"
                       /db_xref="taxon:9606"
                       /chromosome="16"
                       /map="16q21"
                       /clone="A-542B9"
       STS             18500. .18701
                       /note="419, HS024XG1, Chr. 16, Homo sapiens"
                       /db_xref="dbSTS:Z16438"
       STS             38294. .38425
                       /note="32826, A002D07, Chr. -, Homo sapiens"
                       /db_xref="dbSTS:G19948"
       STS             39015. .39185
                       /note="42758, D11S1053, Chr. -, Homo sapiens"
                       /db_xref="dbSTS:G28812"
       STS             41559. .41690
                       /note="7617, STS1-cSRL-30a3-uA/cSRL-30a3-uZ, Chr. -, Homo
```



```
                       sapiens"
                       /db_xref="dbSTS:G02131"
       gene            78482. .78868
                       /gene="A-542B9.1"
                       /pseudo
  BASE COUNT    41566 a  23062 c  23483 g  42329 t     51 others
  ORIGIN

    Query Match          8.4%;  Score 25;  DB 26;  Length 130491;
    Best Local Similarity 100.0%;  Pred. No. 3.55e-05;
    Matches   25;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

  Db  18580 GTGTGTATGTATGTATATATGTAAA 18604
            |||||||||||||||||||||||||
  Qy    175 GTGTGTATGTATGTATATATGTAAA 199


  RESULT    3
  LOCUS       RLU92274      291 bp    DNA              STS       01-APR-1997
  DEFINITION  Rana luteiventris repeat region, (GT)11(ATGT)2(AT)39, clone RALU50,
              sequence tagged site.
  ACCESSION   U92274
  NID         g1916320
  KEYWORDS    STS.
  SOURCE      Rana luteiventris.
    ORGANISM  Rana luteiventris
              Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
              Vertebrata; Amphibia; Batrachia; Anura; Neobatrachia; Ranoidea;
              Ranidae; Rana.
  REFERENCE 1  (bases 1 to 291)
    AUTHORS   Call,D.R. and Hallett,J.G.
    TITLE     Population genetic structure for two anurans (Rana luteiventris and
              Hyla regilla)
    JOURNAL   Unpublished
  REFERENCE 2  (bases 1 to 291)
    AUTHORS   Call,D.R. and Hallett,J.G.
    TITLE     Direct Submission
    JOURNAL   Submitted (07-MAR-1997) Zoology, Washington State Unversity,
              Pullman, WA 99164-4236, USA
  FEATURES             Location/Qualifiers
       source          1. .291
                       /organism="Rana luteiventris"
                       /db_xref="taxon:58176"
                       /clone="RALU50"
       primer_bind     24. .43
                       /note="forward repeat primer"
       repeat_region   64. .172
                       /rpt_family="GT/AT"
                       /rpt_type=tandem
       primer_bind     complement(212. .233)
                       /note="reverse repeat primer"
```

```
BASE COUNT       93 a     36 c     60 g    102 t
ORIGIN

  Query Match           7.7%;  Score 23;  DB 30;  Length 291;
  Best Local Similarity 100.0%;  Pred. No. 1.49e-03;
  Matches   23;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      81 GTGTGTATGTATGTATATATGTA 103
           |||||||||||||||||||||||
Qy     175 GTGTGTATGTATGTATATATGTA 197


RESULT    4
LOCUS       MML41RIBO    1020 bp    RNA             ROD       12-FEB-1997
DEFINITION  M.musculus mRNA for L41 ribosomal like-protein.
ACCESSION   Y11009
NID         g1841447
KEYWORDS    B1 repeat; L41 ribosomal protein.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Mus.
REFERENCE   1  (bases 1 to 1020)
  AUTHORS   Kruecken,F., Santurlides,S., Schmitt-Wrede,H.P.,
            Markmann-Mulisch,U. and Wunderlich,F.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1020)
  AUTHORS   Kruecken,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JAN-1997) J. Kruecken, Division of Molecular
            Parasitology, Heinrich-Heine-University Dusseldorf,
            Universitaetsstr. 1, D-40225 Duesseldorf, FRG
COMMENT     Related sequence X82550.
FEATURES             Location/Qualifiers
     source          1. .1020
                     /organism="Mus musculus"
                     /strain="C57Bl/10"
                     /db_xref="taxon:10090"
                     /dev_stage="9-12 weeks old"
                     /tissue_type="spleen"
                     /clone_lib="lambda ZAPII"
                     /cell_type="T-cell"
     mRNA            1. .158
                     /gene="pM48"
                     /product="L41 ribosomal like-protein"
     gene            1. .158
                     /gene="pM48"
     repeat_region   884. .1020
                     /note="B1-repeat"
BASE COUNT      323 a    213 c    210 g    273 t      1 others
```



```
ORIGIN

  Query Match           7.7%;  Score 23;  DB 28;  Length 1020;
  Best Local Similarity 100.0%;  Pred. No. 1.49e-03;
  Matches   23;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     332 TACATATATACATACATACACAC 354
           |||||||||||||||||||||||
Cp     197 TACATATATACATACATACACAC 175


RESULT    5
LOCUS       AC000398    78616 bp    DNA             ROD       09-JUL-1997
DEFINITION  Genomic sequence from Mouse 11, complete sequence.
ACCESSION   AC000398
NID         g2133881
KEYWORDS    HTG.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae;
            Murinae; Mus.
REFERENCE   1  (bases 1 to 78616)
  AUTHORS   Hawkins,T.L., Reeve,M.P., Christoffersen,A., Birren,B.W.,
            Fasman,K.H. and Lander,E.S.
  TITLE     Genomic sequence from Mouse 11
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 78616)
  AUTHORS   Hawkins,T.L., Reeve,M.P., Christoffersen,A., Birren,B.W.,
            Fasman,K.H., Lander,E.S., McKernan,K., Munro,C., Richardson,P.,
            Barna,N., Brown,K., Cooke,P., Daly,M.J., Forrest,C., Fripp,W.J.,
            Gage,D., Geraigery,K., Hagos,B., Jacotot,L., Lane,M., MacKenzie,J.,
            Marquis,N., McDermott,J., Moloney,N., Morrow,J., Nachman,A.,
            Naylor,J., O'Connor,T., Peterson,K., Rollins,G., Spencer,J.,
            Stilwell,J., Stone,C., Strickland,C., Sydney,K., Wilmer,P. and
            Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-APR-1997) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 78616)
  AUTHORS   Hawkins,T.L., Reeve,M.P., Christoffersen,A., Birren,B.W.,
            Fasman,K.H., Lander,E.S., McKernan,K., Munro,C., Richardson,P.,
            Barna,N., Chang,A., Cooke,P., Daly,M.J., Forrest,C., Fripp,W.J.,
            Gage,D., Geraigery,K., Hagos,B., Jacotot,L., Lane,M., MacKenzie,J.,
            Marquis,N., McDermott,J., Moloney,N., Morrow,J., Nachman,A.,
            Naylor,J., O'Connor,T., Peterson,K., Rollins,G., Spencer,J.,
            Stilwell,J., Stone,C., Strickland,C., Sydney,K., Wilmer,P. and
            Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAY-1997) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
```



```
REFERENCE   4  (bases 1 to 78616)
  AUTHORS   Hawkins,T.L., Reeve,M.P., Christoffersen,A., Birren,B.W.,
            Fasman,K.H., Lander,E.S., McKernan,K., Munro,C., Richardson,P.,
            Barna,N., Chang,A., Cooke,P., Daly,M.J., Forrest,C., Fripp,W.J.,
            Gage,D., Geraigery,K., Hagos,B., Jacotot,L., Lane,M., MacKenzie,J.,
            Marquis,N., McDermott,J., Moloney,N., Morrow,J., Nachman,A.,
            Naylor,J., O'Connor,T., Peterson,K., Rollins,G., Spencer,J.,
            Stilwell,J., Stone,C., Strickland,C., Sydney,K., Wilmer,P. and
            Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUL-1997) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 29, 1997 this sequence version replaced gi:1929457.
            The Staden databases, finishing information, and all
            chromatographic files used in the assembly of this clone are
            available from our anonymous ftp site.

            All repeats were identified using RepeatMasker: Smit, A.F.A. &
            Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html.
FEATURES             Location/Qualifiers
     source          1. .78616
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="5195"
                     /clone_lib="unknown"
                     /chromosome="19"
     repeat_region   complement(2519. .2610)
                     /rpt_family="MIR"
     repeat_region   complement(3390. .3435)
                     /rpt_family="(GA)n"
     repeat_region   complement(3439. .3603)
                     /rpt_family="MMB2"
     repeat_region   3657. .3898
                     /rpt_family="URR1"
     repeat_region   3990. .4273
                     /rpt_family="ORR1A2"
     repeat_region   complement(4274. .4303)
                     /rpt_family="AT_rich"
     repeat_region   4307. .4470
                     /rpt_family="ORR1B-INT"
     repeat_region   4346. .4481
                     /rpt_family="MMB1"
     repeat_region   4496. .4546
                     /rpt_family="ORR1A2"
     repeat_region   complement(5394. .5456)
                     /rpt_family="AT_rich"
     repeat_region   complement(5499. .5563)
                     /rpt_family="(CA)n"
     repeat_region   complement(5728. .5795)
                     /rpt_family="ID5"
```



```
     repeat_region   complement(5969. .5999)
                     /rpt_family="AT_rich"
     repeat_region   6019. .6151
                     /rpt_family="MMB1"
     repeat_region   complement(6182. .6398)
                     /rpt_family="MMB2"
     repeat_region   complement(6423. .6500)
                     /rpt_family="(CA)n"
     repeat_region   complement(6540. .6630)
                     /rpt_family="MMB2"
     repeat_region   complement(6785. .6993)
                     /rpt_family="MMB2"
     repeat_region   complement(7027. .7071)
                     /rpt_family="(CAAA)n"
     repeat_region   complement(7144. .7253)
                     /rpt_family="MMB1"
     repeat_region   complement(8145. .8218)
                     /rpt_family="ID4"
     repeat_region   complement(12847. .13065)
                     /rpt_family="MMB2"
     repeat_region   13935. .14016
                     /rpt_family="(CA)n"
     repeat_region   complement(14409. .14493)
                     /rpt_family="MMPLAM"
     repeat_region   complement(14725. .14796)
                     /rpt_family="ID3"
     repeat_region   complement(16182. .16256)
                     /rpt_family="(CA)n"
     repeat_region   18483. .18570
                     /rpt_family="MIR"
     repeat_region   19634. .19675
                     /rpt_family="(GGA)n"
     repeat_region   20134. .20186
                     /rpt_family="MMSOR2"
     repeat_region   complement(20257. .20304)
                     /rpt_family="MMB1"
     repeat_region   21611. .21700
                     /rpt_family="(CA)n"
     repeat_region   complement(22675. .22734)
                     /rpt_family="ID3"
     repeat_region   complement(23214. .23238)
                     /rpt_family="(CA)n"
     repeat_region   23674. .23732
                     /rpt_family="(CA)n"
     repeat_region   complement(25206. .25316)
                     /rpt_family="MMB1"
     repeat_region   complement(25539. .25846)
                     /rpt_family="L1MB6"
     repeat_region   25847. .25896
                     /rpt_family="(CA)n"
     repeat_region   25847. .25967
```

```
                     /rpt_family="(CATA)n"
     repeat_region   complement(25985. .26100)
                     /rpt_family="MMB1"
     repeat_region   26166. .26241
                     /rpt_family="MLT1A2"
     repeat_region   26166. .26225
                     /rpt_family="MLT1B"
     repeat_region   26275. .26435
                     /rpt_family="MLT1A2"
     repeat_region   complement(26600. .26739)
                     /rpt_family="MMB1F"
     repeat_region   26935. .27134
                     /rpt_family="MMB2"
     repeat_region   complement(27341. .27455)
                     /rpt_family="MMB1"
     repeat_region   27464. .27678
                     /rpt_family="URR1"
     repeat_region   28200. .28305
                     /rpt_family="MMB1F"
     repeat_region   complement(28470. .28492)
                     /rpt_family="AT_rich"
     repeat_region   complement(30279. .30380)
                     /rpt_family="(CA)n"
     repeat_region   31505. .31578
                     /rpt_family="ID2"
     repeat_region   complement(32206. .32317)
                     /rpt_family="(TGGA)n"
     repeat_region   34199. .34236
                     /rpt_family="(GGGA)n"
     repeat_region   34424. .34460
                     /rpt_family="(CA)n"
     repeat_region   complement(34513. .34659)
                     /rpt_family="MMB1"
     repeat_region   complement(35036. .35113)
                     /rpt_family="L1MB4"
     repeat_region   complement(36615. .36664)
                     /rpt_family="(CA)n"
     repeat_region   complement(37333. .37468)
                     /rpt_family="MMB1F"
     repeat_region   37777. .37883
                     /rpt_family="(CATA)n"
     repeat_region   39393. .39462
                     /rpt_family="MMB2"
     repeat_region   complement(40710. .40829)
                     /rpt_family="MMB1"
     repeat_region   complement(41221. .41267)
                     /rpt_family="POLY_A"
     repeat_region   41292. .41420
                     /rpt_family="MMB1"
     repeat_region   complement(41507. .41590)
                     /rpt_family="(CATA)n"
```



```
     repeat_region   complement(41660. .41796)
                     /rpt_family="MMSOR2"
     repeat_region   complement(42459. .42532)
                     /rpt_family="ID5"
     repeat_region   45120. .45463
                     /rpt_family="MTF"
     repeat_region   complement(45489. .46001)
                     /rpt_family="L1MB3"
...
Note: remainder of annotations omitted.

  Query Match             7.7%;  Score 23;  DB 28;  Length 78616;
  Best Local Similarity 100.0%;  Pred. No. 1.49e-03;
  Matches    23;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  37794 TACATATATACATACATACACAC 37816
          |||||||||||||||||||||||
Cp    197 TACATATATACATACATACACAC 175


RESULT    6
LOCUS       AC005104   136569 bp    DNA             HTG       12-JUN-1998
DEFINITION  *** SEQUENCING IN PROGRESS *** Homo sapiens clone RG442F18; HTGS
            phase 1, 3 unordered pieces.
ACCESSION   AC005104
NID         g3212887
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 136569)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 136569)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-1998) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
  COMMENT   ***                                                        ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                        ***
            * This sequence is unfinished. It consists of 3 contigs for
            * which the order is not known; their order in this record is
            * arbitrary. In some cases, the exact lengths of the gaps
            * between the contigs are also unknown; these gaps are  presented
            * as runs of N as a convenience only. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
```



```
            *        1     1704: contig of 1704 bp in length
            *     1705     1721: gap of unknown length
            *     1722    45985: contig of 44264 bp in length
            *    45986    46002: gap of unknown length
            *    46003   136569: contig of 90567 bp in length.
FEATURES             Location/Qualifiers
     source          1. .136569
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RG442F18"
BASE COUNT    37428 a   30733 c   30529 g   37802 t     77 others
ORIGIN

  Query Match             7.7%;  Score 23;  DB 17;  Length 136569;
  Best Local Similarity 100.0%;  Pred. No. 1.49e-03;
  Matches    23;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   7510 GTGTGTATGTATGTATATATGTA 7532
          |||||||||||||||||||||||
Qy    175 GTGTGTATGTATGTATATATGTA 197


RESULT    7
LOCUS       AC004616   149752 bp    DNA             PRI       27-JUN-1998
DEFINITION  Homo sapiens Xp22 BAC GSHB-536K7 (Genome Systems Human BAC library)
            complete sequence.
ACCESSION   AC004616
NID         g3264544
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 149752)
  AUTHORS   Muzny,D., Arenson,A.D., Adams,C., Brundage,E., Bunac,C.,
            Carvelli,K., Chacko,J., Chen,J., Di,W., Ding,Y., Dugan,S.,
            Durbin,J., Forcum,J., Ganesh,R., Garcia,C., Goodman,M.,
            Gorrell,J.H., Haywood,M., Hernandez,J., Jackson,L., Jin,S.,
            Kampal,R., Karpathy,S., Kovar,C., Leal,B., Li,Y., Lichtarge,O.,
            Liu,W., Logan,O., Lu,J., Ly,T., Martinez,C., Oswal,G., Perez,L.,
            Rashid,N.D., Rowland,K., Savage,L., Scherer,S.E., Shen,H.,
            Simon,M., Stovall,K., Timms,K.M., Todd,J., Vo,Q., Williamson,A.,
            Worley,K.C., Yu,W., Chinault,C., Nelson,D. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 149752)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-APR-1998) Molecular and Human Genetics, Baylor
            College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 149752)
```



```
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JUN-1998) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
  COMMENT   On Jun 27, 1998 this sequence version replaced gi:3242669.
            Sequencing is completed to a minimum standard of double strand
            coverage with a minimum of 2 clones and 2 reads with no ambiguities
            or 2 chemistries with a minimum of 2 clones and 3 reads with no
            ambiguities. If the sequence quality does not meet this standard,
            it will be indicated in the annotation.

            The repeat regions shown were identified using RepeatMasker by
            Adrian Smit.

            Sequence similarities were identified using Powerblast by Jinghui
            Zhang.

            Exon/Intron boundaries of identified genes were chosen if there
            were canonical splice junctions that maintained sequence continuity
            across the splice junctions.
FEATURES             Location/Qualifiers
     source          1. .149752
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="GSHB-536K7"
                     /chromosome="X"
                     /clone_lib="Genome Systems Human BAC library"
                     /map="Xp22"
     repeat_region   complement(444. .897)
                     /rpt_family="MLT1C"
     repeat_region   2598. .2956
                     /rpt_family="THE1B"
     repeat_region   2962. .3309
                     /rpt_family="MLT1A2"
     repeat_region   3314. .3682
                     /rpt_family="MLT1C"
     repeat_region   complement(3698. .4143)
                     /rpt_family="L1PB1"
     STS             3988. .4183
                     /standard_name="AFMa082zh1, Chr. X, Homo sapiens"
                     /db_xref="dbSTS:47297"
     repeat_region   4145. .4196
                     /rpt_family="MLT1C"
     repeat_region   4644. .5288
                     /rpt_family="MER82"
     repeat_region   complement(5938. .6236)
                     /rpt_family="AluSc"
     repeat_region   complement(6237. .6528)
                     /rpt_family="AluSx"
     repeat_region   6554. .6916
```

```
                     /rpt_family="MLT1A1"
     repeat_region   7618. .7918
                     /rpt_family="AluY"
     repeat_region   complement(8233. .8388)
                     /rpt_family="L1MB3"
     repeat_region   complement(8415. .8779)
                     /rpt_family="L1PB1"
     repeat_region   8797. .8987
                     /rpt_family="MER82"
     repeat_region   complement(9003. .9337)
                     /rpt_family="AluJo"
     repeat_region   9339. .9425
                     /rpt_family="MER82"
     repeat_region   complement(9433. .9483)
                     /rpt_family="(CA)n"
     repeat_region   complement(10289. .10407)
                     /rpt_family="L1MB6"
     repeat_region   complement(10408. .10440)
                     /rpt_family="U2"
     repeat_region   complement(10446. .10836)
                     /rpt_family="L1MB6"
     repeat_region   11062. .11395
                     /rpt_family="L2"
     repeat_region   complement(11850. .11906)
                     /rpt_family="(GAA)n"
     repeat_region   complement(12103. .12308)
                     /rpt_family="AluJo"
     repeat_region   complement(13892. .13921)
                     /rpt_family="(CA)n"
     repeat_region   13920. .13955
                     /rpt_family="(CA)n"
     gene            complement(join(14144. .14188,25971. .26469))
                     /note="AA662138 na66a02.s1 NCI_CGAP_Pr22 Homo-sapiens cDNA
                     clone IMAGE:1188554 3'"
                     /gene="AA662138"
     repeat_region   complement(14209. .14506)
                     /rpt_family="AluJo"
     repeat_region   complement(14532. .14892)
                     /rpt_family="L1PA16"
     repeat_region   14893. .15064
                     /rpt_family="THE1B"
     repeat_region   complement(15065. .16922)
                     /rpt_family="L1PA6"
     repeat_region   16922. .17158
                     /rpt_family="L1PA6"
     repeat_region   17159. .17366
                     /rpt_family="THE1B"
     repeat_region   complement(17372. .17624)
                     /rpt_family="L1M1"
     repeat_region   18203. .18275
                     /rpt_family="MLT1A1"
```



```
     repeat_region   18276. .18668
                     /rpt_family="MLT1A2"
     repeat_region   18669. .18973
                     /rpt_family="MLT1A1"
     repeat_region   19150. .19337
                     /rpt_family="L1MA9"
     repeat_region   complement(19339. .19726)
                     /rpt_family="MSTA"
     repeat_region   19727. .19933
                     /rpt_family="L1MA7"
     repeat_region   complement(20409. .20557)
                     /rpt_family="L1MA4"
     misc_feature    complement(20659. .20832)
                     /note="Region: AA348446 EST55087 Hippocampus II Homo
                     sapiens cDNA 3' end"
     repeat_region   complement(21310. .21353)
                     /rpt_family="(TAA)n"
     repeat_region   complement(21354. .21651)
                     /rpt_family="AluJo"
     repeat_region   complement(21958. .22134)
                     /rpt_family="MER5A"
     repeat_region   complement(22309. .22574)
                     /rpt_family="LTR8"
     repeat_region   complement(24234. .24351)
                     /rpt_family="L1MC1"
     repeat_region   24359. .24417
                     /rpt_family="L1MC1"
     repeat_region   complement(24487. .24585)
                     /rpt_family="(TA)n"
     repeat_region   24587. .24711
                     /rpt_family="(TA)n"
     repeat_region   24771. .25112
                     /rpt_family="MLT1A1"
     repeat_region   26179. .26233
                     /rpt_family="MIR"
     repeat_region   complement(26521. .26543)
                     /rpt_family="AT_rich"
     repeat_region   complement(26556. .27382)
                     /rpt_family="L1PA13"
     repeat_region   complement(27391. .27691)
                     /rpt_family="AluSx"
     repeat_region   complement(27770. .27797)
                     /rpt_family="AT_rich"
     repeat_region   29768. .30067
                     /rpt_family="AluJo"
     repeat_region   complement(30889. .31275)
                     /rpt_family="L1PA14"
     repeat_region   31275. .31472
                     /rpt_family="L1PA14"
     repeat_region   32159. .32231
                     /rpt_family="MER63A"
```



```
     repeat_region   33102. .33414
                     /rpt_family="THE1C"
     repeat_region   complement(33511. .33682)
                     /rpt_family="(TA)n"
     repeat_region   complement(33693. .33980)
                     /rpt_family="AluJo"
     repeat_region   34306. .34568
                     /rpt_family="L1MA5A"
     repeat_region   34635. .34743
                     /rpt_family="(GGA)n"
     repeat_region   34991. .35054
                     /rpt_family="MER5B"
     repeat_region   35104. .35191
                     /rpt_family="(GGA)n"
     repeat_region   35269. .35397
                     /rpt_family="L1ME"
     repeat_region   35678. .35754
                     /rpt_family="AT_rich"
     repeat_region   complement(36080. .36201)
...
Note: remainder of annotations omitted.

  Query Match             7.7%;  Score 23;  DB 26;  Length 149752;
  Best Local Similarity 100.0%;  Pred. No. 1.49e-03;
  Matches    23;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  89641 GTGTGTATGTATGTATATATGTA 89663
          |||||||||||||||||||||||
Qy    175 GTGTGTATGTATGTATATATGTA 197


RESULT    8
LOCUS       HS1191B2   173139 bp    DNA              HTG       23-SEP-1998
DEFINITION  Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone
            1191B2; HTGS phase 1.
ACCESSION   AL022237
NID         g3646039
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 173139)
  AUTHORS   Cobley,V.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-SEP-1998) Wellcome Trust Genome Campus, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquires:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 24, 1998 this sequence version replaced gi:3559850.
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
```



```
            and the release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc. Order of segments is not known; 800 n's separate
            segments. Unfinished: bK1191B2  Contig_ID: 01414  acc=AL022237
            Length: 1890 bp Unfinished: bK1191B2  Contig_ID: 07996
            acc=AL022237  Length: 65362 bp Unfinished: bK1191B2  Contig_ID:
            03592  acc=AL022237  Length: 1596 bp Unfinished: bK1191B2
            Contig_ID: 02715  acc=AL022237  Length: 1023 bp Unfinished:
            bK1191B2  Contig_ID: 01110  acc=AL022237  Length: 1208 bp
            Unfinished: bK1191B2  Contig_ID: 03390  acc=AL022237  Length: 1814
            bp Unfinished: bK1191B2  Contig_ID: 04346  acc=AL022237  Length:
            1083 bp Unfinished: bK1191B2  Contig_ID: 03609  acc=AL022237
            Length: 1027 bp Unfinished: bK1191B2  Contig_ID: 07343
            acc=AL022237  Length: 1647 bp Unfinished: bK1191B2  Contig_ID:
            07653  acc=AL022237  Length: 1555 bp Unfinished: bK1191B2
            Contig_ID: 03384  acc=AL022237  Length: 49151 bp Unfinished:
            bK1191B2  Contig_ID: 02475  acc=AL022237  Length: 36983 bp.
            ***                                                       ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                       ***
            * This sequence is unfinished. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
FEATURES              Location/Qualifiers
     source           1. .173139
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="22"
                      /clone="1191B2"
BASE COUNT    37674 a  44531 c  44281 g  37849 t   8804 others
ORIGIN

  Query Match             7.7%;  Score 23;  DB 17;  Length 173139;
  Best Local Similarity 100.0%;  Pred. No. 1.49e-03;
  Matches    23;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  20881 TACATATATACATACATACACAC 20903
          |||||||||||||||||||||||
Cp    197 TACATATATACATACATACACAC 175


RESULT    9
LOCUS       HS526I14   199553 bp    DNA              HTG       23-SEP-1998
DEFINITION  Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone
            526I14; HTGS phase 1.
ACCESSION   Z82214
NID         g3649728
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
```

```
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 199553)
  AUTHORS   Bridgeman,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-SEP-1998) Wellcome Trust Genome Campus, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquires:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 25, 1998 this sequence version replaced gi:3355576.
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and the release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc. Order of segments is not known; 800 n's separate
            segments. Unfinished: dJ526I14  Contig_ID: 09662  acc=Z82214
            Length: 65546 bp Unfinished: dJ526I14  Contig_ID: 06705  acc=Z82214
             Length: 1098 bp Unfinished: dJ526I14  Contig_ID: 05695  acc=Z82214
             Length: 1951 bp Unfinished: dJ526I14  Contig_ID: 04941  acc=Z82214
             Length: 1287 bp Unfinished: dJ526I14  Contig_ID: 08724  acc=Z82214
             Length: 1229 bp Unfinished: dJ526I14  Contig_ID: 05206  acc=Z82214
             Length: 124442 bp.
            ***                                                     ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                     ***
            * This sequence is unfinished. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
FEATURES              Location/Qualifiers
     source           1. .199553
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="22"
                      /clone="526I14"
BASE COUNT    44012 a  53470 c  54015 g  44051 t   4005 others
ORIGIN

  Query Match             7.7%;  Score 23;  DB 17;  Length 199553;
  Best Local Similarity 100.0%;  Pred. No. 1.49e-03;
  Matches    23;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  47149 GTGTGTATGTATGTATATATGTA 47171
          |||||||||||||||||||||||
Qy    175 GTGTGTATGTATGTATATATGTA 197


RESULT  10
LOCUS       HS67M12    103268 bp    DNA              PRI       07-JAN-1998
DEFINITION  Human DNA sequence from PAC 67M12 on chromosome 6p22.1-22.3.
            Contains ESTs, STSs and CA and TAA repeat polymorphisms.
ACCESSION   AL008732
```



```
NID         g2760545
KEYWORDS    6p22.1-22.3; CA repeat polymorphism; TAA repeat polymorphism.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 103268)
  AUTHORS   Phillips,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JAN-1998) sanger.ac.uk/HGP/Chr6/) Sanger Centre,
            Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquires:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jan 9, 1998 this sequence version replaced gi:2598469.
            IMPORTANT: This sequence is the entire insert of clone 67M12.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variations annotated may not be found in the sequence submission
            corresponding to the overlapping clone as we submit sequences with
            only a small overlap as described above.
            This sequence was generated from part of bacterial clone contigs of
            human chromosome 6, constructed by the Sanger Centre chromosome 6
            mapping group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6/
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            The true left end of clone 67M12 is at 1 in this sequence. The true
            right end of clone 67M12 is at 103268.
            67M12 is from the library RPCI1 constructed at the Roswell Park
            Cancer Institute by the group of Pieter de Jong.
            For further details see http://bacpac.med.buffalo.edu/.
FEATURES              Location/Qualifiers
     source           1. .103268
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="6"
                      /map="6p22.1-22.3"
                      /clone="67M12"
                      /clone_lib="RPCI1"
     repeat_region    1. .69
                      /note="AluY repeat: matches 233. .301 of consensus;
                      incomplete repeat"
     repeat_region    670. .980
                      /note="AluJb repeat: matches 2. .297 of consensus"
     repeat_region    3265. .3297
                      /note="16 copies of 2 mer 88 % conserved"
     repeat_region    3809. .3952
```



```
                      /note="MIR repeat: matches 166. .22 of consensus"
     repeat_region    4099. .4254
                      /note="MIR repeat: matches 216. .53 of consensus"
     prim_transcript  4553. .4999
                      /note="match: ESTs R18779 H14763 H15082"
     repeat_region    5442. .6219
                      /note="L1MC3 repeat: matches 2482. .1676 of consensus"
     prim_transcript  6225. .6622
                      /note="match: ESTs T75290 R15166 R56507"
     repeat_region    6631. .6718
                      /note="2 copies of 44 mer 88 % conserved"
     prim_transcript  6809. .7650
                      /note="match: ESTs W17044 W16822 N90929 N90913 R41585
                      R56114 R49660 F10641"
     repeat_region    7531. .7696
                      /note="MER21B repeat: matches 430. .263 of consensus"
     repeat_region    7607. .7719
                      /note="MER21A repeat: matches 408. .288 of consensus"
     repeat_region    8584. .8674
                      /note="MIR repeat: matches 235. .144 of consensus"
     repeat_region    9208. .9293
                      /note="AluSp repeat: matches 218. .303 of consensus;
                      incomplete repeat"
     repeat_region    9700. .10051
                      /note="THE1C repeat: matches 371. .1 of consensus"
     repeat_region    10442. .10908
                      /note="L1 repeat: matches 4912. .5378 of consensus"
     repeat_region    10774. .11616
                      /note="L1PB3 repeat: matches 1. .897 of consensus"
     repeat_region    11973. .13775
                      /note="L1 repeat: matches 3584. .5390 of consensus"
     repeat_region    13630. .14511
                      /note="L1PA15 repeat: matches 1. .904 of consensus"
     repeat_region    15951. .16253
                      /note="AluJb repeat: matches 7. .302 of consensus"
     repeat_region    16347. .16648
                      /note="AluSx repeat: matches 302. .1 of consensus"
     repeat_region    16898. .17207
                      /note="AluSx repeat: matches 2. .300 of consensus"
     repeat_region    17420. .17466
                      /note="MIR repeat: matches 94. .140 of consensus"
     repeat_region    17898. .17925
                      /note="AluY repeat: matches 232. .259 of consensus;
                      incomplete repeat"
     repeat_region    18724. .18776
                      /note="MER34 repeat: matches 532. .482 of consensus"
     repeat_region    18947. .19021
                      /note="MER4B repeat: matches 195. .271 of consensus"
     repeat_region    19130. .19430
                      /note="AluSg repeat: matches 1. .300 of consensus"
     repeat_region    19448. .19746
```



```
                      /note="AluY repeat: matches 1. .301 of consensus"
     repeat_region    19775. .19893
                      /note="AluJo repeat: matches 4. .122 of consensus;
                      incomplete repeat"
     repeat_region    19924. .19963
                      /note="20 copies of 2 mer 100 % conserved"
     repeat_region    19969. .20255
                      /note="AluSx repeat: matches 284. .1 of consensus;
                      incomplete repeat"
     repeat_region    20272. .20444
                      /note="AluJb repeat: matches 129. .301 of consensus;
                      incomplete repeat"
     repeat_region    20446. .20479
                      /note="17 copies of 2 mer 88 % conserved"
     repeat_region    20489. .20585
                      /note="MER34 repeat: matches 164. .55 of consensus"
     repeat_region    22087. .22165
                      /note="MIR2 repeat: matches 63. .141 of consensus"
     repeat_region    26651. .26956
                      /note="AluJb repeat: matches 1. .302 of consensus"
     repeat_region    31577. .31645
                      /note="MIR repeat: matches 176. .107 of consensus"
     repeat_region    32445. .32732
                      /note="AluSx repeat: matches 1. .302 of consensus"
     repeat_region    32741. .32784
                      /note="22 copies of 2 mer 98 % conserved"
     repeat_region    32818. .32907
                      /note="45 copies of 2 mer 81 % conserved"
     misc_feature     33140. .33516
                      /note="match: STS G17038"
     repeat_region    35464. .35763
                      /note="AluSx repeat: matches 302. .1 of consensus"
     repeat_region    36160. .36306
                      /note="MIR repeat: matches 2. .157 of consensus"
     repeat_region    37396. .37697
                      /note="AluSx repeat: matches 1. .302 of consensus"
     repeat_region    37729. .37905
                      /note="L1 repeat: matches 5195. .5390 of consensus"
     repeat_region    37781. .38469
                      /note="L1ME2 repeat: matches 23. .774 of consensus"
     repeat_region    38527. .38787
                      /note="AluJb repeat: matches 1. .301 of consensus"
     repeat_region    42100. .42225
                      /note="MIR2 repeat: matches 145. .12 of consensus"
     repeat_region    42617. .43115
                      /note="L1ME3 repeat: matches 402. .910 of consensus"
     repeat_region    44174. .44493
                      /note="AluJo repeat: matches 301. .2 of consensus"
     repeat_region    45485. .45727
                      /note="MER7B repeat: matches 1. .250 of consensus"
     repeat_region    45750. .45859
```

```
                     /note="MER7B repeat: matches 1094. .1205 of consensus"
     repeat_region   45885. .46247
                     /note="THE1B repeat: matches 1. .363 of consensus"
     repeat_region   47549. .47679
                     /note="PLAM_C repeat: matches 132. .1 of consensus"
     repeat_region   48096. .48174
                     /note="MER7B repeat: matches 96. .188 of consensus"
     repeat_region   48129. .48264
                     /note="MER7B repeat: matches 208. .73 of consensus"
     repeat_region   48869. .48980
                     /note="MER7B repeat: matches 1205. .1094 of consensus"
     repeat_region   49054. .49273
                     /note="MER7B repeat: matches 235. .8 of consensus"
     repeat_region   49398. .49464
                     /note="L1PA16 repeat: matches 874. .808 of consensus"
     repeat_region   49678. .50617
                     /note="L1 repeat: matches 4366. .5344 of consensus"
     repeat_region   50512. .51417
                     /note="L1ME1 repeat: matches 1. .910 of consensus"
     repeat_region   52956. .53167
                     /note="MER42c repeat: matches 1537. .1330 of consensus"
     repeat_region   53190. .53884
                     /note="L1ME2 repeat: matches 892. .133 of consensus"
     repeat_region   53560. .53985
                     /note="L1PA9 repeat: matches 475. .2 of consensus"
     repeat_region   53879. .54018
                     /note="L1 repeat: matches 5366. .5210 of consensus"
     repeat_region   54079. .54380
                     /note="AluJb repeat: matches 302. .1 of consensus"
     repeat_region   54443. .55065
                     /note="L1 repeat: matches 5052. .4416 of consensus"
     repeat_region   55063. .55135
                     /note="7SK repeat: matches 193. .121 of consensus"
     repeat_region   55174. .55300
                     /note="7SK repeat: matches 125. .1 of consensus"
     repeat_region   55324. .55944
                     /note="L1 repeat: matches 4411. .3787 of consensus"
     repeat_region   56268. .56304
                     /note="MIR repeat: matches 114. .78 of consensus"
     repeat_region   58347. .58556
                     /note="MIR repeat: matches 7. .223 of consensus"
     repeat_region   59312. .59505
                     /note="AluJb repeat: matches 302. .115 of consensus;
...
Note: remainder of annotations omitted.

  Query Match             7.4%;  Score 22;  DB 26;  Length 103268;
  Best Local Similarity 100.0%;  Pred. No. 9.15e-03;
  Matches    22;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  17996 ACATATATACATACATACACAC 18017
```



```
              ||||||||||||||||||||||
Cp    196 ACATATATACATACATACACAC 175


RESULT   11
LOCUS       AC002324   171952 bp    DNA             ROD       30-MAY-1998
DEFINITION  Mus musculus chromosome 11, clone 475_H_14, complete sequence.
ACCESSION   AC002324
NID         g3169207
KEYWORDS    HTG.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 171952)
  AUTHORS   Birren,B., Fasman,K., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus chromosome 11, clone 475_H_14
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 171952)
  AUTHORS   Hawkins,T.L., Reeve,M.P., Christoffersen,A., Birren,B.W.,
            Fasman,K.H., Lander,E.S., McKernan,K., Munro,C., Richardson,P.,
            Barna,N., Chang,A., Cooke,P., Daly,M.J., Devon,K., Dewar,K.,
            Forrest,C., Gage,D., Geraigery,K., Guitau,G., Hagos,B., Huang,J.,
            Jacotot,L., Lane,M., Lee,K., MacKenzie,J., Marquis,N.,
            McDermott,J., Molla,M., Moloney,N., Morrow,J., Nachman,A.,
            Naylor,J., O'Connor,T., Olotu,A., Peterson,K., Rollins,G.,
            Spencer,J., Stilwell,J., Stone,C., Strickland,C., Sydney,K.,
            Traish,A., Wilmer,F., Zemtseva,I. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JUL-1997) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 171952)
  AUTHORS   Birren,B., Fasman,K., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boatin,C.,
            Boutwell,C., Brown,A., Castle,A., Cerny,J., Cooke,P., Depayre,E.,
            Devon,K., Dewar,K., Donelan,L., Etemadi,S., Ferreira,P.,
            FitzHugh,W., Forrest,C., Funke,R., Gage,D., Gardyna,S.,
            Gensheimer,S., Geraigery,K., Gilmartin,T., Grant,G., Hagos,B.,
            Harris,K., Horton,L., Howland,J.C., Hui,L., Jacotot,L., Kann,L.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Nachman,A., Nahf,R., Naylor,J., O'Connor,T., Pavlin,B.,
            Peterson,K., Riley,R., Roberts,D., Rossello,R., Roy,A., Shyam,R.,
            Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C.,
            Strickland,C., Subramanian,A., Torruella-Miller,I., Vassiliev,H.,
            Vo,A., Wagner,A., Wang,B., Wheeler,J., Wu,Y., Ye,W.J., Zhao,J. and
            Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAY-1998) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 30, 1998 this sequence version replaced gi:3142319.
```



```
            All repeats were identified using RepeatMasker: Smit, A.F.A. &
            Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html.
FEATURES            Location/Qualifiers
     source         1. .171952
                    /organism="Mus musculus"
                    /db_xref="taxon:10090"
                    /clone="475_H_14"
                    /clone_lib="Research Genetics/Cal Tech CITB-HSP-C (plates
                    195-384)"
                    /map="11"
                    /chromosome="11"
     repeat_region  371. .456
                    /rpt_family="PB1D10"
     repeat_region  complement(712. .827)
                    /rpt_family="PB1D10"
     repeat_region  complement(990. .1029)
                    /rpt_family="PB1D9"
     repeat_region  complement(1309. .1451)
                    /rpt_family="B1_MM"
     repeat_region  complement(1882. .2071)
                    /rpt_family="URR1B"
     repeat_region  2523. .2563
                    /rpt_family="B3A"
     repeat_region  2596. .2717
                    /rpt_family="B3"
     repeat_region  complement(4356. .4383)
                    /rpt_family="POLY_A"
     repeat_region  5275. .5494
                    /rpt_family="B3"
     repeat_region  5501. .5674
                    /rpt_family="B3"
     repeat_region  6287. .6402
                    /rpt_family="PB1D10"
     repeat_region  complement(7238. .7371)
                    /rpt_family="L2"
     repeat_region  7659. .7737
                    /rpt_family="B1-F"
     repeat_region  complement(8284. .8429)
                    /rpt_family="B1_MM"
     repeat_region  8576. .8652
                    /rpt_family="L2"
     repeat_region  complement(11209. .11411)
                    /rpt_family="B3"
     repeat_region  complement(11436. .11605)
                    /rpt_family="B3"
     repeat_region  complement(11634. .11836)
                    /rpt_family="B2"
     repeat_region  complement(11861. .12013)
                    /rpt_family="B2"
     repeat_region  complement(12262. .12303)
```



```
                    /rpt_family="PB1D9"
     repeat_region  12306. .12347
                    /rpt_family="L2"
     repeat_region  complement(12573. .12596)
                    /rpt_family="(CCAA)n"
     repeat_region  complement(12639. .12762)
                    /rpt_family="B4"
     repeat_region  complement(12807. .12872)
                    /rpt_family="PB1D7"
     repeat_region  12920. .13046
                    /rpt_family="B1-F"
     repeat_region  complement(13101. .13294)
                    /rpt_family="B2"
     repeat_region  13310. .13532
                    /rpt_family="MER2"
     repeat_region  complement(13534. .13650)
                    /rpt_family="B1_MM"
     repeat_region  complement(13597. .13670)
                    /rpt_family="B4A"
     repeat_region  complement(13684. .13829)
                    /rpt_family="B1-F"
     repeat_region  complement(13804. .13837)
                    /rpt_family="MT2B"
     repeat_region  13897. .13922
                    /rpt_family="MT2B"
     unsure         complement(13899. .14050)
                    /note="Single-stranded terminator coverage."
     repeat_region  13923. .14070
                    /rpt_family="B1_MM"
     repeat_region  complement(14207. .14351)
                    /rpt_family="B1_MM"
     repeat_region  complement(14958. .15399)
                    /rpt_family="MTC"
     repeat_region  16021. .16143
                    /rpt_family="PB1D10"
     repeat_region  16153. .16236
                    /rpt_family="PB1D7"
     repeat_region  complement(17105. .17224)
                    /rpt_family="(CA)n"
     repeat_region  complement(17227. .17290)
                    /rpt_family="(CA)n"
     repeat_region  complement(18858. .18900)
                    /rpt_family="AT_rich"
     repeat_region  complement(18901. .18907)
                    /rpt_family="B1-F"
     repeat_region  complement(18908. .18966)
                    /rpt_family="B4"
     repeat_region  complement(18967. .19010)
                    /rpt_family="B1-F"
     repeat_region  complement(19543. .19690)
                    /rpt_family="RSINE1"
```

```
     repeat_region   complement(19713. .19815)
                     /rpt_family="PB1D10"
     repeat_region   19950. .20032
                     /rpt_family="PB1D9"
     repeat_region   21057. .21132
                     /rpt_family="PB1D7"
     repeat_region   21151. .21201
                     /rpt_family="ID4"
     repeat_region   21438. .21541
                     /rpt_family="B1-F"
     repeat_region   complement(24457. .24557)
                     /rpt_family="PB1D9"
     repeat_region   25092. .25226
                     /rpt_family="B1_MM"
     repeat_region   complement(25295. .25403)
                     /rpt_family="PB1D7"
     repeat_region   complement(25668. .25852)
                     /rpt_family="RSINE1"
     repeat_region   complement(25985. .26117)
                     /rpt_family="B1_MM"
     repeat_region   26126. .26176
                     /rpt_family="B3A"
     repeat_region   complement(26292. .26359)
                     /rpt_family="ID4"
     repeat_region   complement(28326. .28816)
                     /rpt_family="L1_MM"
     repeat_region   29393. .29599
                     /rpt_family="B2"
     repeat_region   29603. .29823
                     /rpt_family="B3"
     repeat_region   29996. .30193
                     /rpt_family="B3"
     repeat_region   complement(30412. .30563)
                     /rpt_family="B1_MM"
     repeat_region   complement(31242. .31293)
                     /rpt_family="PB1D7"
     repeat_region   31515. .31647
                     /rpt_family="B1_MM"
     repeat_region   32490. .32579
                     /rpt_family="PB1D10"
     unsure          33376. .33409
                     /note="Single-stranded terminator coverage."
     repeat_region   33580. .33671
                     /rpt_family="GC_rich"
     repeat_region   complement(35248. .35318)
                     /rpt_family="(TA)n"
     repeat_region   35664. .35809
                     /rpt_family="RSINE1"
     repeat_region   36199. .36304
                     /rpt_family="PB1D9"
     repeat_region   complement(36426. .36535)
```



```
                     /rpt_family="B2"
...
Note: remainder of annotations omitted.

  Query Match             7.4%;  Score 22;  DB 28;  Length 171952;
  Best Local Similarity 100.0%;  Pred. No. 9.15e-03;
  Matches    22;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  59308 TGTGTATGTATGTATATATGTA 59329
          ||||||||||||||||||||||
Qy    176 TGTGTATGTATGTATATATGTA 197


RESULT   12
LOCUS       HSA108WE1     312 bp    DNA             STS       22-MAR-1996
DEFINITION  H.sapiens (D12S1585) DNA segment containing (CA) repeat; clone
            AFMa108we1; single read.
ACCESSION   Z52055
NID         g1233355
KEYWORDS    CA repeat; dinucleotide repeat; GT repeat; microsatellite DNA;
            microsatellite marker; repeat polymorphism; STS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 312)
  AUTHORS   Weissenbach,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-SEP-1995) Genethon, B.P. 60, 91002 Evry Cedex France.
            E-mail: Jean.Weissenbach@genethon.fr
REFERENCE   2  (bases 1 to 312)
  AUTHORS   Dib,C., Faure,S., Fizames,C., Samson,D., Drouot,N., Vignal,A.,
            Millasseau,P., Marc,S., Hazan,J., Seboun,E., Lathrop,M., Gyapay,G.,
            Morissette,J. and Weissenbach,J.
  TITLE     A comprehensive genetic map of the human genome based on 5,264
            microsatellites
  JOURNAL   Nature 380 (6570), 152-154 (1996)
  MEDLINE   96176476
COMMENT     full automatic.
FEATURES             Location/Qualifiers
     source          1. .312
                     /organism="Homo sapiens"
                     /note="cloning vector is M13mp18"
                     /db_xref="taxon:9606"
                     /cell_line="CEPH 134702"
                     /clone_lib="genomic DNA"
                     /chromosome="12"
BASE COUNT      107 a     74 c     51 g     73 t      7 others
ORIGIN

  Query Match             7.0%;  Score 21;  DB 30;  Length 312;
```



```
  Best Local Similarity  95.5%;  Pred. No. 5.37e-02;
  Matches    21;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db    206 ACATATANACATACATACACAC 227
          ||||||| ||||||||||||||||
Cp    196 ACATATATACATACATACACAC 175


RESULT   13
LOCUS       HS22QT017     421 bp    DNA             PRI       06-JUN-1997
DEFINITION  H.sapiens telomeric DNA sequence, clone 22QTEL017, read
            22QTEL000017.seq.
ACCESSION   Z96503
NID         g2181578
KEYWORDS    genomic fragment; subtelomeric DNA.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 421)
  AUTHORS   Flint,J. and Bates,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUN-1997) Jonathan Flint, Institute of Molecular
            Medicine, John Radcliffe Hospital Oxford, OX3 9DS, UK E-mail
            contact: jf@worf.molbiol.ox.ac.uk
REFERENCE   2  (bases 1 to 421)
  AUTHORS   Ning,Y. and Flint,J.
  TITLE     A complete set of human telomeric probes and their clinical
            application application
  JOURNAL   Nat. Genet. 14, 86-89 (1996)
  REMARK    National Institutes of Health, Institute of Molecular Medicine.
COMMENT     Vector: m13mp18.
FEATURES             Location/Qualifiers
     source          1. .421
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /dev_stage="adult"
                     /sex="male"
                     /tissue_type="blood"
                     /clone_lib="TEL"
                     /clone="22QTEL017"
BASE COUNT      118 a    103 c     73 g    124 t      3 others
ORIGIN

  Query Match             7.0%;  Score 21;  DB 26;  Length 421;
  Best Local Similarity 100.0%;  Pred. No. 5.37e-02;
  Matches    21;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    354 TACATATATACATACATACAC 374
          ||||||||||||||||||||
```



```
Cp    197 TACATATATACATACATACAC 177


RESULT   14
LOCUS       HSU50912     7354 bp    DNA             PRI       26-JUN-1997
DEFINITION  Human XIST gene, poly purine-pyrimidine repeat region.
ACCESSION   U50912
NID         g1575007
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 7354)
  AUTHORS   Hendrich,B.D., Plenge,R.M. and Willard,H.F.
  TITLE     Identification and characterization of the human XIST gene
            promoter: implications for models of X chromosome inactivation
  JOURNAL   Nucleic Acids Res. 25 (13), 2661-2671 (1997)
  MEDLINE   97331054
REFERENCE   2  (bases 1 to 7354)
  AUTHORS   Hendrich,B.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAR-1996) Brian D. Hendrich, ICMB, University of
            Edinburgh, Darwin Building, King's Buildings, Edinburgh, Scotland
            EH9 3JR, UK
FEATURES             Location/Qualifiers
     source          1. .7354
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="Xq13.2"
     repeat_region   700. .7354
                     /rpt_family="Alu"
                     /rpt_unit=700. .1023
                     /rpt_unit=1230. .1400
                     /rpt_unit=5150. .5390
                     /rpt_unit=7130. .7354
     misc_feature    1400. .1625
                     /note="EST-like region"
     repeat_region   1867. .2345
                     /rpt_family="poly purine-pyrimidine repeat"
     misc_feature    3629. .3762
                     /note="EST-like region"
     misc_feature    3810. .4250
                     /note="EST-like region"
     misc_feature    4905. .5145
                     /note="EST-like region"
     misc_feature    5420. .5615
                     /note="EST-like region"
BASE COUNT     1823 a   1565 c   1736 g   2230 t
```

```
ORIGIN

  Query Match           7.0%;  Score 21;  DB 27;  Length 7354;
  Best Local Similarity 100.0%;  Pred. No. 5.37e-02;
  Matches   21;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2076 ACATATATACATACATACACA 2096
           |||||||||||||||||||||
Cp    196 ACATATATACATACATACACA 176


RESULT   15
LOCUS       HSXIST1     31520 bp    DNA             PRI       22-AUG-1997
DEFINITION  Human Xq13 5' end of PAC 92E23 containing the X inactivation
            transcipt (XIST) gene, complete sequence.
ACCESSION   U80459
NID         g2341005
KEYWORDS    HTG.
SEGMENT     1 of 2
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 31520)
  AUTHORS   Dugan-Roche,S. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-1996) Molecular and Human Genetics, Baylor
            College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE   2  (bases 1 to 31520)
  AUTHORS   Chiu,M.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-AUG-1997) Molecular and Human Genetics, Baylor
            College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Aug 22, 1997 this sequence version replaced gi:1850353.
FEATURES             Location/Qualifiers
     source          1. .31520
                     /organism="Homo sapiens"
                     /note="PAC 92E23 is about 170 kb long; there are
                     approximately 100 bases between the 5' and 3' segments;
                     Alu repeat region flanks the missing region"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /clone="92E23"
                     /map="Xq13"
     repeat_region   1. .206
                     /note="AluSp"
                     /rpt_family="SINE/Alu"
     repeat_region   221. .568
                     /note="L1"
                     /rpt_family="LINE/L1"
```



```
     repeat_region   569. .864
                     /note="AluSq"
                     /rpt_family="SINE/Alu"
     repeat_region   865. .1133
                     /note="L1"
                     /rpt_family="LINE/L1"
     STS             889. .1222
     repeat_region   complement(1120. .1442)
                     /note="L1M3"
                     /rpt_family="LINE/L1"
     repeat_region   complement(1443. .1743)
                     /note="AluSg"
                     /rpt_family="SINE/Alu"
     repeat_region   complement(1745. .1816)
                     /note="L1M3"
                     /rpt_family="LINE/L1"
     repeat_region   complement(1819. .2118)
                     /note="AluSg"
                     /rpt_family="SINE/Alu"
     repeat_region   complement(2158. .2639)
                     /note="L1PA15"
                     /rpt_family="LINE/L1"
     repeat_region   complement(2660. .2948)
                     /note="AluSp"
                     /rpt_family="SINE/Alu"
     repeat_region   complement(2951. .2976)
                     /note="(CA)n"
                     /rpt_family="Simple_repeat"
     repeat_region   complement(3118. .3392)
                     /note="AluSg"
                     /rpt_family="SINE/Alu"
     repeat_region   complement(3400. .3467)
                     /note="AT_rich"
                     /rpt_family="Low_complexity"
     repeat_region   complement(3479. .3763)
                     /note="AluSx"
                     /rpt_family="SINE/Alu"
     repeat_region   complement(3791. .5349)
                     /note="L1PA2"
                     /rpt_family="LINE/L1"
     repeat_region   complement(5332. .6622)
                     /note="L1PA2"
                     /rpt_family="LINE/L1"
     repeat_region   complement(6623. .6916)
                     /note="AluSg"
                     /rpt_family="SINE/Alu"
     repeat_region   complement(6918. .10394)
                     /note="L1PA2"
                     /rpt_family="LINE/L1"
     repeat_region   complement(10399. .11000)
                     /note="L1P5"
```



```
                     /rpt_family="LINE/L1"
     repeat_region   complement(11002. .11300)
                     /note="AluSx"
                     /rpt_family="SINE/Alu"
     repeat_region   complement(11332. .15484)
                     /note="L1PA2"
                     /rpt_family="LINE/L1"
     repeat_region   complement(15486. .15779)
                     /note="AluY"
                     /rpt_family="SINE/Alu"
     repeat_region   complement(15782. .15952)
                     /note="L1P5"
                     /rpt_family="LINE/L1"
     repeat_region   complement(15959. .16255)
                     /note="AluSx"
                     /rpt_family="SINE/Alu"
     repeat_region   complement(16277. .16565)
                     /note="AluSg"
                     /rpt_family="SINE/Alu"
     repeat_region   complement(16566. .17174)
                     /note="L1P5"
                     /rpt_family="LINE/L1"
     repeat_region   complement(17176. .17464)
                     /note="AluY"
                     /rpt_family="SINE/Alu"
     repeat_region   complement(17465. .18150)
                     /note="L1P5"
                     /rpt_family="LINE/L1"
     repeat_region   complement(18162. .18483)
                     /note="L1PA13"
                     /rpt_family="LINE/L1"
     repeat_region   complement(18484. .18924)
                     /note="L1P5"
                     /rpt_family="LINE/L1"
     repeat_region   complement(18932. .19234)
                     /note="AluSp"
                     /rpt_family="SINE/Alu"
     repeat_region   complement(19235. .19593)
                     /note="L1P5"
                     /rpt_family="LINE/L1"
     repeat_region   complement(19594. .19773)
                     /note="AluSx"
                     /rpt_family="SINE/Alu"
     repeat_region   19774. .20954
                     /note="SVA"
                     /rpt_family="Other"
     repeat_region   complement(20955. .21091)
                     /note="AluSx"
                     /rpt_family="SINE/Alu"
     repeat_region   complement(21106. .21562)
                     /note="L1PA7"
```



```
                     /rpt_family="LINE/L1"
     repeat_region   21563. .22052
                     /note="L1PA7"
                     /rpt_family="LINE/L1"
     repeat_region   complement(22051. .22162)
                     /note="L1"
                     /rpt_family="LINE/L1"
     repeat_region   complement(22194. .22673)
                     /note="L1"
                     /rpt_family="LINE/L1"
     repeat_region   22675. .22974
                     /note="AluSg"
                     /rpt_family="SINE/Alu"
     repeat_region   complement(23133. .23225)
                     /note="L1M1/2"
                     /rpt_family="LINE/L1"
     repeat_region   23229. .23362
                     /note="FLAM_C"
                     /rpt_family="SINE/Alu"
     repeat_region   complement(23378. .23595)
                     /note="L1M3/4"
                     /rpt_family="LINE/L1"
     repeat_region   complement(23652. .23835)
                     /note="L1M1/2"
                     /rpt_family="LINE/L1"
     repeat_region   23839. .23978
                     /note="(TA)n"
                     /rpt_family="Simple_repeat"
     repeat_region   complement(23988. .24264)
                     /note="(CATA)n"
                     /rpt_family="Simple_repeat"
     repeat_region   complement(24334. .24449)
                     /note="L1M1/2"
                     /rpt_family="LINE/L1"
     repeat_region   complement(24507. .24673)
                     /note="L1M1/2"
                     /rpt_family="LINE/L1"
     repeat_region   complement(24892. .25159)
                     /note="L1M1/2"
                     /rpt_family="LINE/L1"
     repeat_region   27085. .27356
                     /note="AluSx"
                     /rpt_family="SINE/Alu"
     repeat_region   complement(28224. .28630)
                     /note="L1M1/2"
                     /rpt_family="LINE/L1"
     repeat_region   complement(28639. .29076)
                     /note="L1"
                     /rpt_family="LINE/L1"
     repeat_region   complement(29082. .29378)
                     /note="AluSp"
```

```
                     /rpt_family="SINE/Alu"
     repeat_region   complement(29379. .29481)
                     /note="L1"
                     /rpt_family="LINE/L1"
     repeat_region   complement(29495. .29744)
                     /note="AluJo"
                     /rpt_family="SINE/Alu"
     repeat_region   complement(29746. .29860)
...
Note: remainder of annotations omitted.

  Query Match           7.0%;  Score 21;  DB 26;  Length 31520;
  Best Local Similarity 100.0%;  Pred. No. 5.37e-02;
  Matches   21;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  24010 ACATATATACATACATACACA 24030
          |||||||||||||||||||||
Cp    196 ACATATATACATACATACACA 176


RESULT   16
LOCUS       HS321I15    75793 bp    DNA             PRI       03-APR-1998
DEFINITION  Human DNA sequence from PAC 321I15 on chromosome Xq21.3.
ACCESSION   AL008632
NID         g3036776
KEYWORDS    Xq21.3.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 75793)
  AUTHORS   Mistry,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-OCT-1997) sanger.ac.uk/HGP/ChrX/) Sanger Centre,
            Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquires:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Apr 8, 1998 this sequence version replaced gi:2578085.
            IMPORTANT: This sequence is the entire insert of clone 321I15.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variations annotated may not be found in the sequence submission
            corresponding to the overlapping clone as we submit sequences with
            only a small overlap as described above.
            This sequence was generated from part of bacterial clone contigs of
            human chromosome X, constructed by the Sanger Centre chromosome X
            mapping group. Further information can be found at
            http://www.sanger.ac.uk/HGP/ChrX/
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
```



```
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            The true left end of clone 321I15 is at 1 in this sequence. The
            true right end of clone 321I15 is at 75793.
            321I15 is from the library RPCI1 constructed at the Roswell Park
            Cancer Institute by the group of Pieter de Jong.
            For further details see http://bacpac.med.buffalo.edu/.
FEATURES            Location/Qualifiers
     source         1. .75793
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="X"
                    /map="Xq21.3"
                    /clone="321I15"
                    /clone_lib="RPCI1"
     repeat_region  1184. .1696
                    /note="L1PA5 repeat: matches 378. .892 of consensus"
     repeat_region  2499. .2800
                    /note="AluSq repeat: matches 1. .302 of consensus"
     repeat_region  4436. .4469
                    /note="17 copies of 2 mer 82 % conserved"
     repeat_region  5196. .5346
                    /note="MER4A2 repeat: matches 1. .152 of consensus"
     repeat_region  5347. .5646
                    /note="MER4A2 repeat: matches 211. .506 of consensus"
     repeat_region  6135. .6220
                    /note="MER34 repeat: matches 530. .450 of consensus"
     repeat_region  7224. .7954
                    /note="LTR8 repeat: matches 690. .1 of consensus"
     repeat_region  9239. .9272
                    /note="17 copies of 2 mer 82 % conserved"
     repeat_region  9487. .9788
                    /note="AluSq repeat: matches 1. .302 of consensus"
     repeat_region  10896. .11248
                    /note="THE1C repeat: matches 370. .1 of consensus"
     repeat_region  11987. .12612
                    /note="LTR8 repeat: matches 1. .558 of consensus"
     repeat_region  13703. .14143
                    /note="MER4B repeat: matches 611. .194 of consensus"
     repeat_region  14353. .14597
                    /note="MER4B repeat: matches 219. .1 of consensus"
     repeat_region  16357. .16402
                    /note="23 copies of 2 mer 89 % conserved"
     repeat_region  16825. .16950
                    /note="AluJo repeat: matches 136. .11 of consensus;
                    incomplete repeat"
     repeat_region  18670. .18972
                    /note="AluJo repeat: matches 301. .2 of consensus"
     repeat_region  19098. .19184
                    /note="AluJb repeat: matches 1. .87 of consensus;
```



```
                    incomplete repeat"
     repeat_region  19189. .19340
                    /note="MER39 repeat: matches 311. .460 of consensus"
     repeat_region  19990. .20292
                    /note="AluSq repeat: matches 1. .303 of consensus"
     repeat_region  21851. .21989
                    /note="L1PB2 repeat: matches 183. .41 of consensus"
     repeat_region  21851. .21953
                    /note="L1MD2 repeat: matches 183. .81 of consensus"
     repeat_region  22141. .22409
                    /note="L1ME3 repeat: matches 911. .635 of consensus"
     repeat_region  23025. .23532
                    /note="2 copies of 254 mer 100 % conserved"
     repeat_region  23884. .24289
                    /note="MLT1D repeat: matches 57. .503 of consensus"
     repeat_region  26286. .26319
                    /note="17 copies of 2 mer 88 % conserved"
     repeat_region  26910. .26955
                    /note="23 copies of 2 mer 83 % conserved"
     repeat_region  28242. .28531
                    /note="AluY repeat: matches 1. .290 of consensus"
     repeat_region  29316. .29645
                    /note="AluJo repeat: matches 302. .1 of consensus"
     repeat_region  30847. .30979
                    /note="FLAM_C repeat: matches 133. .1 of consensus"
     repeat_region  30997. .31030
                    /note="17 copies of 2 mer 82 % conserved"
     repeat_region  32102. .32229
                    /note="AluJo repeat: matches 129. .1 of consensus;
                    incomplete repeat"
     repeat_region  32751. .33112
                    /note="MLT1A2 repeat: matches 1. .374 of consensus"
     repeat_region  33120. .33192
                    /note="MER5B repeat: matches 155. .84 of consensus"
     repeat_region  33812. .34808
                    /note="L1 repeat: matches 4372. .5384 of consensus"
     repeat_region  34668. .35665
                    /note="L1MA4 repeat: matches 4. .1046 of consensus"
     repeat_region  36335. .36634
                    /note="AluJo repeat: matches 1. .302 of consensus"
     repeat_region  38132. .38448
                    /note="MER4B repeat: matches 611. .297 of consensus"
     repeat_region  38453. .38682
                    /note="MER4B repeat: matches 231. .1 of consensus"
     repeat_region  39063. .39247
                    /note="AluJb repeat: matches 1. .177 of consensus;
                    incomplete repeat"
     repeat_region  39702. .39837
                    /note="AluJb repeat: matches 166. .301 of consensus;
                    incomplete repeat"
     repeat_region  39884. .40234
```



```
                    /note="MLT1A1 repeat: matches 365. .1 of consensus"
     repeat_region  40875. .41053
                    /note="MER5B repeat: matches 1. .173 of consensus"
     repeat_region  41288. .41418
                    /note="AluJo repeat: matches 132. .1 of consensus;
                    incomplete repeat"
     repeat_region  41628. .41914
                    /note="L1PB3 repeat: matches 896. .591 of consensus"
     repeat_region  41746. .42317
                    /note="L1MA5 repeat: matches 758. .136 of consensus"
     repeat_region  42058. .42427
                    /note="L1PB2 repeat: matches 395. .29 of consensus"
     repeat_region  42411. .44898
                    /note="L1 repeat: matches 5377. .2924 of consensus"
     repeat_region  44915. .45196
                    /note="AluJo repeat: matches 17. .293 of consensus;
                    incomplete repeat"
     repeat_region  45240. .45848
                    /note="L1 repeat: matches 2875. .2257 of consensus"
     repeat_region  45872. .46170
                    /note="AluJb repeat: matches 5. .302 of consensus"
     repeat_region  46258. .46557
                    /note="AluJo repeat: matches 300. .1 of consensus"
     repeat_region  46572. .47422
                    /note="L1 repeat: matches 2198. .1356 of consensus"
     repeat_region  47151. .47767
                    /note="MER25 repeat: matches 2135. .1510 of consensus"
     repeat_region  48136. .48214
                    /note="MSTA repeat: matches 79. .1 of consensus"
     repeat_region  48215. .49832
                    /note="MST-INTERNAL repeat: matches 1651. .1 of consensus"
     repeat_region  49833. .50232
                    /note="MSTA repeat: matches 426. .1 of consensus"
     repeat_region  50697. .50903
                    /note="MER25 repeat: matches 1504. .1297 of consensus"
     repeat_region  51475. .51694
                    /note="MER25 repeat: matches 1296. .1079 of consensus"
     repeat_region  51723. .52016
                    /note="AluSx repeat: matches 297. .1 of consensus"
     repeat_region  52028. .52484
                    /note="MER25 repeat: matches 1057. .602 of consensus"
     repeat_region  52672. .53166
                    /note="MER25 repeat: matches 548. .49 of consensus"
     repeat_region  53615. .53733
                    /note="FLAM_A repeat: matches 128. .11 of consensus"
     repeat_region  53875. .54925
                    /note="THE1B-INTERNAL repeat: matches 107. .1167 of
                    consensus"
     repeat_region  54929. .55248
                    /note="THE1C repeat: matches 54. .368 of consensus"
     repeat_region  55252. .56389
```

```
                     /note="L1 repeat: matches 3226. .2078 of consensus"
     repeat_region   56380. .56619
                     /note="L1MB5 repeat: matches 242. .2 of consensus"
     repeat_region   56472. .57179
                     /note="L1 repeat: matches 5390. .4668 of consensus"
     repeat_region   57183. .57486
                     /note="AluSq repeat: matches 1. .303 of consensus"
     repeat_region   57509. .59386
                     /note="L1 repeat: matches 4680. .2773 of consensus"
     repeat_region   59391. .59423
                     /note="11 copies of 3 mer 85 % conserved"
     repeat_region   59435. .59746
                     /note="AluJo repeat: matches 1. .302 of consensus"
     repeat_region   59770. .59802
                     /note="AluSx/g repeat: matches 36. .4 of consensus;
                     incomplete repeat"
     repeat_region   59806. .60759
...
Note: remainder of annotations omitted.

  Query Match             7.0%;  Score 21;  DB 26;  Length 75793;
  Best Local Similarity 100.0%;  Pred. No. 5.37e-02;
  Matches    21;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  10099 TGTGTATGTATGTATATATGT 10119
          |||||||||||||||||||||
Qy    176 TGTGTATGTATGTATATATGT 196


RESULT   17
LOCUS       AC004940   203583 bp    DNA             HTG       12-JUN-1998
DEFINITION  *** SEQUENCING IN PROGRESS *** Homo sapiens clone DJ0978E18; HTGS
            phase 1, 20 unordered pieces.
ACCESSION   AC004940
NID         g1213057
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 203583)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 203583)
  AUTHORS   Wateraton,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-1998) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     ***                                                      ***
```



```
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                  ***
            * This sequence is unfinished. It consists of 20 contigs for
            * which the order is not known; their order in this record is
            * arbitrary. In some cases, the exact lengths of the gaps
            * between the contigs are also unknown; these gaps are  presented
            * as runs of N as a convenience only. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
            *        1     1478: contig of 1478 bp in length
            *     1479     1497: gap of unknown length
            *     1498    11497: contig of 10000 bp in length
            *    11498    11516: gap of unknown length
            *    11517    12974: contig of 1458 bp in length
            *    12975    12993: gap of unknown length
            *    12994    14494: contig of 1501 bp in length
            *    14495    14513: gap of unknown length
            *    14514    16055: contig of 1542 bp in length
            *    16056    16074: gap of unknown length
            *    16075    17698: contig of 1624 bp in length
            *    17699    17717: gap of unknown length
            *    17718    19199: contig of 1482 bp in length
            *    19200    19218: gap of unknown length
            *    19219    20872: contig of 1654 bp in length
            *    20873    20891: gap of unknown length
            *    20892    22362: contig of 1471 bp in length
            *    22363    22381: gap of unknown length
            *    22382    24386: contig of 2005 bp in length
            *    24387    24405: gap of unknown length
            *    24406    26861: contig of 2456 bp in length
            *    26862    26880: gap of unknown length
            *    26881    28760: contig of 1880 bp in length
            *    28761    28779: gap of unknown length
            *    28780    30381: contig of 1602 bp in length
            *    30382    30400: gap of unknown length
            *    30401    32390: contig of 1990 bp in length
            *    32391    32409: gap of unknown length
            *    32410    34243: contig of 1834 bp in length
            *    34244    34262: gap of unknown length
            *    34263    54985: contig of 20723 bp in length
            *    54986    55003: gap of unknown length
            *    55004    90645: contig of 35642 bp in length
            *    90646    90663: gap of unknown length
            *    90664   200400: contig of 109737 bp in length
            *   200401   200418: gap of unknown length
            *   200419   202020: contig of 1602 bp in length
            *   202021   202038: gap of unknown length
            *   202039   203583: contig of 1545 bp in length.
FEATURES             Location/Qualifiers
     source          1. .203583
                     /organism="Homo sapiens"
```



```
                     /db_xref="taxon:9606"
                     /clone="DJ0978E18"
BASE COUNT    64295 a  38222 c  37636 g  63073 t    357 others
ORIGIN

  Query Match             7.0%;  Score 21;  DB 17;  Length 203583;
  Best Local Similarity 100.0%;  Pred. No. 5.37e-02;
  Matches    21;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  53664 ATTTACATATATACATACATA 53684
          |||||||||||||||||||||
Cp    200 ATTTACATATATACATACATA 180


RESULT   18
LOCUS       CHKMCW82      182 bp    DNA             VRT       09-NOV-1995
DEFINITION  Gallus domesticus DNA microsatellite marker MCW82.
ACCESSION   L43637
NID         g1063498
KEYWORDS    CA-microsatellite; polymorphic microsatellite.
SOURCE      Gallus gallus (clone: O71) (tissue library: lambda ZAP) female
            adult blood DNA.
  ORGANISM  Gallus gallus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Archosauria; Aves; Neognathae; Galliformes;
            Phasianidae; Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 182)
  AUTHORS   Groenen,M.A.M., Crooijmans,R.P.M. and van der Poel,J.J.
  TITLE     Preliminary linkage map of the chicken (Gallus domesticus) genome
            based on microsatellite markers: 77 new markers mapped
  JOURNAL   Unpublished (1996)
FEATURES             Location/Qualifiers
     source          1. .182
                     /organism="Gallus gallus"
                     /db_xref="taxon:9031"
                     /clone="071"
                     /dev_stage="adult"
                     /germline
                     /sex="female"
                     /tissue_type="blood"
                     /tissue_lib="lambda ZAP"
     primer_bind     1. .21
     repeat_region   18. .79
                     /rpt_family="CA- and TA-microsatellite"
     primer_bind     100. .120
BASE COUNT       62 a      12 c      41 g      67 t
ORIGIN

  Query Match             6.7%;  Score 20;  DB 20;  Length 182;
  Best Local Similarity 100.0%;  Pred. No. 3.01e-01;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```



```
Db     54 GTGTGTATGTATGTATATAT 73
          ||||||||||||||||||||
Qy    175 GTGTGTATGTATGTATATAT 194


RESULT   19
LOCUS       G37914        225 bp    DNA             STS       17-APR-1998
DEFINITION  c3m52 Plasmodium falciparum haploid Plasmodium falciparum STS
            genomic, sequence tagged site.
ACCESSION   G37914
NID         g3056658
KEYWORDS    STS.
SOURCE      malaria parasite.
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 225)
  AUTHORS   Su,X-z., Ferdig,M.T. and Wellems,T.E.
  TITLE     Integrated genetic and physical maps of Plasmodium falciparum
  JOURNAL   Unpublished (1998)
COMMENT
            Contact: Thomas E. Wellems
            LPD
            NIAID
            Bldg. 4, Room 126, NIH Campus, Bethesda, MD 20892-0425
            Tel: (301) 496-4021
            Fax: (301) 402-0079
            Email: tew@helix.nih.gov
            Primer A: GAAAGCAATTATCCAAGGTT
            Primer B: GTCAAAGACCCTAAAGGCAC
            STS size: 225
            PCR Profile:
            Initial heat: 2min at 94oC
            Cycles:       30 cycles, each having steps 20sec at 94oC;
                          10sec at 45oC; 10sec at 40oC; 30sec at 60oC
            Machine:      Perkin Elmer GeneAmp 9600
            Protocol:
                  Template:        1-10 ng
                  Primer:          5 pM each
                  dNTPs:           0.3 ul of 10 mM stock solution
                  Taq Polymerase:  0.5 unit
                  Total Volume:    15 ul

            Buffer:
                   10 mM Tris HCl, pH 8.3
                   1.5 mM MgCl2
                   50 mM KCl.
FEATURES             Location/Qualifiers
     source          1. .225
                     /organism="Plasmodium falciparum"
                     /strain="HB3xDd2"
```

```
                     /db_xref="taxon:5833"
                     /map="3"
                     /clone_lib="Plasmodium falciparum haploid"
                     /dev_stage="erythrocytic haploid stage"
     STS             1. .225
     primer_bind     1. .20
     primer_bind     complement(206. .225)
BASE COUNT       97 a      9 c     19 g    100 t
ORIGIN

  Query Match             6.7%;  Score 20;  DB 30;  Length 225;
  Best Local Similarity 100.0%;  Pred. No. 3.01e-01;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    136 TGTATGTATGTATATATGTA 155
          ||||||||||||||||||||
Qy    178 TGTATGTATGTATATATGTA 197


RESULT   20
LOCUS       I31098        264 bp    DNA             PAT       20-DEC-1996
DEFINITION  Sequence 10 from patent US 5582979.
ACCESSION   I31098
NID         g1821889
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 264)
  AUTHORS   Weber,J.L.
  TITLE     Length polymorphisms in (dC-dA).sub.n.(dG-dT).sub.n sequences and
            method of using the same
  JOURNAL   Patent: US 5582979-A 10 10-DEC-1996;
FEATURES             Location/Qualifiers
     source          1. .264
                     /organism="unknown"
BASE COUNT      105 a     65 c     48 g     46 t
ORIGIN

  Query Match             6.7%;  Score 20;  DB 21;  Length 264;
  Best Local Similarity 100.0%;  Pred. No. 3.01e-01;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    139 ATATATACATACATACACAC 158
          ||||||||||||||||||||
Cp    194 ATATATACATACATACACAC 175


RESULT   21
LOCUS       HSMFD22       264 bp    DNA             PRI       05-AUG-1991
DEFINITION  Human MFD22 dinucleotide repeat DNA.
```



```
ACCESSION   X54567
NID         g34571
KEYWORDS    dinucleotide repeat.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 264)
  AUTHORS   Weber,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-SEP-1990) Weber J.L., Marshfield Medical Research
            Foundation, 1000 N Oak, Marshfield WI 54449, U.S.A.
REFERENCE   2  (bases 1 to 264)
  AUTHORS   Weber,J.L. and May,P.E.
  TITLE     Dinucleotide repeat polymorphism at the D20S27 locus
  JOURNAL   Nucleic Acids Res. 18 (8), 2202 (1990)
  MEDLINE   90245698
FEATURES             Location/Qualifiers
     source          1. .264
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="blood"
                     /sex="Male"
     misc_feature    1. .264
                     /note="Only one strand sequenced"
                     /evidence=experimental
     misc_feature    89. .107
                     /standard_name="PCR primer"
                     /evidence=experimental
     repeat_region   153. .193
                     /rpt_type=TANDEM
                     /rpt_family="(dC-dA)n.(dG-dT)n"
     misc_feature    complement(220. .238)
                     /standard_name="PCR primer"
                     /evidence=experimental
BASE COUNT      105 a     65 c     48 g     46 t
ORIGIN

  Query Match             6.7%;  Score 20;  DB 25;  Length 264;
  Best Local Similarity 100.0%;  Pred. No. 3.01e-01;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    139 ATATATACATACATACACAC 158
          ||||||||||||||||||||
Cp    194 ATATATACATACATACACAC 175


RESULT   22
LOCUS       HS240XD10     265 bp    DNA             STS       04-NOV-1995
DEFINITION  H.sapiens DNA segment containing (CA) repeat; clone AFM240xd10;
            single read.
```



```
ACCESSION   Z66857
NID         g1051640
KEYWORDS    CA repeat; dinucléotide repeat; GT repeat; microsatellite DNA;
            microsatellite marker; repeat polymorphism; STS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 265)
  AUTHORS   Weissenbach,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-OCT-1995) Genethon, B.P. 60, 91002 Evry Cedex France.
            E-mail: Jean.Weissenbach@genethon.fr
COMMENT     full automatic.
FEATURES             Location/Qualifiers
     source          1. .265
                     /organism="Homo sapiens"
                     /note="cloning vector is M13mp18"
                     /db_xref="taxon:9606"
                     /cell_line="CEPH 134702"
                     /clone_lib="genomic DNA"
                     /chromosome="4"
BASE COUNT      105 a     64 c     47 g     47 t      2 others
ORIGIN

  Query Match             6.7%;  Score 20;  DB 30;  Length 265;
  Best Local Similarity 100.0%;  Pred. No. 3.01e-01;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    139 ATATATACATACATACACAC 158
          ||||||||||||||||||||
Cp    194 ATATATACATACATACACAC 175


RESULT   23
LOCUS       HS324XH5      392 bp    DNA             STS       04-NOV-1995
DEFINITION  H.sapiens DNA segment containing (CA) repeat; clone AFM324xh5;
            single read.
ACCESSION   Z67039
NID         g1051822
KEYWORDS    CA repeat; dinucleotide repeat; GT repeat; microsatellite DNA;
            microsatellite marker; repeat polymorphism; STS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 392)
  AUTHORS   Weissenbach,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-OCT-1995) Genethon, B.P. 60, 91002 Evry Cedex France.
            E-mail: Jean.Weissenbach@genethon.fr
```



```
COMMENT     full automatic.
FEATURES             Location/Qualifiers
     source          1. .392
                     /organism="Homo sapiens"
                     /note="cloning vector is M13mp18"
                     /db_xref="taxon:9606"
                     /cell_line="CEPH 134702"
                     /clone_lib="genomic DNA"
                     /chromosome="1"
BASE COUNT      115 a     94 c     75 g    108 t
ORIGIN

  Query Match             6.7%;  Score 20;  DB 30;  Length 392;
  Best Local Similarity 100.0%;  Pred. No. 3.01e-01;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    211 ATATATACATACATACACAC 230
          ||||||||||||||||||||
Cp    194 ATATATACATACATACACAC 175


RESULT   24
LOCUS       MMU31315      546 bp    DNA             ROD       22-AUG-1995
DEFINITION  Mus musculus ACAT tetranucleotide repeat sequence, chromosome 2,
            marker um-m7.
ACCESSION   U31315
NID         g950207
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;
            Murinae; Mus.
REFERENCE   1  (bases 1 to 546)
  AUTHORS   Chamberlain,J.S. and Kiousis,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-1995) Jeffrey S. Chamberlain, Human Genetics,
            University of Michigan, 1150 W. Medical Center Drive, Ann Arbor, MI
            48109, USA
FEATURES             Location/Qualifiers
     source          1. .546
                     /organism="Mus musculus"
                     /note="marker um-m7"
                     /db_xref="taxon:10090"
                     /chromosome="2"
     repeat_region   311. .427
                     /note="complex repeat"
                     /rpt_unit=ACAT
BASE COUNT      141 a    141 c    112 g    149 t      3 others
ORIGIN
```

```
  Query Match           6.7%;  Score 20;  DB 28;  Length 546;
  Best Local Similarity 100.0%;  Pred. No. 3.0le-01;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    361 TACATATATACATACATACA 380
          ||||||||||||||||||||
Cp    197 TACATATATACATACATACA 178


RESULT   25
LOCUS       MMC35FLA     1516 bp    DNA             ROD       20-MAY-1992
DEFINITION  M.musculus DNA for C3 5'-flanking region.
ACCESSION   X62903
NID         g50232
KEYWORDS    5' flanking region; c3 gene; complement protein C3; cytokine.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;
            Murinae; Mus.
REFERENCE   1  (bases 1 to 1516)
  AUTHORS   Wetsel,R.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-OCT-1991) R.A. Wetsel, Washington University, School
            of Med., Department of Pediatrics, Box 8116, 400 South Kensington,
            St. Louis, MO 63110, USA
REFERENCE   2  (bases 1 to 1516)
  AUTHORS   Kawamura,N., Singer,L., Wetsel,R.A. and Colten,H.R.
  TITLE     Cis- and trans-acting elements required for constitutive and
            cytokine-regulated expression of the mouse complement C3 gene
  JOURNAL   Biochem. J. 283 (Pt 3), 705-712 (1992)
  MEDLINE   92272667
COMMENT     For related sequences see X62904, J00367-8 & K02408.
FEATURES             Location/Qualifiers
     source          1. .1516
                     /organism="Mus musculus"
                     /strain="B10.D2/nSnJ"
                     /db_xref="taxon:10090"
                     /dev_stage="fully developed"
                     /tissue_type="liver"
                     /chromosome="17"
     TATA_signal     194. .200
     TATA_signal     845. .851
     TATA_signal     1056. .1062
     enhancer        1368. .1417
                     /note="including IL-1/IL-6 responsive elements"
                     /evidence=experimental
     TATA_signal     1428. .1433
BASE COUNT      468 a    398 c    291 g    359 t
ORIGIN
```



```
  Query Match           6.7%;  Score 20;  DB 28;  Length 1516;
  Best Local Similarity 100.0%;  Pred. No. 3.0le-01;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    107 TACATATATACATACATACA 126
          ||||||||||||||||||||
Cp    197 TACATATATACATACATACA 178


RESULT   26
LOCUS       ATHG3P       1799 bp    DNA             PLN       27-FEB-1995
DEFINITION  Arabidopsis thaliana chloroplast glyceraldehyde-3-phosphate
            dehydrogenase B subunit gene, 5'end.
ACCESSION   L14749
NID         g484109
KEYWORDS    glyceraldehyde-3-phosphate dehydrogenase B subunit; light
            responsive element.
SOURCE      Chloroplast Arabidopsis thaliana (strain Columbia, organelle
            Chloroplast Arabidopsis thaliana) adult leaves DNA.
  ORGANISM  Chloroplast Arabidopsis thaliana
            Eukaryotae; mitochondrial eukaryotes; Viridiplantae;
            Charophyta/Embryophyta.group; Embryophyta; Magnoliophyta;
            Magnoliopsida; Capparales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1799)
  AUTHORS   Kwon,H.B., Park,S.C., Peng,H.P., Goodman,H.M., Dewdney,J. and
            Shih,M.C.
  TITLE     Identification of a light-responsive region of the nuclear gene
            encoding the B subunit of chloroplast glyceraldehyde 3-phosphate
            dehydrogenase from Arabidopsis thaliana
  JOURNAL   Plant Physiol. 105 (1), 357-367 (1994)
  MEDLINE   94302174
FEATURES             Location/Qualifiers
     source          1. .1799
                     /organism="Arabidopsis thaliana"
                     /chloroplast
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /dev_stage="adult"
                     /tissue_type="leaves"
     TATA_signal     1713. .1719
     CDS             1783. .>1799
                     /codon_start=1
                     /product="glyceraldehyde-3-phosphate dehydrogenase B
                     subunit"
                     /db_xref="PID:g484110"
                     /translation="MATHAA"
BASE COUNT      482 a    357 c    317 g    643 t
ORIGIN

  Query Match           6.7%;  Score 20;  DB 23;  Length 1799;
  Best Local Similarity 100.0%;  Pred. No. 3.0le-01;
```



```
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1700 GTGTGTATGTATGTATATAT 1719
          ||||||||||||||||||||
Qy    175 GTGTGTATGTATGTATATAT 194


RESULT   27
LOCUS       RATSCG10A    2227 bp    DNA             ROD       22-SEP-1992
DEFINITION  Rattus norvegicus neural-restrictive silencer element (scg-10)
            gene.
ACCESSION   M90489
NID         g206863
KEYWORDS    silencer element.
SOURCE      Rattus norvegicus DNA.
  ORGANISM  Rattus norvegicus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;
            Murinae; Rattus.
REFERENCE   1  (bases 1 to 2227)
  AUTHORS   Mori,N., Schoenherr,C.J., Vandenbergh,D.J. and Anderson,D.J.
  TITLE     A common silencer element in the SCG10 and type II Na+ channel
            genes binds a fact or present in nonneuronal cells but not in
            neuronal cells
  JOURNAL   Neuron 9, 45-54 (1992)
  MEDLINE   92337920
FEATURES             Location/Qualifiers
     source          1. .2227
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
     misc_feature    627. .649
                     /gene="scg-10"
                     /note="silencer sequence"
     gene            627. .2056
                     /gene="scg-10"
     TATA_signal     2051. .2056
                     /gene="scg-10"
BASE COUNT      629 a    484 c    456 g    658 t
ORIGIN

  Query Match           6.7%;  Score 20;  DB 28;  Length 2227;
  Best Local Similarity 100.0%;  Pred. No. 3.0le-01;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    997 ATATATACATACATACACAC 1016
          ||||||||||||||||||||
Cp    194 ATATATACATACATACACAC 175


RESULT   28
LOCUS       PFU27339     5430 bp    mRNA            INV       27-FEB-1996
```



```
DEFINITION  Plasmodium falciparum erythrocyte membrane protein 1, Malayan Camp
            variant 2 (MCvar-2 PfEMP1) mRNA, partial cds.
ACCESSION   U27339
NID         g914920
KEYWORDS    .
SOURCE      malaria parasite.
  ORGANISM  Plasmodium falciparum
            Eukaryotae; mitochondrial eukaryotes; Alveolata; Apicomplexa;
            Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 5430)
  AUTHORS   Baruch,D.I., Pasloske,B.L., Singh,H.B., Bi,X., Ma,X.C., Feldman,M.,
            Taraschi,T.F. and Howard,R.J.
  TITLE     Cloning the P. falciparum gene encoding PfEMP1, a malarial variant
            antigen and adherence receptor on the surface of parasitized human
            erythrocytes
  JOURNAL   Cell 82 (1), 77-87 (1995)
  MEDLINE   95330812
REFERENCE   2  (bases 1 to 5430)
  AUTHORS   Baruch,D.I.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAY-1995) Dror I. Baruch, Affymax Research Institute,
            3410 Central Expressway, Santa Clara, CA 95051, USA
FEATURES             Location/Qualifiers
     source          1. .5430
                     /organism="Plasmodium falciparum"
                     /strain="Malayan Camp (MC132 K+C+R+)"
                     /db_xref="taxon:5833"
                     /dev_stage="asexual"
     5'UTR           1. .244
     gene            245. .5430
                     /gene="MCvar-2 PfEMP1"
     CDS             245. .>5430
                     /gene="MCvar-2 PfEMP1"
                     /standard_name="Malayan Camp variant 2 Plasmodium
                     falciparum erythrocyte membrane protein 1"
                     /note="antigenic phenotype and cytoadherence properties of
                     the malaria parasitized erythrocytes are associated with
                     the transcribed PfEMP1 gene"
                     /codon_start=1
                     /function="variant surface antigen and adherence receptor
                     of P. falciparum parasitized erythrocytes"
                     /product="PfEMP1 variant 2 of strain MC"
                     /db_xref="PID:g914921"
                     /translation="MGGGNGGGGTKDKDAKHALDRIGEEVYKEKVENDAEKYKKALKG
                     NLQEAKGIGELASSPNPCKLVEDIYNNRLKRKRYPCANRQTVRFSDETGGQCTFNRIK
                     DSENNDNSIGACAPYRRLHLCDYNLEKMGKTSTTKHDLLLDVCMAARYEGDSIKTHYT
                     KHELTNPDTKSQLCTILARSFADIGDIVRGKDLYLGYDDKEKDERKKLENNLIEIPKK
                     IHENLGTQDAKDHYKKDEENYTQLRED WTANRSTVWKAITCHAGESDKYFRKTCCSG
                     EWTDDKCRCKDEEGKNETNEVPTYFDYVPQYLRWFEEWAEDFCRKRKKKIENAIKNCR
                     GEKGNERYCDLNGYNCEETARGAEIFVKGDDCHKCSVACDRFVKWIDNQRKEFDKQKK
                     KYDEEINKTHGTTIITGNGKINNLYVGHFYKILKKYYPTVDKSLQKLNDEAICKKPPN
```

VGNEKASTVDFNNEVNTTFSHTTYCEACPWCGAQKEKNGGGWKAKEKSCAKKKERIFN
KENSTDIKILTPEKGRSKTLEKLKTFCKDGQKIKNDIWKCHYDDNGTDDQTDDSNDCV
LGDWGNLTKEDKIMSYHAFFWWWVHDMLIDSIKWRDEHGRCINKDKGKTCIKGCNKKC
ICFQKWVEQKKTEWGKIKDHFRKQKDIPKDWTHDDFLQTLLMKDLLLEIIQDTYGDAN
EIKRIEALLEQAGVGGIDFAALAGLYTKGFVAEKDTTIDKLLQHEQKEADKCLKTHTD
DTCPPQEDRSVARSESATVPSPPADPKATEEVDANASSDGEDDFEEEEEEEDEGEEE
AEEVQEEKTDESATEAVAPSPPGTTQDGVKPASQEDDVKVCSIVDKALKGKLDDACTL
KYGKTAPTSWKCIPSDTKSVATTGSDTTGSGSICVPPRRRKLYVGKLHDWAGGETTEA
KSQETSGGQKTPSGNESSPSEKLPQGPTPETTKETPESSLLHAFVSPPRLRRFLPWHK
FKEQWKAQHGAGATGLQLPGVTVDDSDPDPQTQLRRGNIPNDFLRQMFYTLGDYRDIC
IGGDRDIVGDTIVSITEGESTKKKISKIIEGFLKKQTVTSPSPRDTSSRTPVHPQTSV
SKTPQQTWWEANGPHIWNGMICALTEDSGAIGQPPQKVEDADKVLEKLKPNTANGIK
WYLKEDNTSSAMPTSSSSSSGSNDPINTPKLTEFVEIPTFFRYLHEWGQNFCKERMKR
LKQIYKECKVGENGYGRGRKQKTPQCSCYGEDCEDQLSKYSYDTVADLECPKCAKHCR
WYKKWIEKKKDEFTEQEKAFPKQKDYYVNGNNKGGGDNGFCITLKSLSDAAQFLEKLG
SCKKDNSEDNGNDKLNFSQPNETFVPATNCKPCSEFKIDCKENGKCKNGGGGTNETCN
GTTPITSENFKQKGQTAKEFVMRVSDNNPNGFDDLNEACQNAGIFKSIRKDEWECGKV
CGYEVCIPEKGNGVTTSGENNDQIITIRGLVAHWVQNFLDDYNKIKHKISHCKNSSEG
YTCIKNCVEQWISTKRTEWTNIKILLNEQYKDNPDYNVKTILQDLQSQIDFNKAIKPC
GTLTKPEDSCGINGAESSEKKNGHETYDAIDCMLNRLQDKIDDCNKNHAQNGGENQAKC
EKHSAPDEDDDDEAIEEENPVTQPNICPKPPEPKAEEKGGCEPAEKKEEVEEKEPEKTVN
TVAAPTEKEAAGDPAGPAADSEENPEEEKAPEPEVETKKDKAPVKPTPASPPPLGLHL
YPP"

BASE COUNT 2122 a 904 c 1157 g 1247 t
ORIGIN

Query Match 6.7%; Score 20; DB 18; Length 5430;
Best Local Similarity 100.0%; Pred. No. 3.01e-01;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 98 TACATATATACATACATACA 117
||||||||||||||||||||
Cp 197 TACATATATACATACATACA 178

RESULT 29
LOCUS HSXLRPGN2 6608 bp DNA PRI 26-JUN-1997
DEFINITION H.sapiens RP3 gene (XLRP gene 2).
ACCESSION X94767
NID g1418965
KEYWORDS RP3 gene; X-linked retinitis pigmentosa.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
Homo.
REFERENCE 1 (bases 1 to 6608)
AUTHORS Bauer,D.
TITLE Direct Submission
JOURNAL Submitted (03-JAN-1996) D. Bauer, Institut fuer Molekulare
Biotechnologie, Genomanalyse, Beutenbergstr. 11, D- 07745 Jena, FRG

REFERENCE 2 (bases 1 to 6608)
AUTHORS Roepman,R., Bauer,D., Rosenberg,T., van Duijnhoven,G., van de
Vosse,E., Platzer,M., Rosenthal,A., Ropers,H.H., Cremers,F.P. and
Berger,W.
TITLE Identification of a gene disrupted by a microdeletion in a patient
with X-linked retinitis pigmentosa (XLRP)
JOURNAL Hum. Mol. Genet. 5 (6), 827-833 (1996)
MEDLINE 96372820
FEATURES Location/Qualifiers
source 1. .6608
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="cosmid LL242D12"
/clone_lib="YAC ICRFy90E0701"
/chromosome="X"
/map="p21.1-p11.4"
mRNA join(<1. .757,3148. .3330,4772. .>4838)
/evidence=experimental
/product="RP3"
BASE COUNT 2093 a 1278 c 1213 g 2012 t 12 others
ORIGIN

Query Match 6.7%; Score 20; DB 26; Length 6608;
Best Local Similarity 100.0%; Pred. No. 3.01e-01;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 4851 TACATATATACATACATACA 4870
||||||||||||||||||||
Cp 197 TACATATATACATACATACA 178

RESULT 30
LOCUS PNPHYT 7749 bp DNA PLN 14-SEP-1993
DEFINITION P.nudum gene for phytochrome.
ACCESSION X74931
NID g400481
KEYWORDS phy gene; phytochrome.
SOURCE Psilotum nudum.
ORGANISM Psilotum nudum
Eukaryotae; mitochondrial eukaryotes; Viridiplantae;
Charophyta/Embryophyta group; Embryophyta; Psilotophyta;
Psilotales; Psilotaceae; Psilotum.
REFERENCE 1 (bases 1 to 7749)
AUTHORS Marx,S.
TITLE Direct Submission
JOURNAL Submitted (06-SEP-1993) S. Marx, Botanisches Institut, Universitaet
Koeln, Gyrhofstrasse 15, 50931 Koeln, FRG
REFERENCE 2 (bases 1 to 7749)
AUTHORS Schneider-Poetsch,H., Marx,S., Kolukisaoglu,U., Hanelt,S. and
Braun,B.
TITLE Phytochrome evolution: Phytochrome in Mosses and Ferns

JOURNAL Unpublished
FEATURES Location/Qualifiers
source 1. .7749
/organism="Psilotum nudum"
/db_xref="taxon:3240"
/clone="K8"
gene 280. .4691
/gene="psphyg"
exon 280. .2329
/gene="psphyg"
/number=1
/product="phytochrome"
intron 2330. .3879
/gene="psphyg"
/number=1
exon 3880. .4691
/gene="psphyg"
/number=2
/product="phytochrome"
intron 4692. .>7749
BASE COUNT 2214 a 1442 c 1749 g 2344 t
ORIGIN

Query Match 6.7%; Score 20; DB 23; Length 7749;
Best Local Similarity 100.0%; Pred. No. 3.01e-01;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 6030 TACATATATACATACATACA 6049
||||||||||||||||||||
Cp 197 TACATATATACATACATACA 178

RESULT 31
LOCUS FTPPCA 9012 bp DNA PLN 21-JUL-1993
DEFINITION F.trinervia ppcA1 gene for phosphoenolpyruvate carboxylase.
ACCESSION X64143 S43592
NID g18463
KEYWORDS phosphoenolpyruvate carboxylase; ppcA1 gene.
SOURCE Flaveria trinervia.
ORGANISM Flaveria trinervia
Eukaryotae; mitochondrial eukaryotes; Viridiplantae;
Charophyta/Embryophyta group; Embryophyta; Magnoliophyta;
Magnoliopsida; Asterales; Asteraceae; Flaveria.
REFERENCE 1 (bases 1 to 9012)
AUTHORS Westhoff,P.
TITLE Direct Submission
JOURNAL Submitted (17-JAN-1992) P. Westhoff, Inst. fuer Entwicklungs- und
Molekularbiologie der Pflanzen, Heinrich-Heine-Uni Duesseldorf,
Universitaetstrasse 1, 4000 Duesseldorf 1, FRG
REFERENCE 2 (bases 1 to 9012)
AUTHORS Hermans,J. and Westhoff,P.

TITLE Homologous genes for the C4 isoform of phosphoenolpyruvate
carboxylase in a C3 and a C4 Flaveria species
JOURNAL Mol. Gen. Genet. 234 (2), 275-284 (1992)
MEDLINE 92374996
FEATURES Location/Qualifiers
source 1. .9012
/organism="Flaveria trinervia"
/db_xref="taxon:4227"
/clone_lib="lambda DASH EcoRI"
/clone="lnFtppc8"
CAAT_signal 2041. .2045
TATA_signal 2077. .2083
exon 2119. .2352
/gene="ppcA1"
/number=1
prim_transcript 2119. .8496
/gene="ppcA1"
gene 2119. .8496
/gene="ppcA1"
CDS join(2185. .2352,3185. .3576,3654. .3738,3830. .4056,
4145. .4247,4326. .4413,4496. .4650,5390. .6388,7005. .7391,
7996. .8295)
/gene="ppcA1"
/EC_number="4.1.1.31"
/codon_start=1
/product="phosphoenolpyruvate carboxylase"
/db_xref="PID:g18464"
/db_xref="SWISS-PROT:Q01648"
/translation="MANRNVEKLASIDAQLRLLVPGKVSEDDKLVEYDALLLDKFLDI
LQDLHGEDLKEAVQQCYELSAEYEGKHDPKKLDELGSLLTSLDPGDSIVIAKAFSHML
NLANLAEEVQIAYRRRIKLKSGDFADEANATTESDIEETFKRLVHKLNKSPEEVFDAL
KNQTVELVLTAHPTQSIRRSLLQKHGRIRNCLAQLYAKDITPDDKQELDEALHREIQA
AFRTDEIRRTPPTPQDEMRAGMSYFHETIWKGVPKFLRRVDTALKNIGINERFPYNAP
LIQFSSWMGGDRDGNPRVTPEVTRDVCLLARMDTSMMYFSQIEDLMIEMSMWRCNSE
LRVRAEELYRTARKDVKHYIEFWKQIPPNQPYRVILGDVRDKLYNTRERSRHLLVDGK
SDIPDEAVYTNVEQLLEPLELCYRSLCDSGDHVIADGSLLDFLRQVSTFGLSLVKLDI
RQESDRHTEVLDAITQHLGIGSYREWSEEKRQEWLLAELSGKRPLIGPDLPKTEEVKD
CLDTFNVLAELPSDCFGAYIISMATSTSDVLAVELLQREYHIKHPLRVVPLFEKLADL
EAAPAAMTRLFSMDWYRNRIDGKQEVMIGYSDSGKDAGRFSAAWQLYKTQEQIVKIAK
EFGVKLVIFHGRGGTVGRGGGPTHLALLSQPPDTINGSLRVTVQGEVIEQSFGEEHLC
FRTLQRFCAATLEHGMNPPISPRPEWRELMDQMAVVATEEYRSVVFKEPRFVEYFRLA
TPELEPGRMNIGSRPSKRKPSGGIESLRAIPWIFSWTQTRFHLPVWLGFGAAFKHAIQ
KDSKNLQMLQEMYKTWPFFRVTIDLVEMVFAKGNPGIAALNDKLLVSEDLRPFGESLR
ANYEETKNLLLKIAGHKDLLEGDPYLKQGIRLRDPYITTLNVCQAYTLKRIRDPNYHV
TLRPHISKEYAAEPSKPADELIHLNPTSEYAPGLEDTLILTMKGIAAGMQNTG"
intron 2353. .3184
/gene="ppcA1"
/number=1
exon 3185. .3576
/gene="ppcA1"
/number=2

```
     intron          3577. .3653
                     /gene="ppcA1"
                     /number=2
     exon            3654. .3738
                     /gene="ppcA1"
                     /number=3
     intron          3739. .3829
                     /gene="ppcA1"
                     /number=3
     exon            3830. .4056
                     /gene="ppcA1"
                     /number=4
     intron          4057. .4144
                     /gene="ppcA1"
                     /number=4
     exon            4145. .4247
                     /gene="ppcA1"
                     /number=5
     intron          4248. .4325
                     /gene="ppcA1"
                     /number=5
     exon            4326. .4413
                     /gene="ppcA1"
                     /number=6
     intron          4414. .4495
                     /gene="ppcA1"
                     /number=6
     exon            4496. .4650
                     /gene="ppcA1"
                     /number=7
     intron          4651. .5389
                     /gene="ppcA1"
                     /number=7
     exon            5390. .6388
                     /gene="ppcA1"
                     /number=8
     intron          6389. .7004
                     /gene="ppcA1"
                     /number=8
     exon            7005. .7391
                     /gene="ppcA1"
                     /number=9
     intron          7392. .7995
                     /gene="ppcA1"
                     /number=9
     exon            7996. .8496
                     /gene="ppcA1"
                     /number=10
     polyA_signal    8453. .8458
                     /gene="ppcA1"
     polyA_site      8496
```



```
                     /gene="ppcA1"
                     /note="by similarity to cDNA clone pcFtppc5.5"
  BASE COUNT     2817 a   1532 c   1597 g   3064 t      2 others
  ORIGIN

    Query Match           6.7%;  Score 20;  DB 23;  Length 9012;
    Best Local Similarity 100.0%;  Pred. No. 3.01e-01;
    Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

  Db    439 TGTATGTATGTATATATGTA 458
            ||||||||||||||||||||
  Qy    178 TGTATGTATGTATATATGTA 197


  RESULT   32
  LOCUS       PFU27338    10213 bp    DNA              INV       27-FEB-1996
  DEFINITION  Plasmodium falciparum erythrocyte membrane protein 1, Malayan Camp
              variant 1 (MCvar-1 PfEMP1) gene, complete cds.
  ACCESSION   U27338
  NID         g914918
  KEYWORDS    .
  SOURCE      malaria parasite.
    ORGANISM  Plasmodium falciparum
              Eukaryotae; mitochondrial eukaryotes; Alveolata; Apicomplexa;
              Haemosporida; Plasmodium.
  REFERENCE   1  (bases 1 to 10213)
    AUTHORS   Baruch,D.I., Pasloske,B.L., Singh,H.B., Bi,X., Ma,X.C., Feldman,M.,
              Taraschi,T.F. and Howard,R.J.
    TITLE     Cloning the P. falciparum gene encoding PfEMP1, a malarial variant
              antigen and adherence receptor on the surface of parasitized human
              erythrocytes
    JOURNAL   Cell 82 (1), 77-87 (1995)
    MEDLINE   95330812
  REFERENCE   2  (bases 1 to 10213)
    AUTHORS   Baruch,D.I.
    TITLE     Direct Submission
    JOURNAL   Submitted (19-MAY-1995) Dror I. Baruch, Affymax Research Institute,
              3410 Central Expressway, Santa Clara, CA 95051, USA
  FEATURES             Location/Qualifiers
       source          1. .10213
                       /organism="Plasmodium falciparum"
                       /strain="Malayan Camp (MC132 K+C+R+)"
                       /db_xref="taxon:5833"
                       /dev_stage="asexual"
                       /pop_variant="MC var-1 PfEMP1"
       exon            1. .7672
                       /number=1
       5'UTR           1. .244
       variation       203
                       /note="Clone A2 of MC-PfEMP1"
                       /replace="A"
```



```
     variation       240
                     /note="Clone A2 of MC-PfEMP1"
                     /replace="G"
     gene            245. .9744
                     /gene="MCvar-1 PfEMP1"
     CDS             join(245. .7672,8398. .9744)
                     /gene="MCvar-1 PfEMP1"
                     /standard_name="Malayan Camp variant 1 Plasmodium
                     falciparum erythrocyte membrane protein 1"
                     /note="antigenic phenotype and cytoadherence properties of
                     the malaria parasitized erythrocytes are associated with
                     the transcribed PfEMP1 gene"
                     /codon_start=1
                     /function="variant surface antigen and adherence receptor
                     of P. falciparum parasitized erythrocytes"
                     /product="PfEMP1 variant 1 of strain MC"
                     /db_xref="PID:g914919"
                     /translation="MGGGNGGGGTKDKDAKHALDRIGEEVYKEKVENDAEKYKKALKG
                     NLQEAKGIGELASSPNPCKLVEDYYNNRLKRKRYPCANRQTVRFSDEYGGQCTFNRIK
                     DSENNDNSIGACAPYRRLHLCDYNLEKMGKTSTTKHDLLLDVCMAAKYEGDSIKTHYT
                     KHELTNPDTKSQLCTILARSPADIGDIVRGKDLYLGYDDKEKDERKKLENNLIEIFKK
                     IHENLGTQDAKDHYKKDEENYYQLREDWWTANRSTVWKAITCHAGESDKYFRKTCCSG
                     EWTDDKCRCKDEEGKNETNEVPTYFDYVPQYLRWFEEWAEDFCRKRKKKIENAIKNCR
                     GEKGNERYCDLNGYNCEETARGAEIFVKGDDCHKCSVACDRFVKWIDNQRKEPDKQKK
                     KYDEINKTHGTTITTGMGKINNLYVGHFYKLKKYYPTVDKSLQKLNDEAICKKPPN
                     VGNEASTVDPNNEVNTTPSHTTYCEACPNGAQEEKNGGGTKAKEKSCAKKKERIFN
                     KENSTDIKILTPEKGRSKTLERLKTPCKDGQKIKNDIWKCHYDDNGTDDQTDDSNDCV
                     LGDWGNLTKEDKIMSYNAFPWMWVHDMLIDSIKWRDEHGRCINKDKGKTCIKGCNKKC
                     ICFQKWVEQKKTEWGKIKDHPRKQKDIPKDWTHDDFLQTLLMKDLLLEIIQDTYGDAN
                     EIKRIEALLEQAGVGGIDFAALAGLYTKGFVAEKDTTIDKLLQHEQKEADKCLKTHTD
                     DTCPPQEDRSVARSESATVPSPPADPKATEEVDANASSDDEDDFEEEEEEEEDEGEEE
                     AEEVQEERTDESATEAVAPSPPGTTQDGVKPASQEDDVKVCSIVDKALKGKLDDACTL
                     KYGKTAPTSWKCIPSGNNTTTESTTKPGAAGTPSGKDTGSICVPPRRRKLYVGKLHDW
                     AGGETTEAKSQETSGGQKTPSGNESSPSEKLPQGPTPETTKETPESSLLHAFVSPPRL
                     RRFLPWHKFKEQWKAQNGAGATGQQTIGTLDGGGEETPDKLLKTGHIPPDPLRQMFY
                     TLGDYRDILVGNTDIVVHTSGNKEDMQIMEAIQKKIEQILPTSGGSPSPPRVTQTQHS
                     VENPRKTWWNENGKKIWEGMVCALTYWTDTPSGTAPTQIQEVRTKLRDENSKNPKIPQ
                     YKYDQVKLDDTSDAKTTGSPVPSGEKITPLTDPISRPPYFRYLEEWGETPCKERKKRL
                     EKIKEECRGDRTGHEHCSGDGYDCTRTDADRNDKFVDLNCRDCHIQCRKYRKWIDIKF
                     DEYHKQEKKYQGEYDKLTKDKSSGGDNNCCKDIEKHKSAAVFLKELKHCKNGQTSENK
                     GNQEDQLNKLDFDKIPQTFSPSTYCKACPVYGVNCNGNKRGRGGTNGCTTNNEPENKE
                     NDKGAASTISILINDGSTNGATNGTTGTTDETLKECSDKYAFPKGLRKQEWTCQKKYG
                     VNQCNLTNRVNDTYFDKDIVPNEFFQRWLRYFVHDYNILKHKIDPCIKKEKQDKTEHK
                     CINGCNIKCECVRKWLEIKGNEWGNIKKHYNINSNDDKETIAYNVKSIFVDQGLFDTD
                     YKKAQKVVEDEKERKKIWGCTGHDECSEREKEENKNPITNLISELQDKITSCQNKHNP
                     NGKTACDPFPSPTPEETDPLDDDTPDPLDDDQHTEQPKFCPPPPPPPMTCVEKIAKELR
                     VEAGKINNELKGNGKDFNGKCNNVKKKNGAVIGEESCKFEQTTEHSVNNINNKCKDN
                     QNERFKIGQKWNPKYIGTIRKDLCIPPRREHMCLDDLSMLGRTTISDSSALLKKIQEA
                     AKSERDDIIRKLLEQNSCDEHRICDAMKYSPADLGDIIRGRDLWNKNSKQKGLQKRLE
                     YAFINIYNKLQNDKNKYEKDRPKYLQLRSDWWDANRKHIWNAMTCNAPDDAKFLKKNP
                     NDTSGSSSSKGIMTTHSNCGYDKEPPDYDYIPQPPRWMQEWSESPCKLLNEEMEQFEK
```



```
                     TCGECKKNSITCEDDRNGTNCENCKMQCEKYKKLIHNWKLGFDKYKEIYNEIYNNKDS
                     KINSNEYFKKPLEKLKDKCKELNSSDKCIDEATHCTKYKFSNSENKNHNNYAFKNPPK
                     EYEKACKCDAPDPLDNCPKDSATYEKACNTLLPTKLCESKTFNNDDDSWDTSFVQTSP
                     RDNTCVLVPPRRRQICLKNITTKLRSIEKIDDPKAPLMTSAYNEGKLLCELYKKDRDV
                     TLQAMKYSPYDYGDIVKGTDLISTAPLDKLKTKLNVLLKGDGTNEIKEDRGKWWTENR
                     TRVWHAMLCGYKAAGGKIEERDCSLPDDNTHQPLRWFREWSEHPCAKRQKLFNEVKRE
                     CASAQCIIETGTIDPPVCEEACTQYRDYITRKIQEYRLLNYQYNTNFNEKKAEVTKAP
                     EYFNDKCNDKCNCLSKYIDIEKKWKNMYDSFDDNDLKNKCICRQIKPKRPPKKVKPEE
                     EHTPSEQDTPPPLPPKPDDLPPPAEEPFNRDILEKTIPFGIALALGSIAFLFLKKKTK
                     SSVGNLFQILHIPKSDYDIPTKLSPNRYIPYTSGKYRGKRYIYLEGDSGTDSGYTDHY
                     SDITSSSESEYEELDINDIYVPGSPKYKTLIEVVLEPSGNNTTASGKNTPSDTQNDIQ
                     NDGIPSSKITDNEWNTLKDEFISNMLQNEPNTEPNMLGYNVDNNTHPTTSRHNVEEKP
                     FIMSIHDRDLYSGEEYSYNVKMVNNDIPISARNGNYSGIDLINDSLNSNKVDIYDELL
                     KRKDNELFGTNHTKKNTSTNSVAKNTNTDPINQLNLFHTWLDRHRDMCEKWDTNNKK
                     EELLDKLKEEWNKDNNSGNINPSGNTPPTSDIPSGKQSDIPSDNNIHSDIPYVLNTDV
                     SIQIHMDNPKPINEPSNMDTYPNNSSMDTILEDLDKPFNEPYYYDVQDDIYIDVHDHD
                     TSTVDTNAMDEPSKVQIEMDVNTKLVKEKYPIADLWDI"
     variation       555
                     /gene="MCvar-1 PfEMP1"
                     /note="Clone A2 of MC-PfEMP1"
                     /replace="C"
     variation       557
                     /gene="MCvar-1 PfEMP1"
                     /note="Clone A2 of MC-PfEMP1"
                     /replace="A"
     variation       660
                     /gene="MCvar-1 PfEMP1"
                     /note="Clone A2 of MC-PfEMP1"
                     /replace="G"
     variation       1328
                     /gene="MCvar-1 PfEMP1"
                     /note="Clone B2 of MC-PfEMP1"
                     /replace="G"
     variation       1753
                     /gene="MCvar-1 PfEMP1"
                     /note="Clone C2 of MC-PfEMP1"
                     /replace="T"
     variation       1985
                     /gene="MCvar-1 PfEMP1"
                     /note="Clone C2 of MC-PfEMP1"
                     /replace="A"
     variation       2021
                     /gene="MCvar-1 PfEMP1"
                     /note="Clone C2 of MC-PfEMP1"
                     /replace="G"
     variation       2094
                     /gene="MCvar-1 PfEMP1"
                     /note="Clone C2 of MC-PfEMP1"
                     /replace="A"
     variation       2136
                     /gene="MCvar-1 PfEMP1"
```

```
                     /note="Clone C2 of MC-PfEMP1"
                     /replace="G"
     variation       2218
                     /gene="MCvar-1 PfEMP1"
                     /note="Clone C2 of MC-PfEMP1"
                     /replace="A"
     variation       2410
                     /gene="MCvar-1 PfEMP1"
                     /note="clone C2 of MC-PfEMP1"
                     /replace="T"
     variation       2602
                     /gene="MCvar-1 PfEMP1"
                     /note="Clone D2 of MC-PfEMP1"
                     /replace="A"
     intron          7673. .8397
                     /gene="MCvar-1 PfEMP1"
     exon            8398. .10213
                     /number=2
     3'UTR           9745. .10213
BASE COUNT     4170 a   1429 c   1869 g   2745 t
ORIGIN

  Query Match           6.7%;  Score 20;  DB 18;  Length 10213;
  Best Local Similarity 100.0%;  Pred. No. 3.01e-01;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     98 TACATATATACATACATACA 117
          ||||||||||||||||||||
Cp    197 TACATATATACATACATACA 178


RESULT   33
LOCUS       MMU62765    10824 bp    DNA             ROD       21-MAY-1997
DEFINITION  Mus musculus meprin A alpha subunit (Mepla) gene, partial cds.
ACCESSION   U62765
NID         g2108049
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae;
            Murinae; Mus.
REFERENCE   1  (bases 1 to 10824)
  AUTHORS   Jiang,W., Gorbea,C.M., Flannery,A.V., Beynon,R.J., Grant,G.A. and
            Bond,J.S.
  TITLE     The alpha subunit of meprin A. Molecular cloning and sequencing,
            differential expression in inbred mouse strains, and evidence for
            divergent evolution of the alpha and beta subunits
  JOURNAL   J. Biol. Chem. 267 (13), 9185-9193 (1992)
  MEDLINE   92250517
  REMARK    This paper describes the cDNA (exon) sequence
```



```
            Erratum:[J Biol Chem 1992 Jul 5;267(19):13779]
REFERENCE   2  (bases 1 to 10824)
  AUTHORS   Jiang,W. and Flannery,A.V.
  TITLE     Correlation of the exon/intron organization to the secondary
            structures of the protease domain of mouse meprin alpha subunit
  JOURNAL   Gene 189 (1), 65-71 (1997)
  MEDLINE   97305147
REFERENCE   3  (bases 1 to 10824)
  AUTHORS   Jiang,W. and Flannery,A.V.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUL-1996) Biochemistry & Molecular Biology,
            Pennsylvania State University College of Medicine, Hershey, PA
            17033, USA
FEATURES             Location/Qualifiers
     source          1. .10824
                     /organism="Mus musculus"
                     /strain="129"
                     /db_xref="taxon:10090"
                     /chromosome="17"
                     /map="centromeric to H-2"
     mRNA            join(<1. .76,2995. .3112,6222. .6397,10601. .>10822)
                     /gene="Mepla"
     gene            1. .10824
                     /gene="Mepla"
     exon            <1. .76
                     /gene="Mepla"
     CDS             join(<1. .76,2995. .3112,6222. .6397,10601. .>10822)
                     /gene="Mepla"
                     /EC_number="3.4.24.18"
                     /function="metalloendopeptidase"
                     /note="protease domain"
                     /codon_start=1
                     /product="meprin A alpha subunit"
                     /db_xref="PID:g2108050"
                     /translation="RTRNAMRDPSSRWKLPIPYILADNLELNAKGAILHAPEMFRLKS
                     CVDFKPYEGESSYIIFQKLSGCWSMIGDQQVGQNISIGEGCDFKATIEHEILHALGFP
                     HEQSRTDRDDYVNIWWDQIITDYEHNFNTYDDNTITDLNTPYDYESLMHYGPFSFNKN
                     ESIPIYTTKIPEFNTIIGQLPDFSAIDLIRLNRMYNC"
     intron          77. .2994
                     /gene="Mepla"
     exon            2995. .3112
                     /gene="Mepla"
     intron          3113. .6221
                     /gene="Mepla"
     exon            6222. .6397
                     /gene="Mepla"
     intron          6398. .10600
                     /gene="Mepla"
     exon            10601. .10822
                     /gene="Mepla"
     intron          10823. .>10824
```



```
                     /gene="Mepla"
BASE COUNT     2997 a   2296 c   2185 g   3346 t
ORIGIN

  Query Match           6.7%;  Score 20;  DB 28;  Length 10824;
  Best Local Similarity 100.0%;  Pred. No. 3.01e-01;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   3279 TGTATGTATGTATATATGTA 3298
          ||||||||||||||||||||
Qy    178 TGTATGTATGTATATATGTA 197


RESULT   34
LOCUS       HSN104C7    28865 bp    DNA             PRI       18-AUG-1998
DEFINITION  Human DNA sequence from cosmid N104C7 on chromosome 22, complete
            sequence.
ACCESSION   Z82246
NID         g1694826
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 28865)
  AUTHORS   Hunt,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-NOV-1996) E-mail enquiries: humquery@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 28, 1996 this sequence version replaced gi:1666415.
            IMPORTANT:  This sequence is not the entire insert of clone N104C7.
             It may be shorter because we only sequence overlapping sections
            once, or longer because we arrange for a small overlap between
            neighbouring submissions.
            This sequence has been finished according to sequence map criteria
            as follows.  An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu).  Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            The true left end of clone N104C7 is at 1 in this sequence. The
            true left end of clone N117B5 is at 28766.
            N104C7 is from the human chromosome 22-specific cosmid library
            LL22NCO3, constructed at the Biomedical Sciences Division, Lawrence
            Livermore National Laboratory, Livermore, CA 94550 under the
            auspices of the National Laboratory Gene Library Project sponsored
            by the US Department of Energy.  The source of the flow sorted
            chromosomes was a human/hamster hybrid containing chromosomes Y, 22
            and 9.
            VECTOR:  Lawrist 16.
            This sequence was generated from part of bacterial clone contigs of
```



```
            human chromosome 22, constructed by the Sanger Centre Chromosome 22
            Mapping Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr22.
FEATURES             Location/Qualifiers
     source          1. .28865
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="22"
                     /map="22q11.2-qter"
                     /clone="N104C7"
                     /clone_lib="LL22NCO3"
     repeat_region   744. .831
                     /note="MIR repeat: matches 82. .171 of consensus"
     repeat_region   948. .1087
                     /note="2 copies of 70 mer 83 % conserved"
     repeat_region   1154. .1413
                     /note="LTR8 repeat: matches 408. .668 of consensus"
     repeat_region   2410. .2492
                     /note="MIR repeat: matches 127. .45 of consensus"
     repeat_region   4110. .4409
                     /note="AluSg repeat: matches 2. .300 of consensus"
     repeat_region   5137. .5282
                     /note="MIR repeat: matches 114. .262 of consensus"
     repeat_region   5396. .5467
                     /note="MIR repeat: matches 73. .145 of consensus"
     repeat_region   5750. .5938
                     /note="MIR repeat: matches 197. .4 of consensus"
     repeat_region   7164. .7285
                     /note="L1ME3A repeat: matches 907. .783 of consensus"
     repeat_region   7215. .7977
                     /note="L1ME3 repeat: matches 856. .111 of consensus"
     repeat_region   7424. .7977
                     /note="L1MC3 repeat: matches 696. .111 of consensus"
     repeat_region   8132. .8340
                     /note="MIR repeat: matches 262. .53 of consensus"
     repeat_region   8894. .9196
                     /note="MER44B repeat: matches 505. .202 of consensus"
     repeat_region   9192. .9321
                     /note="MER44A repeat: matches 148. .1 of consensus"
     repeat_region   9322. .9502
                     /note="MER20 repeat: matches 196. .25 of consensus"
     repeat_region   11388. .11483
                     /note="MIR2 repeat: matches 120. .24 of consensus"
     repeat_region   11518. .11586
                     /note="MIR repeat: matches 164. .90 of consensus"
     repeat_region   11653. .11954
                     /note="AluSp repeat: matches 303. .2 of consensus"
     repeat_region   12785. .12961
                     /note="MSTA repeat: matches 1. .201 of consensus"
     repeat_region   12963. .13258
                     /note="AluSx repeat: matches 1. .296 of consensus"
```

```
     repeat_region   13259. .13473
                     /note="MSTA repeat: matches 189. .424 of consensus"
     repeat_region   14157. .14322
                     /note="MIR repeat: matches 36. .214 of consensus"
     repeat_region   14382. .14683
                     /note="AluSq repeat: matches 303. .1 of consensus"
     repeat_region   14850. .14959
                     /note="MIR2 repeat: matches 13. .124 of consensus"
     repeat_region   15452. .15540
                     /note="MIR repeat: matches 196. .114 of consensus"
     repeat_region   15664. .15955
                     /note="AluSx repeat: matches 290. .1 of consensus"
     repeat_region   16493. .16783
                     /note="AluY repeat: matches 296. .6 of consensus"
     repeat_region   17137. .17660
                     /note="MER1A repeat: matches 519. .1 of consensus"
     repeat_region   17862. .18154
                     /note="AluSx repeat: matches 294. .1 of consensus"
     repeat_region   18444. .18744
                     /note="AluY repeat: matches 301. .1 of consensus"
     repeat_region   19098. .19125
                     /note="14 copies of 2 mer 89 % conserved"
     repeat_region   19631. .19927
                     /note="AluSc repeat: matches 297. .1 of consensus"
     repeat_region   19986. .20096
                     /note="L1ME1 repeat: matches 506. .397 of consensus"
     repeat_region   20144. .20190
                     /note="AluSx/g repeat: matches 299. .253 of consensus;
                     incomplete repeat"
     repeat_region   20399. .20715
                     /note="AluY repeat: matches 1. .301 of consensus"
     repeat_region   21892. .22193
                     /note="AluY repeat: matches 1. .301 of consensus"
     repeat_region   22615. .22823
                     /note="MIR repeat: matches 262. .56 of consensus"
     repeat_region   22978. .23114
                     /note="MIR repeat: matches 4. .146 of consensus"
     repeat_region   23963. .23994
                     /note="16 copies of 2 mer 84 % conserved"
     repeat_region   24038. .24234
                     /note="L1PA4 repeat: matches 891. .696 of consensus"
     repeat_region   25241. .25537
                     /note="AluSx repeat: matches 291. .1 of consensus"
     repeat_region   25815. .25989
                     /note="MIR repeat: matches 21. .206 of consensus"
     repeat_region   26211. .26507
                     /note="AluY repeat: matches 1. .299 of consensus"
     repeat_region   26645. .26959
                     /note="AluJo repeat: matches 300. .1 of consensus"
     repeat_region   27078. .27134
                     /note="L1MB7 repeat: matches 922. .867 of consensus"
```



```
     repeat_region   27169. .27448
                     /note="AluSg repeat: matches 1. .280 of consensus;
                     incomplete repeat"
     repeat_region   27469. .27738
                     /note="L1MB7 repeat: matches 872. .606 of consensus"
     repeat_region   27749. .28049
                     /note="AluSp repeat: matches 302. .1 of consensus"
     repeat_region   28054. .28134
                     /note="L1MB6 repeat: matches 613. .532 of consensus"
     repeat_region   28454. .28758
                     /note="AluSg repeat: matches 300. .2 of consensus"
BASE COUNT     7441 a    6632 c    6511 g    8280 t       1 others
ORIGIN

  Query Match           6.7%;  Score 20;  DB 27;  Length 28865;
  Best Local Similarity 100.0%;  Pred. No. 3.01e-01;
  Matches   20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   3188 TGTATGTATGTATATATGTA 3207
         ||||||||||||||||||||
Qy    178 TGTATGTATGTATATATGTA 197


RESULT   35
LOCUS       PFSC03015   30101 bp    DNA             INV       04-DEC-1997
DEFINITION  Plasmodium falciparum DNA *** SEQUENCING IN PROGRESS *** from
            contig 3-15, complete sequence.
ACCESSION   Z98550
NID         g2665376
KEYWORDS    HTG.
SOURCE      malaria parasite.
  ORGANISM  Plasmodium falciparum
            Eukaryotae; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 30101)
  AUTHORS   Lawson,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-DEC-1997) P.falciparum Genome Sequencing Consortium,
            The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA, UK
COMMENT     On Dec 7, 1997 this sequence version replaced gi:2624342.
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and the release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc.
            For more information about this sequence or the Malaria Project,
            see http://www.sanger.ac.uk/Projects/P_falciparum.
FEATURES             Location/Qualifiers
     source          1. .30101
                     /organism="Plasmodium falciparum"
```



```
                     /db_xref="taxon:5833"
                     /chromosome="3"
                     /clone="3-15"
BASE COUNT    11832 a    2797 c    2879 g   12593 t
ORIGIN

  Query Match           6.7%;  Score 20;  DB 18;  Length 30101;
  Best Local Similarity 100.0%;  Pred. No. 3.01e-01;
  Matches   20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  27940 TGTATGTATGTATATATGTA 27959
          ||||||||||||||||||||
Qy    178 TGTATGTATGTATATATGTA 197


RESULT   36
LOCUS       PFSC03013   31162 bp    DNA             INV       16-FEB-1998
DEFINITION  Plasmodium falciparum DNA *** SEQUENCING IN PROGRESS *** from
            contig 3-13, complete sequence.
ACCESSION   Z98548
NID         g2894369
KEYWORDS    HTG.
SOURCE      malaria parasite.
  ORGANISM  Plasmodium falciparum
            Eukaryotae; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 31162)
  AUTHORS   Lawson,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-FEB-1998) P.falciparum Genome Sequencing Consortium,
            The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA, UK
COMMENT     On Feb 18, 1998 this sequence version replaced gi:2665374.
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and the release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc.
            For more information about this sequence or the Malaria Project,
            see http://www.sanger.ac.uk/Projects/P_falciparum.
FEATURES             Location/Qualifiers
     source          1. .31162
                     /organism="Plasmodium falciparum"
                     /db_xref="taxon:5833"
                     /chromosome="3"
                     /clone="3-13"
BASE COUNT    12311 a    2869 c    2946 g   13027 t       9 others
ORIGIN

  Query Match           6.7%;  Score 20;  DB 18;  Length 31162;
  Best Local Similarity 100.0%;  Pred. No. 3.01e-01;
```



```
  Matches   20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  29003 TGTATGTATGTATATATGTA 29022
          ||||||||||||||||||||
Qy    178 TGTATGTATGTATATATGTA 197


RESULT   37
LOCUS       AC004700    37056 bp    DNA             PRI       19-MAY-1998
DEFINITION  Homo sapiens chromosome 19, cosmid F14498, complete sequence.
ACCESSION   AC004700
NID         g3138891
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 37056)
  AUTHORS   Lamerdin,J.E., McCready,P.M., Skowronski,E., Adamson,A.W.,
            Burkhart-Schultz,K., Gordon,L., Kyle,A., Ramirez,M., Stilwagen,S.,
            Phan,H., Velasco,N., Do,L., Regala,W., Terry,A., Garnes,J.,
            Danganan,L., Poundstone,P., Christensen,M., Georgescu,A., Avila,J.,
            Liu,S., Attix,C., Andreise,T., Trankheim,M., Amico-Keller,G.,
            Coefield,J., Duarte,S., Lucas,S., Bruce,R., Thomas,P., Quan,G.,
            Kronmiller,B., Arellano,A., Montgomery,M., Ow,D., Nolan,M.,
            Trong,S., Kobayashi,A., Olsen,A.S. and Carrano,A.V.
  TITLE     Sequence analysis of a 2.5 Mb region in 19q13.2 containing a
            clustered CEA/PSG gene family
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 37056)
  AUTHORS   Lamerdin,J.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAY-1998) Joint Genome Institute, Lawrence Livermore
            National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
COMMENT     Map and sequence oriented from q centromere to telomere. Cosmid
            F14498 overlaps cosmid R26371 to the left and BAC 338531 to the
            right. Additional chr 19 map and sequence information may be
            obtained at: http://www-bio.llnl.gov/bbrp/genome/genome.html.
FEATURES             Location/Qualifiers
     source          1. .37056
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="F14498"
                     /chromosome="19"
                     /map="19q13.2 from BCKDHA-D19S217"
                     /cell_line="UV5HL9-5B"
                     /clone_lib="LL19NC02 F chromosome 19-specific cosmid
                     library"
                     /note="Cosmid library constructed at LLNL from flow-sorted
                     chromosomes from hybrid UV5HL9-5B, which carries
                     chromosome 19 as its only human chromosome."
```

```
     repeat_region   complement(2. .644)
                     /rpt_family="L1MB3"
     repeat_region   complement(1082. .1517)
                     /rpt_family="LOR1b"
     repeat_region   complement(2075. .2381)
                     /rpt_family="AluY"
     repeat_region   complement(3718. .3738)
                     /rpt_family="AT_rich"
     repeat_region   3840. .3947
                     /rpt_family="L1PA5"
     repeat_region   complement(4707. .4882)
                     /rpt_family="MER5B"
     repeat_region   complement(5531. .5609)
                     /rpt_family="(GGGA)n"
     repeat_region   complement(7712. .8184)
                     /rpt_family="LTR1"
     repeat_region   9211. .9613
                     /rpt_family="MER39b_j"
     repeat_region   complement(9614. .9919)
                     /rpt_family="AluJo"
     repeat_region   9920. .10138
                     /rpt_family="MER39b_j"
     repeat_region   10144. .10322
                     /rpt_family="AluJo"
     repeat_region   10323. .10360
                     /rpt_family="MER39b_j"
     repeat_region   10927. .11180
                     /rpt_family="LINE2"
     repeat_region   11798. .12342
                     /rpt_family="MER41A"
     repeat_region   12776. .12947
                     /rpt_family="MER41C"
     repeat_region   12985. .13178
                     /rpt_family="L1M4"
     repeat_region   13205. .13642
                     /rpt_family="L1M4"
     repeat_region   complement(13660. .14032)
                     /rpt_family="L1"
     repeat_region   complement(14044. .14683)
                     /rpt_family="L1"
     repeat_region   complement(14817. .15135)
                     /rpt_family="AluY"
     misc_feature    complement(16787. .16965)
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 1, quality: excellent, score: 93.000"
     repeat_region   16894. .17059
                     /rpt_family="MER20"
     repeat_region   complement(17803. .17870)
                     /rpt_family="AT_rich"
     repeat_region   complement(18012. .18123)
                     /rpt_family="(TA)n"
```



```
     repeat_region   complement(18124. .18309)
                     /rpt_family="(TA)n"
     repeat_region   complement(18414. .18525)
                     /rpt_family="(TA)n"
     repeat_region   complement(19040. .19323)
                     /rpt_family="MLT1B"
     repeat_region   complement(19325. .19538)
                     /rpt_family="MER30"
     repeat_region   complement(19539. .19652)
                     /rpt_family="MLT1B"
     repeat_region   complement(20255. .20301)
                     /rpt_family="(CA)n"
     repeat_region   complement(21081. .21625)
                     /rpt_family="MER41A"
     repeat_region   21937. .22174
                     /rpt_family="L1ME2"
     repeat_region   22289. .23120
                     /rpt_family="L1ME2"
     repeat_region   23170. .23682
                     /rpt_family="MLT2B"
     repeat_region   complement(23836. .23887)
                     /rpt_family="LINE2"
     repeat_region   complement(24885. .25001)
                     /rpt_family="(TAGA)n"
     misc_feature    complement(25816. .25889)
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 1, quality: excellent, score: 91.000"
     repeat_region   26089. .26502
                     /rpt_family="MSTB"
     repeat_region   complement(26817. .26919)
                     /rpt_family="(GAAAA)n"
     repeat_region   complement(28760. .28790)
                     /rpt_family="AT_rich"
     repeat_region   complement(29152. .29215)
                     /rpt_family="(TAAA)n"
     repeat_region   complement(29237. .29516)
                     /rpt_family="AluJo"
     repeat_region   29658. .30042
                     /rpt_family="MLT1B"
     misc_feature    complement(30564. .30731)
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 1, quality: good, score: 70.000"
     repeat_region   complement(32183. .32286)
                     /rpt_family="FLAM_A"
     misc_feature    complement(32235. .32375)
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 1, quality: good, score: 57.000"
     misc_feature    complement(32509. .32628)
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 2, quality: excellent, score: 95.000"
     repeat_region   complement(32895. .33052)
```



```
                     /rpt_family="L1ME"
     repeat_region   33235. .33351
                     /rpt_family="LINE2"
     repeat_region   33513. .33911
                     /rpt_family="MSTA"
     misc_feature    complement(33973. .34040)
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 0, quality: excellent, score: 79.000"
     misc_feature    complement(34878. .35022)
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 1, quality: good, score: 54.000"
     repeat_region   34906. .34946
                     /rpt_family="MIR"
     repeat_region   35519. .36286
                     /rpt_family="L1MB5"
     repeat_region   36291. .36589
                     /rpt_family="AluY"
     repeat_region   36606. .36655
                     /rpt_family="L1MB3"
BASE COUNT    10735 a    7589 c    7577 g   11155 t
ORIGIN

  Query Match           6.7%;  Score 20;  DB 26;  Length 37056;
  Best Local Similarity 100.0%;  Pred. No. 3.01e-01;
  Matches   20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  24906 TGTATGTATGTATATATGTA 24925
          ||||||||||||||||||||
Qy    178 TGTATGTATGTATATATGTA 197


RESULT   38
LOCUS       HSU75A4     39132 bp    DNA             PRI       10-APR-1997
DEFINITION  Human DNA sequence from cosmid U75A4 on chromosome X.
ACCESSION   Z82255
NID         g1934910
KEYWORDS    X.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 39132)
  AUTHORS   McLaren,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-APR-1997) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1RQ, UK. E-mail enquires: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Apr 13, 1997 this sequence version replaced gi:1781020.
            IMPORTANT:  This sequence is the entire insert of clone U75A4. The
            true left end of clone U75A4 is at 1 in this sequence. The true
```



```
            right end of clone U75A4 is at 39132.
            U75A4 is from the Lawrence Livermore National Laboratory
            flow-sorted
            X chromosome cosmid library LL0XNC01.
FEATURES    Location/Qualifiers
     source          1. .39132
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="X"
                     /clone="U75A4"
                     /clone_lib="LL0XNC01"
     repeat_region   1. .658
                     /note="MST-INTERNAL repeat: matches 667. .1 of consensus"
     repeat_region   661. .1023
                     /note="THE1C repeat: matches 369. .3 of consensus"
     repeat_region   1026. .1175
                     /note="L1MB3 repeat: matches 386. .535 of consensus"
     repeat_region   1281. .1653
                     /note="L1MA9 repeat: matches 529. .908 of consensus"
     repeat_region   3056. .3255
                     /note="L1 repeat: matches 2119. .2315 of consensus"
     repeat_region   3245. .4500
                     /note="L1 repeat: matches 2367. .3685 of consensus"
     repeat_region   4505. .5037
                     /note="MLT1D repeat: matches 505. .3 of consensus"
     repeat_region   5044. .5883
                     /note="L1 repeat: matches 3689. .4560 of consensus"
     repeat_region   6363. .7152
                     /note="L1MA8 repeat: matches 1035. .233 of consensus"
     repeat_region   7146. .7540
                     /note="MSTA repeat: matches 5. .425 of consensus"
     repeat_region   7534. .7772
                     /note="L1MB1 repeat: matches 249. .1 of consensus"
     repeat_region   7630. .8059
                     /note="L1 repeat: matches 5390. .4954 of consensus"
     repeat_region   8070. .8409
                     /note="L1ME1 repeat: matches 560. .910 of consensus"
     repeat_region   8420. .8573
                     /note="AluJo repeat: matches 290. .137 of consensus;
                     incomplete repeat"
     repeat_region   8651. .10906
                     /note="L1 repeat: matches 3073. .5390 of consensus"
     repeat_region   10759. .11773
                     /note="L1MA9 repeat: matches 1. .1017 of consensus"
     repeat_region   12034. .12433
                     /note="L1MD2 repeat: matches 571. .164 of consensus"
     repeat_region   12454. .12703
                     /note="MLT2CB repeat: matches 499. .270 of consensus"
     repeat_region   13195. .13641
                     /note="MLT2CB repeat: matches 44. .481 of consensus"
```

```
     repeat_region   13717. .13760
                     /note="MLT2E repeat: matches 349. .392 of consensus"
     repeat_region   14590. .14625
                     /note="9 copies of 4 mer 89 % conserved"
     repeat_region   17031. .17392
                     /note="THE1B repeat: matches 1. .362 of consensus"
     repeat_region   17395. .18913
                     /note="THE1B-INTERNAL repeat: matches 1. .1580 of
                     consensus"
     repeat_region   18914. .19276
                     /note="THE1B repeat: matches 1. .364 of consensus"
     repeat_region   19278. .19652
                     /note="L1 repeat: matches 1665. .2042 of consensus"
     repeat_region   19950. .22116
                     /note="L1 repeat: matches 2073. .4324 of consensus"
     repeat_region   22114. .22191
                     /note="L1PB1 repeat: matches 610. .528 of consensus"
     repeat_region   22249. .22461
                     /note="L1PB1 repeat: matches 681. .902 of consensus"
     repeat_region   22470. .23459
                     /note="L1 repeat: matches 4306. .5321 of consensus"
     repeat_region   23444. .23970
                     /note="MER25 repeat: matches 3. .550 of consensus"
     repeat_region   24151. .25153
                     /note="MER25 repeat: matches 581. .1783 of consensus"
     repeat_region   25359. .25648
                     /note="L1 repeat: matches 11. .299 of consensus"
     repeat_region   25679. .30512
                     /note="L1 repeat: matches 559. .5390 of consensus"
     repeat_region   30363. .30895
                     /note="L1PA2 repeat: matches 1. .532 of consensus"
     repeat_region   31466. .31874
                     /note="LTR7 repeat: matches 1. .412 of consensus"
     repeat_region   31891. .32262
                     /note="L1PA5 repeat: matches 517. .888 of consensus"
     repeat_region   32248. .32656
                     /note="MER25 repeat: matches 1722. .2133 of consensus"
     repeat_region   32477. .36385
                     /note="L1 repeat: matches 1446. .5362 of consensus"
     repeat_region   36339. .36807
                     /note="L1PA9 repeat: matches 440. .910 of consensus"
     repeat_region   36818. .36973
                     /note="MLT1A2 repeat: matches 152. .315 of consensus"
     repeat_region   36992. .37838
                     /note="L1PB1 repeat: matches 67. .902 of consensus"
     repeat_region   37853. .38430
                     /note="L1MA9 repeat: matches 66. .669 of consensus"
     repeat_region   38427. .38677
                     /note="L1MA8 repeat: matches 778. .1037 of consensus"
BASE COUNT    14797 a   7194 c   6999 g  10142 t
ORIGIN
```



```
  Query Match           6.7%;  Score 20;  DB 26;  Length 39132;
  Best Local Similarity 100.0%;  Pred. No. 3.01e-01;
  Matches   20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1240 ATATATACATACATACACAC 1259
           ||||||||||||||||||||
Cp    194 ATATATACATACATACACAC 175


RESULT   39
LOCUS       MMU35323    53419 bp    DNA             ROD       03-MAY-1996
DEFINITION  Mus musculus H2-M alpha chain (H2-Ma) gene, H2-M beta 2 chain
            (H2-Mb2) gene, H2-M beta 1 chain (H2-Mb1) gene, low molecular
            weight protein 2 Lmp2 (Lmp2) gene, complete cds.
ACCESSION   U35323
NID         g1209533
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;
            Murinae; Mus.
REFERENCE   1  (bases 1 to 53419)
  AUTHORS   Peleraux,A., Karlsson,L., Chambers,J. and Peterson,P.A.
  TITLE     Genomic organization of a mouse MHC class II region including the
            H2-M and Lmp2 loci
  JOURNAL   Immunogenetics 43 (4), 204-214 (1996)
  MEDLINE   96163020
REFERENCE   2  (bases 1 to 53419)
  AUTHORS   Peleraux,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-SEP-1995) Annick Peleraux, The R.W. Johnson
            Pharmaceutical Research Institute, Scripps Research Institute,
            10666 North Torrey Pines Rd., La Jolla, CA 92037, USA
FEATURES             Location/Qualifiers
     source          1. .53419
                     /organism="Mus musculus"
                     /strain="BALB/c"
                     /note="MHC class II, I region"
                     /db_xref="taxon:10090"
     repeat_unit     156. .283
                     /note="B1 repeat"
     gene            861. .4007
                     /gene="H2-Ma"
     CDS             join(861. .948,2300. .2584,3089. .3367,3593. .3721,4003. .4007)
                     /gene="H2-Ma"
                     /note="MHC class II alpha chain gene; Allele: d haplotype;
                     MHC class II alpha chain"
                     /codon_start=1
                     /product="H2-M alpha chain"
```



```
                     /db_xref="PID:g1209534"
                     /translation="MEHEQKSGAVLLRLLRLLWLLPHSWAVLEASTPVLWDDPQNHTF
                     RHTLPCQDGIPNIGLSETYDEDELFSFDFSQNTRVPRLPDFAEWAQGQGDASAIAFDR
                     SFCEMLWHEVSPKLECGQIPVSRGLPVAEVFTLKPLEFGKPNTLVCFISNLFPPTLTVN
                     WQLHSAPVEGASPTSISAVDGLTPQAFSYLNFTPEPPDLYSCTVTHEIDRYTAIAYWV
                     PQNALPSDLLENALCGVAPGLGVLGTIIGIVFFLCSQRPCSGD"
     repeat_unit     complement(5778. .5961)
                     /note="B2 repeat"
     repeat_unit     complement(6138. .6267)
                     /note="B1 repeat"
     repeat_unit     complement(9227. .9356)
                     /note="B1 repeat"
     gene            10879. .16559
                     /gene="H2-Mb2"
     CDS             join(10879. .10933,13108. .13389,14008. .14292,15964. .16080,
                     16248. .16283,16549. .16559)
                     /gene="H2-Mb2"
                     /note="MHC class II beta chain gene;  Allele: d haplotype;
                     MHC class II beta chain"
                     /codon_start=1
                     /product="H2-M beta 2 chain"
                     /db_xref="PID:g1209535"
                     /translation="MAALWLLLLVLSLHCMGAGGFVAHVESTCVLNDAGTPQDFTYCV
                     SFNKDLLACWDPDVGKIVPCEPGVLSRLAEIISNILNEQESLIHRLQNGLQDCATHTQ
                     PFWDVLTHRTRAPSVRVAQTTPFNTREPVMLACTVWGFYPADVTITWMKNGQLVPSHS
                     NKEKTAQPNGDWTYQTVSYLALTPSYGDVYTCVVQHSGTSEPIRGDWTPGLSPIQTVK
                     VSVSAATLGLGFIIFCVGPFRWRKSHSSSYTPLPGSTYPEGRH"
     repeat_unit     complement(12595. .12726)
                     /gene="H2-Mb2"
                     /note="B1 repeat"
     repeat_unit     14981. .15375
                     /gene="H2-Mb2"
                     /note="MT repeat"
     repeat_unit     complement(16707. .16838)
                     /note="B1 repeat"
     gene            19195. .25981
                     /gene="H2-Mb1"
     CDS             join(19195. .19249,21400. .21681,23194. .23478,25385. .25501,
                     25670. .25705,25971. .25981)
                     /gene="H2-Mb1"
                     /note="Description: MHC class II beta chain gene;  Allele:
                     d haplotype;  MHC class II beta chain"
                     /codon_start=1
                     /product="H2-M beta 1 chain"
                     /db_xref="PID:g1209536"
                     /translation="MAALWLLLLVLSLHCMGAGGFVAHVESTCVLDDAGTPQDFTYCV
                     SFNKDLLACWDPIVGKIVPCEPGVLYPLAENFSRILNKEESLLQRLQNGLPDCASHTQ
                     PFWNALTHRTRPPSVRVAQTTPFNTREPVMLACTVWGFYPADVTITWMKNGQLVPSHS
                     NKEKTAQPNGDWTYQTVSYLALTPSYGDVYTCVVQHSGTSEPIRGDWTPGLSPIQTVK
                     VSVSAATLGLGFIIFCVGPFRWRKSHSSSYTPLSGSTYPEGRH"
     repeat_unit     complement(20887. .21018)
```



```
                     /gene="H2-Mb1"
                     /note="B1 repeat"
     repeat_unit     complement(22505. .22676)
                     /gene="H2-Mb1"
                     /note="B2 repeat"
     repeat_unit     24405. .24797
                     /gene="H2-Mb1"
                     /note="MT repeat"
     repeat_unit     complement(28569. .28740)
                     /note="B2 repeat"
     repeat_unit     complement(29493. .29624)
                     /note="B1 repeat"
     repeat_unit     29643. .29774
                     /note="B1 repeat"
     repeat_unit     complement(29813. .29993)
                     /note="B2 repeat"
     repeat_unit     complement(31332. .31509)
                     /note="B2 repeat"
     repeat_unit     33565. .33694
                     /note="B1 repeat"
     repeat_unit     complement(37294. .37425)
                     /note="B1 repeat"
     repeat_unit     43604. .43736
                     /note="B1 repeat"
     repeat_unit     complement(43801. .43976)
                     /note="B2 repeat"
     repeat_unit     complement(44006. .44140)
                     /note="B1 repeat"
     repeat_unit     complement(44591. .44926)
                     /note="MT repeat"
     repeat_unit     47216. .47392
                     /note="B2 repeat"
     gene            complement(47866. .52957)
                     /gene="Lmp2"
     CDS             complement(join(47866. .47993,48730. .48871,49254. .49383,
                     49921. .50052,51378. .51445,52898. .52957))
                     /gene="Lmp2"
                     /note="proteasome subunit gene;  Allele: d haplotype;  MHC
                     encoded proteasome subunit"
                     /codon_start=1
                     /product="low molecular weight protein 2 Lmp2"
                     /db_xref="PID:g1209537"
                     /translation="MLRAGAPTAGSFRTEEVHTGTTIMAVEFDGGVVVGSDSRVSAGT
                     AVVNRVFDKLSPLHQRIFCALSGSAADAQAIADMAAYQLELHGLELEEPPLVLAAANV
                     VKNISYKYREDLLAHLIVAGWDQREGGQVYGTMGGMLIRQPFTIGGSGSSYIYGYVDA
                     AYKPGMTPEECRRFTTNAITLAMNRDGSSGGVIYLVTITAAGVDHRVILGDELPKFYD
                     E"
     repeat_unit     complement(49003. .49135)
                     /gene="Lmp2"
                     /note="B1 repeat"
     repeat_unit     50307. .50435
```

```
                     /gene="Lmp2"
                     /note="B1 repeat"
     repeat_unit     complement(50481. .50658)
                     /gene="Lmp2"
                     /note="B2 repeat"
     repeat_unit     50744. .50875
                     /gene="Lmp2"
                     /note="B1 repeat"
BASE COUNT    13151 a  12611 c  13314 g  14338 t     .  5 others
ORIGIN

  Query Match           6.7%;  Score 20;  DB 28;  Length 53419;
  Best Local Similarity 100.0%;  Pred. No. 3.01e-01;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   9117 TGTGTATGTATGTATATATG 9136
           ||||||||||||||||||||
Qy    176 TGTGTATGTATGTATATATG 195


RESULT   40
LOCUS       AC004300    58663 bp    DNA              INV       01-OCT-1998
DEFINITION  Drosophila melanogaster DNA sequence (P1 DS08209 (D186)), complete
            sequence.
ACCESSION   AC004300 AC003835
NID         g3687207
KEYWORDS    HTG.
SOURCE      Drosophila melanogaster (Subclones in sac from P1 clone DS08209
            (D186)) DNA.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea;
            Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 58663)
  AUTHORS   Celniker,S.E., George,R.A., Galle,R.F., Hoskins,R.A.,
            Svirskas,R.R., Harris,N.L., Agbayani,A., Arcaina,T.T., Baxter,E.,
            Blazej,R.G., Chavez,C., Chew,M., Doyle,C.M., Farfan,D.E.,
            Flanagan,J., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,S.H., Lee,B., Lomotan,M.A., Mak,J., Mazda,P., Moshrefi,A.R.,
            Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,E.,
            Snir,E., Twomey,B., Wan,K.H., Zhang,R., Zieran,L.L. and Rubin,G.M.
  TITLE     Sequencing of Drosophila chromosome 2R, region 57E3-57F1
  JOURNAL   Unpublished (1997)
REFERENCE   2  (bases 1 to 58663)
  AUTHORS   Celniker,S.E., George,R.A., Galle,R., Svirskas,R.R., Hoskins,R.A.,
            Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Chaves,C.,
            Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A.,
            Hummasti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B.,
            Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S., Moshrefi,A.R.,
            Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,E.,
            Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R.,
```



```
            Zieran,L.L. and Kimmel,B.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-1998) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Oct 1, 1998 this sequence version replaced gi:3659489.
            Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://fruitfly.berkeley.edu/sequence/) or send
            email to bdgp@fruitfly.berkeley.edu.
            Library location: 49-86.
FEATURES                Location/Qualifiers
     source             1. .58663
                        /organism="Drosophila melanogaster"
                        /db_xref="taxon:7227"
                        /chromosome="2R"
                        /map="57E3-57F1"
                        /clone="P1 DS08209 (D186)"
BASE COUNT    16416 a  13461 c  12978 g  15808 t
ORIGIN

  Query Match           6.7%;  Score 20;  DB 18;  Length 58663;
  Best Local Similarity 100.0%;  Pred. No. 3.01e-01;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  17790 TGTATGTATGTATATATGTA 17809
          ||||||||||||||||||||
Qy    178 TGTATGTATGTATATATGTA 197


RESULT   41
LOCUS       AF021335  . 62031 bp    DNA              ROD       01-OCT-1997
DEFINITION  Mus musculus T cell receptor gamma locus, TCR gamma 2 and gamma 4
            gene clusters.
ACCESSION   AF021335
NID         g2444469
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 62031)
  AUTHORS   Parlee,M.J., Zhou,Q. and Koop,B.F.
  TITLE     Genomic sequence and evolutionary analysis of the gamma 2 and gamma
            4 clusters of the mouse TCR gamma locus
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 62031)
  AUTHORS   Parlee,M.J., Zhou,Q. and Koop,B.F.
```



```
  TITLE     Direct Submission
  JOURNAL   Submitted (26-AUG-1997) Biology, University of Victoria, P.O. Box
            3020, Victoria, BC V8W 3N5, Canada
FEATURES                Location/Qualifiers
     source             1. .62031
                        /organism="Mus musculus"
                        /db_xref="taxon:10090"
     enhancer           complement(251. .558)
                        /standard_name="Gamma 2 Enhancer"
     repeat_region      complement(850. .1791)
                        /standard_name="Lx3"
                        /rpt_family="LINE"
     repeat_region      complement(1850. .2191)
                        /standard_name="L1"
                        /rpt_family="LINE"
     repeat_region      2745. .2939
                        /standard_name="L1MA6"
                        /rpt_family="LINE"
     CDS                complement(join(4814. .4956,5522. .5551,6975. .>7304))
                        /note="TCR C gamma 2"
                        /codon_start=3
                        /product="TRGC1S2"
                        /db_xref="PID:g2444470"
                        /translation="KKLDADISPKPTIFLPSVAETNLHKTGTYLCLLEKFFPDVIRVY
                        WKEKNGNTILDSQEGDTLKTKGTYMKFSWLTVPERAMGKEHSCIVKHENNKGGADQEI
                        FFPSIKKVATTCWQDKNDVLQFQFTSTSAYYTYLLLLLKSVIYLAIISFSLLRRTSVC
                        GNEKKS"
     repeat_region      complement(7360. .7458)
                        /standard_name="PB1D9"
                        /rpt_family="SINE"
     repeat_region      complement(9486. .9536)
                        /standard_name="B4"
                        /rpt_family="SINE"
     CDS                complement(<10998. .>11057)
                        /note="TCR J gamma 2"
                        /codon_start=3
                        /product="TRGJ1S2"
                        /db_xref="PID:g2444471"
                        /translation="SSGFHKVFAEGTKLIVIPS"
     repeat_region      11928. .12012
                        /standard_name="RSINE1"
                        /rpt_family="SINE"
     repeat_region      12243. .12685
                        /standard_name="LTRIS_MM"
                        /rpt_family="LTR/Retroviral"
     repeat_region      12862. .13045
                        /standard_name="HERVR"
                        /rpt_family="LTR/Retroviral"
     repeat_region      13114. .14782
                        /standard_name="HERVR"
                        /rpt_family="LTR/Retroviral"
```



```
     repeat_region      14784. .22407
                        /standard_name="L1"
                        /rpt_family="LINE"
     repeat_region      22417. .22685
                        /standard_name="HERVR"
                        /rpt_family="LTR/Retroviral"
     repeat_region      22725. .23120
                        /standard_name="HERVR"
                        /rpt_family="LTR/Retroviral"
     repeat_region      23659. .24437
                        /standard_name="HERVR"
                        /rpt_family="LTR/Retroviral"
     repeat_region      26830. .31500
                        /standard_name="L1_MM"
                        /rpt_family="LINE"
     repeat_region      complement(32368. .32528)
                        /standard_name="B4"
                        /rpt_family="SINE"
     repeat_region      complement(34731. .34841)
                        /standard_name="RMER5"
     repeat_region      complement(34847. .35089)
                        /standard_name="RLTRETN_MM"
                        /rpt_family="LTR/Retroviral"
     CDS                complement(join(<36330. .36636,36741. .36783))
                        /note="TCR V gamma 1.2"
                        /codon_start=1
                        /product="TRGV5S2"
                        /db_xref="PID:g2444472"
                        /translation="MLLLRWFTSCCLWVFGLGQLEQTELSVTRETDENVQISCIVYLP
                        YFSNTAIHWYRQKTNQQFEYLIYVATNYNQRPLGGKHKKIEASKDFKSSTSTLEINYL
                        KKEDEATYYCAVWM"
     repeat_region      37463. .37608
                        /standard_name="B1_MM"
                        /rpt_family="SINE"
     repeat_region      38478. .38640
                        /standard_name="L1M1"
                        /rpt_family="LINE"
     gene               38681. .38823
                        /note="V gamma pseudogene fragment"
                        /gene="TRGV5S4P"
                        /pseudo
     repeat_region      complement(38907. .39299)
                        /standard_name="L1MA5A"
                        /rpt_family="LINE"
     repeat_region      39471. .39556
                        /standard_name="MIR"
                        /rpt_family="SINE"
     repeat_region      complement(39630. .39658)
                        /standard_name="B1-F"
                        /rpt_family="SINE"
     CDS                join(39772. .39814,39907. .>40210)
```

```
                     /note="TCR V gamma 1.1"
                     /codon_start=1
                     /product="TRGV5S1"
                     /db_xref="PID:g2444473"
                     /translation="MLLLRWPTPCCLWVPGLGQLEQTELSVTRETDESAQISCIVSLP
                     YFSNTAIHWYRQKAKKPEYLIYVSTNYNQRPLGGKNKKIEASKDPQTSTSTLKINYLK
                     KEDEATYYCAVWI"
     CDS             <41908. .>41968
                     /note="TCR J gamma 4"
                     /codon_start=3
                     /product="TRGJ2"
                     /db_xref="PID:g2444474"
                     /translation="RHIMGGDICQRDKARSNSP"
     CDS             join(<44314. .44643,49291. .49344,50085. .50129,51956. .52098)
                     /note="TCR C gamma 4"
                     /codon_start=3
                     /product="TRGC2"
                     /db_xref="PID:g2444475"
                     /translation="KRTDSDFSPKPTIFLPSAAETNLHKAGTYLCLLEKFFPEVIRVY
                     WKEKDGEKILESQEGNTIKTNDRYMKFSWLTVTEDSMAKEHSCIVKHENNKRGVDQEI
                     LPPPIGKAFTTINVNPRDSVLRHENVNNATDLEDCMKGRKDMLQLQVTTTYAFYTYLI
                     LFFKSMVHLAFVVFCLFRRAAMSCDDQRS"
     repeat_region   complement(46831. .46906)
                     /standard_name="B1-F"
                     /rpt_family="LINE"
     repeat_region   47219. .47262
                     /standard_name="RSINE2"
                     /rpt_family="SINE"
     repeat_region   49751. .49819
                     /standard_name="PB1D9"
                     /rpt_family="SINE"
     repeat_region   complement(54224. .54340)
                     /standard_name="MIR"
                     /rpt_family="SINE"
     repeat_region   55036. .55167
                     /standard_name="B4"
                     /rpt_family="SINE"
     enhancer        55953. .56258
                     /standard_name="Gamma 4 enhancer"
                     /evidence=not_experimental
     repeat_region   complement(59211. .59305)
                     /standard_name="B2"
                     /rpt_family="SINE"
     repeat_region   complement(59397. .59583)
                     /standard_name="B2"
                     /rpt_family="SINE"
     repeat_region   complement(61130. .61194)
                     /standard_name="MER58A"
                     /rpt_family="DNA/MER1_type"
BASE COUNT    19882 a  12849 c  12337 g  16963 t
ORIGIN
```



```
  Query Match           6.7%;  Score 20;  DB 28;  Length 62031;
  Best Local Similarity 100.0%;  Pred. No. 3.0le-01;
  Matches   20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    802 TGTATGTATGTATATATGTA 821
          ||||||||||||||||||||
Qy    178 TGTATGTATGTATATATGTA 197


RESULT   42
LOCUS       AC004726     77612 bp    DNA             HTG       06-JUN-1998
DEFINITION  *** SEQUENCING IN PROGRESS *** DS06958; D292; Drosophila
            melanogaster; Chromosome 2L; Region 30A-; Library location 73_46;
            HTGS phase 1, 15 unordered pieces.
ACCESSION   AC004726
NID         g3192486
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea;
            Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 77612)
  AUTHORS   Celniker,S.E., George,R.A., Galle,R., Svirskas,R.R., Hoskins,R.A.,
            Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Chavez,C.,
            Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A.,
            Hummasti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B.,
            Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S., Moshrefi,A.R.,
            Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,E.,
            Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R.,
            Zieran,L.L. and Kimmel,B.E.
  TITLE     Sequencing of Drosophila melanogaster
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 77612)
  AUTHORS   Celniker,S.E., George,R.A., Galle,R., Svirskas,R.R., Hoskins,R.A.,
            Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Chavez,C.,
            Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A.,
            Hummasti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B.,
            Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S., Moshrefi,A.R.,
            Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,E.,
            Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R.,
            Zieran,L.L. and Kimmel,B.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAY-1998) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Jun 6, 1998 this sequence version replaced gi:3150299.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www-hgc.lbl.gov/sequence-archive.html) or
            send email to drosophila@mhgc.lbl.gov. All contigs in this
```



```
            submission meet the following cutoffs: length >= 400 bases, phrap
            computed error rate <= 1/10.
            ***                                                          ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                          ***
            * This sequence is unfinished. It consists of 15 contigs for
            * which the order is not known; their order in this record is
            * arbitrary. In some cases, the exact lengths of the gaps
            * between the contigs are also unknown; these gaps are  presented
            * as runs of N as a convenience only. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
            *       1      865: contig of 865 bp in length
            *     866      945: gap of unknown length
            *     946     1980: contig of 1035 bp in length
            *    1981     2060: gap of unknown length
            *    2061     3348: contig of 1288 bp in length
            *    3349     3428: gap of unknown length
            *    3429     5437: contig of 2009 bp in length
            *    5438     5517: gap of unknown length
            *    5518     7190: contig of 1673 bp in length
            *    7191     7270: gap of unknown length
            *    7271    10200: contig of 2930 bp in length
            *   10201    10280: gap of unknown length
            *   10281    12690: contig of 2410 bp in length
            *   12691    12770: gap of unknown length
            *   12771    16648: contig of 3878 bp in length
            *   16649    16728: gap of unknown length
            *   16729    20033: contig of 3305 bp in length
            *   20034    20113: gap of unknown length
            *   20114    25142: contig of 5029 bp in length
            *   25143    25222: gap of unknown length
            *   25223    32093: contig of 6871 bp in length
            *   32094    32173: gap of unknown length
            *   32174    39215: contig of 7042 bp in length
            *   39216    39295: gap of unknown length
            *   39296    48290: contig of 8995 bp in length
            *   48291    48370: gap of unknown length
            *   48371    61861: contig of 13491 bp in length
            *   61862    61941: gap of unknown length
            *   61942    77612: contig of 15671 bp in length.
FEATURES             Location/Qualifiers
     source          1. .77612
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="DS06958; D292"
                     /chromosome="2"
                     /clone_lib="73_46"
                     /map="30A-"
BASE COUNT    23247 a  16070 c  15373 g  21798 t   1124 others
ORIGIN
```



```
  Query Match           6.7%;  Score 20;  DB 17;  Length 77612;
  Best Local Similarity 100.0%;  Pred. No. 3.0le-01;
  Matches   20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  13520 TGTATGTATGTATATATGTA 13539
          ||||||||||||||||||||
Qy    178 TGTATGTATGTATATATGTA 197


RESULT   43
LOCUS       HS344F17     85832 bp    DNA             PRI       22-MAY-1997
DEFINITION  Human DNA sequence from PAC 344F17 on chromosome 6q22.1-6q22.33
            contains STS.
ACCESSION   Z95326
NID         g2117157
KEYWORDS    6q22.1-6q22.33.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 85832)
  AUTHORS   Palmer,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAY-1997) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquires: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On May 25, 1997 this sequence version replaced gi:2094786.
            IMPORTANT:  This sequence is the entire insert of clone 344F17.
            This sequence was generated from part of bacterial clone contigs of
            human chromosome 6, constructed by the Sanger Centre chromosome 6
            mapping group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6/
            This sequence has been finished according to sequence map criteria
            as follows.  An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu).  Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            The true left end of clone 344F17 is at 1 in this sequence. The
            true right end of clone 344F17 is at 85832.
            344F17 is from the library RPCI3 constructed at the Roswell Park
            Cancer
            Institute by the group of Pieter de Jong.
            For further details see http://bacpac.med.buffalo.edu/.
FEATURES             Location/Qualifiers
     source          1. .85832
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
```

```
                     /map="6q22.1-6q22.33"
                     /clone="344F17"
                     /clone_lib="RPCI-3"
     repeat_region   546. .729
                     /note="MIR repeat: matches 73. .260 of consensus"
     repeat_region   840. .935
                     /note="AluY repeat: matches 206. .301 of consensus;
                     incomplete repeat"
     repeat_region   2872. .3055
                     /note="MIR repeat: matches 66. .262 of consensus"
     repeat_region   8472. .8774
                     /note="AluJo repeat: matches 1. .302 of consensus"
     repeat_region   9412. .9443
                     /note="16 copies of 2 mer 91 % conserved"
     repeat_region   9557. .9856
                     /note="AluY repeat: matches 300. .1 of consensus"
     repeat_region   10713. .11882
                     /note="MER7B repeat: matches 1204. .39 of consensus"
     repeat_region   11907. .12222
                     /note="MLT1A1 repeat: matches 333. .5 of consensus"
     repeat_region   14194. .14311
                     /note="MIR repeat: matches 234. .113 of consensus"
     repeat_region   14449. .15052
                     /note="L1PA10 repeat: matches 301. .912 of consensus"
     repeat_region   15400. .15718
                     /note="L1MB8 repeat: matches 915. .597 of consensus"
     repeat_region   19454. .19572
                     /note="MIR2 repeat: matches 145. .25 of consensus"
     repeat_region   20444. .20699
                     /note="MIR repeat: matches 3. .262 of consensus"
     repeat_region   21963. .22261
                     /note="AluJb repeat: matches 297. .1 of consensus"
     repeat_region   22813. .22948
                     /note="MLT1G repeat: matches 510. .370 of consensus"
     repeat_region   23095. .23217
                     /note="MLT1G repeat: matches 178. .57 of consensus"
     repeat_region   23692. .23891
                     /note="L1MA5A repeat: matches 833. .1042 of consensus"
     repeat_region   24593. .24834
                     /note="MIR repeat: matches 13. .262 of consensus"
     repeat_region   25726. .26110
                     /note="MSTD repeat: matches 1. .385 of consensus"
     repeat_region   26714. .26877
                     /note="MER4A2 repeat: matches 464. .308 of consensus"
     repeat_region   26878. .27114
                     /note="L1PA6 repeat: matches 648. .890 of consensus"
     repeat_region   27132. .27230
                     /note="MER4B repeat: matches 445. .347 of consensus"
     repeat_region   27443. .27669
                     /note="MIR repeat: matches 262. .28 of consensus"
     repeat_region   27831. .28375
```



```
                     /note="MLT2D repeat: matches 553. .1 of consensus"
     repeat_region   29384. .29685
                     /note="AluY repeat: matches 1. .301 of consensus"
     repeat_region   32329. .32984
                     /note="L1ME1 repeat: matches 252. .906 of consensus"
     repeat_region   33401. .33708
                     /note="AluSx repeat: matches 302. .1 of consensus"
     repeat_region   33838. .33895
                     /note="29 copies of 2 mer 85 % conserved"
     repeat_region   34951. .35083
                     /note="MIR repeat: matches 105. .248 of consensus"
     repeat_region   37320. .37367
                     /note="2 copies of 24 mer 96 % conserved"
     repeat_region   37741. .38198
                     /note="MER7B repeat: matches 1. .439 of consensus"
     repeat_region   38237. .38348
                     /note="MER7B repeat: matches 1094. .1205 of consensus"
     repeat_region   38724. .39045
                     /note="MER44A repeat: matches 2. .323 of consensus"
     repeat_region   40883. .41186
                     /note="AluJb repeat: matches 302. .1 of consensus"
     repeat_region   41482. .41775
                     /note="AluJo repeat: matches 300. .1 of consensus"
     repeat_region   45411. .45452
                     /note="MIR repeat: matches 92. .51 of consensus"
     repeat_region   46824. .46952
                     /note="AluJo repeat: matches 253. .123 of consensus;
                     incomplete repeat"
     repeat_region   49584. .50088
                     /note="MLT2CB repeat: matches 501. .1 of consensus"
     repeat_region   52649. .52954
                     /note="AluJb repeat: matches 1. .302 of consensus"
     repeat_region   52965. .53020
                     /note="14 copies of 4 mer 93 % conserved"
     repeat_region   53029. .53168
                     /note="AluSq repeat: matches 1. .135 of consensus;
                     incomplete repeat"
     repeat_region   53807. .53937
                     /note="L1PA9 repeat: matches 780. .908 of consensus"
     repeat_region   53836. .53939
                     /note="L1PA4 repeat: matches 790. .893 of consensus"
     repeat_region   57091. .57173
                     /note="MIR2 repeat: matches 143. .61 of consensus"
     repeat_region   58713. .59015
                     /note="AluSq repeat: matches 1. .303 of consensus"
     repeat_region   59105. .59339
                     /note="L1MB2 repeat: matches 566. .326 of consensus"
     repeat_region   59357. .59655
                     /note="AluJo repeat: matches 302. .1 of consensus"
     repeat_region   59656. .59958
                     /note="L1MB6 repeat: matches 303. .5 of consensus"
```



```
     repeat_region   59807. .60067
                     /note="L1 repeat: matches 5390. .5137 of consensus"
     repeat_region   66880. .67187
                     /note="AluSx repeat: matches 2. .302 of consensus"
     repeat_region   67198. .67435
                     /note="L1 repeat: matches 5078. .4844 of consensus"
     repeat_region   67445. .67576
                     /note="L1MA7 repeat: matches 880. .1011 of consensus"
     repeat_region   70591. .70650
                     /note="15 copies of 4 mer 98 % conserved"
     repeat_region   70666. .70884
                     /note="MER31 repeat: matches 313. .74 of consensus"
     repeat_region   71332. .71581
                     /note="AluJo repeat: matches 299. .86 of consensus;
                     incomplete repeat"
     repeat_region   72054. .72656
                     /note="L1MB7 repeat: matches 920. .303 of consensus"
     repeat_region   72661. .72960
                     /note="AluJo repeat: matches 4. .301 of consensus"
     repeat_region   74276. .74409
                     /note="MIR2 repeat: matches 146. .14 of consensus"
     repeat_region   75106. .75222
                     /note="L1ME3A repeat: matches 905. .771 of consensus"
     repeat_region   75724. .76635
                     /note="L1MB5 repeat: matches 30. .915 of consensus"
     misc_feature    77277. .77579
                     /note="match: STS G02927"
     repeat_region   78519. .78590
                     /note="18 copies of 4 mer 82 % conserved"
     repeat_region   79182. .79308
                     /note="MER34 repeat: matches 533. .418 of consensus"
     repeat_region   79260. .79336
                     /note="MER39 repeat: matches 516. .435 of consensus"
     repeat_region   79467. .79820
                     /note="MER39 repeat: matches 380. .13 of consensus"
     repeat_region   81314. .81689
                     /note="MLT1B repeat: matches 1. .388 of consensus"
     repeat_region   82854. .83143
                     /note="AluSq repeat: matches 2. .303 of consensus"
     repeat_region   83173. .83248
                     /note="19 copies of 4 mer 99 % conserved"
     repeat_region   83805. .83928
                     /note="L1ME3A repeat: matches 243. .375 of consensus"
     repeat_region   83966. .84055
                     /note="L1MA8 repeat: matches 441. .530 of consensus"
     repeat_region   83966. .84138
                     /note="L1MA5 repeat: matches 441. .631 of consensus"
BASE COUNT    26981 a  16526 c  15977 g  26348 t
ORIGIN

  Query Match           6.7%;  Score 20;  DB 26;  Length 85832;
```



```
  Best Local Similarity 100.0%;  Pred. No. 3.01e-01;
  Matches   20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  70598 TACATATATACATACATACA 70617
          ||||||||||||||||||||
Cp    197 TACATATATACATACATACA 178


RESULT   44
LOCUS       AC005250    94336 bp    DNA             PRI       03-JUL-1998
DEFINITION  Homo sapiens BAC clone RG318M05 from 7q22-q31.1, complete sequence.
ACCESSION   AC005250
NID         g3287719
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 94336)
  AUTHORS   Courtne,L., Nelson,J., Wohldmann,P., Bauer,C. and Morris,M.
  TITLE     The sequence of Homo sapiens BAC clone RG318M05
  JOURNAL   Unpublished (1998)
REFERENCE   2  (bases 1 to 94336)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUL-1998) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     SUBMITTED BY:  WUGSC
            Genome Sequencing Center
            Department of Genetics
            Washington University
            St. Louis MO 63108, USA
            http://genome.wustl.edu/gsc
            mailto:sapiens@watson.wustl.edu

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded or sequenced with an alternate
            chemistry; an attempt was made to resolve all sequencing problems,
            such as compressions and repeats; all regions were covered by
            sequence from more than one subclone; and the assembly was
            confirmed by restriction digest.

            MAPPING INFORMATION:
            The sequence of this clone was established as part of a mapping and
            sequencing collaboration between the NHGRI Chromosome 7 Mapping
            Project (Eric D. Green, Director), John D. McPherson in the
```

Department of Genetics (Washington University), and the Washington
University Genome Sequencing Center. For additional information
about the map position of this sequence, see
http://www.nhgri.nih.gov/DIR/GTB/CHR7 , send
mailto:egreen@nhgri.nih.gov , or see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
Clone RG318M05 is from a release of the human BAC library
CITB-HS-A. The library contains cloned DNA from human sperm. See:
Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J.
Kim et al., Genomics 34:213-8 (1996). The clone is available from
Research Genetics, Inc. (http://www.resgen.com).
VECTOR: pBeloBAC11
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:
The actual start of this clone is at base position 1 of RG318M05;
actual end is at base position 94336 of RG318M05.

FEATURES Location/Qualifiers
source 1. .94336
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q22-q31.1"
/clone_lib="CITB-HS-A"
/clone="RG318M05"
repeat_region 1773. .2057
/rpt_family="Alu"
repeat_region 2369. .2480
/rpt_family="(GGAA)n"
STS complement(2377. .2568)
/db_xref="GI:1113400"
repeat_region 2901. .2940
/rpt_family="MIR"
repeat_region 2941. .3042
/rpt_family="U6"
repeat_region 3043. .3311
/rpt_family="L1"
repeat_region 3944. .4244
/rpt_family="Alu"
misc_feature 5791. .6006
/note="match to EST N53637 (NID:g1194803) yz04f07.r1"
repeat_region 5898. .6176
/rpt_family="Alu"
repeat_region 6443. .6717
/rpt_family="MIR"
repeat_region 6795. .6846
/rpt_family="Alu"
repeat_region 6847. .7842
/rpt_family="Retroviral"
repeat_region 7843. .8080

/rpt_family="Alu"
repeat_region 9096. .9231
/rpt_family="MIR"
repeat_region 11498. .11794
/rpt_family="Alu"
repeat_region 11829. .12130
/rpt_family="Alu"
repeat_region 12955. .13739
/rpt_family="MER21_g"
repeat_region 14302. .14460
/rpt_family="MIR"
repeat_region 14673. .14968
/rpt_family="Alu"
repeat_region 15263. .15566
/rpt_family="MER1_type"
repeat_region 15861. .16164
/rpt_family="Alu"
repeat_region 16611. .16911
/rpt_family="Alu"
repeat_region 17048. .17348
/rpt_family="Alu"
repeat_region 18240. .18329
/rpt_family="L1"
repeat_region 20655. .21668
/rpt_family="L1"
repeat_region 21694. .22861
/rpt_family="L1"
misc_feature complement(23815. .24220)
/note="match to EST H66826 (NID:g1025566) yr85h09.s1"
misc_feature complement(23967. .24188)
/note="match to EST AA676859 (NID:g2657381) zj69c12.s1"
repeat_region 24989. .25345
/rpt_family="L1"
repeat_region 25332. .25787
/rpt_family="L1"
repeat_region 25788. .25911
/rpt_family="7SK"
repeat_region 25951. .26018
/rpt_family="7SK"
repeat_region 26019. .26912
/rpt_family="L1"
repeat_region 26915. .27352
/rpt_family="L1"
repeat_region 29116. .29263
/rpt_family="L1"
repeat_region 29353. .29478
/rpt_family="L1"
repeat_region 29568. .29928
/rpt_family="L1"
repeat_region 30036. .30334
/rpt_family="Alu"

repeat_region 30399. .30684
/rpt_family="Alu"
repeat_region 31015. .31399
/rpt_family="L2"
repeat_region 32068. .32635
/rpt_family="L2"
repeat_region 32707. .33004
/rpt_family="Alu"
repeat_region 33095. .33290
/rpt_family="MIR"
repeat_region 33334. .33469
/rpt_family="L2"
repeat_region 33878. .34179
/rpt_family="Alu"
repeat_region 34183. .34291
/rpt_family="Alu"
repeat_region 35145. .35274
/rpt_family="L2"
repeat_region 35408. .35715
/rpt_family="Alu"
repeat_region 35727. .35922
/rpt_family="L2"
repeat_region 36037. .36137
/rpt_family="Alu"
repeat_region 36352. .36987
/rpt_family="L1"
repeat_region 37004. .37302
/rpt_family="Alu"
repeat_region 37304. .38183
/rpt_family="L1"
repeat_region 38918. .39062
/rpt_family="L2"
repeat_region 39104. .39259
/rpt_family="Alu"
repeat_region 39722. .40028
/rpt_family="Alu"
repeat_region 40051. .40367
/rpt_family="L2"
repeat_region 40701. .40746
/rpt_family="L2"
repeat_region 41558. .42083
/rpt_family="L2"
repeat_region 42283. .42396
/rpt_family="L2"
repeat_region 42890. .43202
/rpt_family="Alu"
STS complement(44972. .45361)
/db_xref="GI:1113728"
repeat_region 47958. .48034
/rpt_family="MER1_type"
repeat_region 48044. .48340

/rpt_family="Alu"
repeat_region 49350. .49400
/rpt_family="MER86"
repeat_region 49500. .49792
/rpt_family="Alu"
...
Note: remainder of annotations omitted.

Query Match 6.7%; Score 20; DB 26; Length 94336;
Best Local Similarity 100.0%; Pred. No. 3.01e-01;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 55737 ATATATACATACATACACAC 55756
||||||||||||||||||||
Cp 194 ATATATACATACATACACAC 175

RESULT 45
LOCUS AC005207 99411 bp DNA PRI 24-JUL-1998
DEFINITION Homo sapiens chromosome 17, clone hRPC.1081_P_3, complete sequence.
ACCESSION AC005207
NID g3342221
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 99411)
AUTHORS Birren,B., Fasman,K., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 17, clone hRPC.1081_P_3
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 99411)
AUTHORS Birren,B., Fasman,K., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boatin,C.,
Boutwell,C., Brown,A., Castle,A., Cerny,J., Cooke,P., Depayre,E.,
Devon,K., Dewar,K., Donelan,L., Etemadi,S., Ferreira,P.,
FitzHugh,W., Forrest,C., Funke,R., Gage,D., Gardyna,S.,
Gensheimer,S., Geraigery,K., Gilmartin,T., Grant,G., Hagos,B.,
Harris,K., Horton,L., Howland,J.C., Hui,L., Jacotot,L., Kann,L.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
Nachman,A., Nahf,R., Naylor,J., O'Connor,T., Pavlin,B.,
Peterson,K., Riley,R., Roberts,D., Rossello,R., Roy,A., Shyam,R.,
Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C.,
Strickland,C., Subramanian,A., Torruella-Miller,I., Vassiliev,H.,
Vo,A., Wagner,A., Wang,B., Wheeler,J., Wu,Y., Ye,W.J., Zhao,J. and
Zody,M.
TITLE Direct Submission
JOURNAL Submitted (30-JUN-1998) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 99411)

```
  AUTHORS   Birren,B., Fasman,K., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boatin,C.,
            Boutwell,C., Brown,A., Castle,A., Cerny,J., Cooke,P., Depayre,E.,
            Devon,K., Dewar,K., Donelan,L., Etemadi,S., Ferreira,P.,
            FitzHugh,W., Forrest,C., Funke,R., Gage,D., Gardyna,S.,
            Gensheimer,S., Geraigery,K., Gilmartin,T., Grant,G., Hagos,B.,
            Harris,K., Horton,L., Howland,J.C., Jacotot,L., Kann,L.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Nachman,A., Nahf,R., Naylor,J., Niloff,M., O'Connor,T., Pavlin,B.,
            Peterson,K., Riley,R., Roberts,D., Rossello,R., Roy,A., Shyam,R.,
            Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C.,
            Subramanian,A., Torruella-Miller,I., Vassiliev,H., Vo,A.,
            Wagner,A., Wang,B., Wheeler,J., Wu,Y., Ye,W.J., Zhao,J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUL-1998) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 27, 1998 this sequence version replaced gi:3337315.
            All repeats were identified using RepeatMasker: Smit, A.F.A. &
            Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html.
FEATURES             Location/Qualifiers
     source          1. .99411
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="hRPC.1081_P_3"
                     /clone_lib="Peter de Jong/ human PAC library"
                     /map="17"
                     /chromosome="17"
     repeat_region   complement(1. .61)
                     /rpt_family="AluS"
     repeat_region   complement(80. .117)
                     /rpt_family="L1MA1"
     repeat_region   complement(111. .379)
                     /rpt_family="L1MCa"
     repeat_region   380. .688
                     /rpt_family="AluYa8"
     repeat_region   complement(689. .1491)
                     /rpt_family="L1MCa"
     repeat_region   complement(1492. .1797)
                     /rpt_family="AluSg"
     repeat_region   complement(1798. .1855)
                     /rpt_family="L1MCa"
     repeat_region   complement(1856. .2161)
                     /rpt_family="AluSq"
     repeat_region   complement(2162. .2696)
                     /rpt_family="L1MCa"
     repeat_region   complement(2697. .2966)
                     /rpt_family="AluJo"
     repeat_region   complement(2967. .3337)
                     /rpt_family="L1MCa"
```



```
     repeat_region   complement(3338. .3593)
                     /rpt_family="AluSq"
     repeat_region   complement(3594. .4742)
                     /rpt_family="L1MCa"
     repeat_region   complement(5101. .5249)
                     /rpt_family="L2"
     repeat_region   complement(5285. .5740)
                     /rpt_family="LTR10A"
     repeat_region   complement(5755. .5979)
                     /rpt_family="L2"
     repeat_region   complement(6011. .6720)
                     /rpt_family="MER21B"
     repeat_region   complement(6872. .7541)
                     /rpt_family="LTR8"
     repeat_region   complement(7575. .7988)
                     /rpt_family="L1MD3"
     repeat_region   8046. .8137
                     /rpt_family="L1MC4"
     repeat_region   complement(8468. .9016)
                     /rpt_family="MER41A"
     repeat_region   9272. .9364
                     /rpt_family="MIR"
     repeat_region   complement(9434. .9559)
                     /rpt_family="FLAM_A"
     repeat_region   9650. .9728
                     /rpt_family="MIR"
     repeat_region   9901. .10062
                     /rpt_family="AluJb"
     repeat_region   10355. .10655
                     /rpt_family="AluSx"
     repeat_region   11968. .12252
                     /rpt_family="AluSx"
     repeat_region   complement(12778. .12835)
                     /rpt_family="L2"
     repeat_region   complement(12879. .12925)
                     /rpt_family="L1MD3"
     repeat_region   12926. .13231
                     /rpt_family="AluSp"
     repeat_region   complement(13232. .13243)
                     /rpt_family="L1MD3"
     repeat_region   complement(13304. .13576)
                     /rpt_family="L2"
     repeat_region   13647. .13974
                     /rpt_family="L2"
     repeat_region   complement(13994. .14823)
                     /rpt_family="L2"
     repeat_region   14824. .14940
                     /rpt_family="AluSg"
     repeat_region   14941. .14971
                     /rpt_family="(CAAA)n"
     repeat_region   14972. .15104
```



```
                     /rpt_family="AluSg"
     repeat_region   complement(15105. .15180)
                     /rpt_family="L2"
     repeat_region   15181. .15487
                     /rpt_family="AluSg"
     repeat_region   complement(15488. .15585)
                     /rpt_family="L2"
     repeat_region   complement(15627. .15726)
                     /rpt_family="L2"
     repeat_region   complement(15727. .15975)
                     /rpt_family="AluJo"
     repeat_region   complement(15980. .16036)
                     /rpt_family="L1PA10"
     repeat_region   16040. .16060
                     /rpt_family="AT_rich"
     repeat_region   complement(16109. .16344)
                     /rpt_family="AluSg"
     unsure          complement(16286. .16350)
                     /note="Single-stranded terminator coverage."
     repeat_region   16425. .16736
                     /rpt_family="AluSx"
     repeat_region   16761. .16884
                     /rpt_family="AluJo"
     repeat_region   16885. .17185
                     /rpt_family="AluSx"
     repeat_region   17186. .17332
                     /rpt_family="AluJo"
     repeat_region   complement(17333. .17638)
                     /rpt_family="AluSx"
     repeat_region   17639. .17668
                     /rpt_family="AluJo"
     repeat_region   17669. .17745
                     /rpt_family="purine-rich"
     unsure          complement(17697. .17837)
                     /note="Single-stranded terminator coverage."
     repeat_region   17801. .17955
                     /rpt_family="AluJo"
     repeat_region   complement(17993. .18847)
                     /rpt_family="L1PA13"
     unsure          complement(18083. .18139)
                     /note="Single-stranded terminator coverage."
     repeat_region   complement(18848. .19141)
                     /rpt_family="AluY"
     repeat_region   complement(19142. .19245)
                     /rpt_family="L1PA13"
     repeat_region   complement(19339. .19407)
                     /rpt_family="L1MC/D"
     repeat_region   19478. .19770
                     /rpt_family="AluSx"
     repeat_region   19862. .19886
                     /rpt_family="AT_rich"
```



```
     repeat_region   complement(19915. .19939)
                     /rpt_family="(CA)n"
     repeat_region   complement(20542. .20848)
                     /rpt_family="AluY"
     unsure          complement(20963. .21003)
                     /note="Single-stranded coverage."
     repeat_region   21047. .21195
                     /rpt_family="AluJo"
     repeat_region   21196. .21495
                     /rpt_family="AluJb"
     repeat_region   21496. .21666
                     /rpt_family="AluJo"
     repeat_region   21778. .21814
                     /rpt_family="AT_rich"
...
Note: remainder of annotations omitted.

  Query Match           6.7%;  Score 20;  DB 26;  Length 99411;
  Best Local Similarity 100.0%;  Pred. No. 3.01e-01;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  62085 GTGTGTATGTATGTATATAT 62104
          ||||||||||||||||||||
Qy    175 GTGTGTATGTATGTATATAT 194


RESULT   46
LOCUS       AP000030   100000 bp    DNA             PRI       13-MAY-1998
DEFINITION  Homo sapiens genomic DNA, chromosome 21q11.1, segment 1/28,
            complete sequence.
ACCESSION   AP000030
NID         g3132340
KEYWORDS    HTG.
SOURCE      Homo sapiens DNA, clone:T1040-119B8f_1.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 100000)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Shiba,T. and Sakaki,Y.
  TITLE     Homo sapiens genomic DNA, chromosome 21q
  JOURNAL   Published Only in DataBase (1998) In press
REFERENCE   2  (bases 1 to 100000)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Shiba,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-MAY-1998) to the DDBJ/EMBL/GenBank databases.
            Masahira Hattori, Kitasato University, Department of Science, JST
            Sequencing Laboratory; Kitasato 1-15-1, Sagamihara 228, Japan
            (E-mail:hattori@hgc.ims.u-tokyo.ac.jp, Tel:0427-78-9732,
            Fax:0427-78-9561)
FEATURES             Location/Qualifiers
     source          1. .100000
```

```
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /clone="T1040-119B8f_1"
                     /map="21q22.1"
BASE COUNT    29226 a  23198 c  23040 g  24536 t
ORIGIN

  Query Match           6.7%;  Score 20;  DB 26;  Length 100000;
  Best Local Similarity  90.9%;  Pred. No. 3.01e-01;
  Matches    20;  Conservative    0;  Mismatches   2;  Indels   0;  Gaps   0;

Db  25687 AATACCTGCTCCTCAAATGCAG 25708
          || |||||||| ||||||||||
Cp    279 AANACCTGCTCNTCAAATGCAG 258


RESULT   47
LOCUS       AF037352   105726 bp    DNA             ROD       29-JAN-1998
DEFINITION  Mus musculus T cell receptor gamma locus, TCR gamma 1 and gamma 3
            gene clusters.
ACCESSION   AF037352
NID         g2815645
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 105726)
  AUTHORS   Zhou,Q., Parlee,M. and Koop,B.
  TITLE     Organization, structure and evolutionary implication of 157kb of
            DNA of the murine T-cell receptor gamma locus
  JOURNAL   Unpublished (1997)
REFERENCE   2  (bases 1 to 105726)
  AUTHORS   Zhou,Q.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-1997) Center for Environmental Health, Department
            of Biology, University of Victoria, PO Box 3020, Victoria, B.C. V8W
            3N5, Canada
FEATURES             Location/Qualifiers
     source          1. .105726
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
     repeat_region   complement(132. .353)
                     /note="B3A"
                     /rpt_family="SINE/B2"
     repeat_region   complement(399. .505)
                     /note="PB1"
                     /rpt_family="SINE/Alu"
     gene            1311. .>1761
                     /gene="TCRGV4S1"
```



```
     CDS             join(1311. .1353,1454. .>1761)
                     /gene="TCRGV4S1"
                     /codon_start=1
                     /product="TCR V gamma 5"
                     /db_xref="PID:g2815646"
                     /translation="MLWALALLLAFLPAGRQTSSNLEERIMSITKLEGSSAIMTCDTH
                     RTGTYIHWYRPQKGRAPEHLLYYNFVSSTTVVDSRFNLEKYHVYEGPDKRYKFVLRNV
                     EESDSALYYCASWAG"
     repeat_region   complement(2014. .2488)
                     /note="L1_MM"
                     /rpt_family="LINE/L1"
     repeat_region   4767. .4835
                     /note="MER20"
                     /rpt_family="DNA/MER1_type"
     gene            8236. .8759
                     /gene="TCRGV3S1"
     CDS             join(8236. .8347,8455. .8759)
                     /gene="TCRGV3S1"
                     /codon_start=1
                     /product="TCR V gamma 2"
                     /db_xref="PID:g2815647"
                     /translation="MKNPGSQALLPLYLPWEANLADENPLLKVVIFLCLLTPGHGKLE
                     QPEISISRPRDETAQISCKVFIESFRSVTIHWYRQKPNQGLEFLLYVLATPTHIFLDK
                     ETKKMEASKNPSASTSILTIYSLEEEDEAIYYCSYG"
     repeat_region   11312. .11423
                     /note="B4"
                     /rpt_family="SINE/B4"
     gene            13613. .>14174
                     /gene="TCRGV2S1"
     CDS             join(13613. .13706,13863. .>14174)
                     /gene="TCRGV2S1"
                     /codon_start=1
                     /product="TCR V gamma 4"
                     /db_xref="PID:g2815648"
                     /translation="MGLLLQVFTLASLRICKCMAFRGIEDTGKRAKGSSLTSPLGSYV
                     IKRKGNTAFLKCQIKTSVQKPDAYIHWYQEKPGQRLQRMLCSSSKENIVYEKDFSDER
                     YEARTWQSDLSSVLTIHQVTEEDTGTYYCACWD"
     gene            15523. .>15968
                     /gene="TCRGV1S1"
     CDS             join(15523. .15565,15673. .>15968)
                     /gene="TCRGV1S1"
                     /codon_start=1
                     /product="TCR V gamma 3"
                     /db_xref="PID:g2815649"
                     /translation="MSTSWLFLLSLTCVYGDSWISQDQLSFTRRPNKTVHISCKLSGV
                     PLHNTIVHWYQLKEGEPLRRIFYGSVKTYKQDKSHSRLEIDEKDDGTFYLIINNVVTS
                     DEATYYCACWD"
     repeat_region   complement(16132. .16172)
                     /note="MIR"
                     /rpt_family="SINE/MIR"
     repeat_region   complement(17129. .17175)
```



```
                     /note="B3A"
                     /rpt_family="SINE/B2"
     repeat_region   complement(19570. .19715)
                     /note="RSINE1"
                     /rpt_family="SINE/B4"
     repeat_region   complement(20306. .20836)
                     /note="L1_MM"
                     /rpt_family="LINE/L1"
     repeat_region   complement(20866. .20964)
                     /note="L1_MM"
                     /rpt_family="LINE/L1"
     repeat_region   20965. .21298
                     /note="RLTRETN_MM"
                     /rpt_family="LTR/retroviral"
     repeat_region   21300. .21521
                     /note="MMETN"
                     /rpt_family="LTR/retroviral"
     repeat_region   complement(21530. .21970)
                     /note="L1_MM"
                     /rpt_family="LINE/L1"
     repeat_region   21971. .22343
                     /note="L1_MM"
                     /rpt_family="LINE/L1"
     repeat_region   22479. .23557
                     /note="MMETN"
                     /rpt_family="LTR/retroviral"
     repeat_region   23572. .23690
                     /note="MMETN"
                     /rpt_family="LTR/retroviral"
     repeat_region   23847. .27007
                     /note="MMETN"
                     /rpt_family="LTR/retroviral"
     repeat_region   27008. .27349
                     /note="RLTRETN_MM"
                     /rpt_family="LTR/retroviral"
     repeat_region   complement(27350. .27755)
                     /note="L1_MM"
                     /rpt_family="LINE/L1"
     repeat_region   complement(28035. .28118)
                     /note="PB1D10"
                     /rpt_family="SINE/Alu"
     repeat_region   complement(28158. .28463)
                     /note="L1MA5A"
                     /rpt_family="LINE/L1"
     repeat_region   complement(28587. .28997)
                     /note="L1MA5A"
                     /rpt_family="LINE/L1"
     repeat_region   complement(29447. .29906)
                     /note="L1MA5A"
                     /rpt_family="LINE/L1"
     repeat_region   complement(29913. .30113)
```



```
                     /note="B3"
                     /rpt_family="SINE/B2"
     repeat_region   complement(30114. .30242)
                     /note="L1"
                     /rpt_family="LINE/L1"
     repeat_region   complement(30908. .31751)
                     /note="Lx9"
                     /rpt_family="LINE/L1"
     repeat_region   complement(32599. .32750)
                     /note="RSINE1"
                     /rpt_family="SINE/B4"
     CDS             <33598. .>33658
                     /codon_start=1
                     /product="TCR J gamma 1"
                     /db_xref="PID:g2815650"
                     /translation="DSSGFHKVFAEGTKLIVIPS"
     repeat_region   37204. .37302
                     /note="PB1D7"
                     /rpt_family="SINE/Alu"
     CDS             join(<37357. .37607,39119. .39163,39731. .39873)
                     /codon_start=1
                     /product="TCR C gamma 1"
                     /db_xref="PID:g2815651"
                     /translation="DKRLDADISPKPTIFLPSVAETNLHKTGTYLCLLEKFFPDVIRV
                     YWKEKDGNTILDSQEGDTLKTNDTYMKFSWLTVPERAMGKEHRCIVKHENNKGGADQE
                     IFFPSIKKVAVSTRPTTCWQDKNDVLQLQFTITSAIYTYLLLLLKSVIYLAIISFSLL
                     RRTSVCGNEKKS"
     repeat_region   40401. .40538
                     /note="B1_MM"
                     /rpt_family="SINE/Alu"
     repeat_region   41216. .41550
                     /note="L1"
                     /rpt_family="LINE/L1"
     repeat_region   41574. .42545
                     /note="Lx4"
                     /rpt_family="LINE/L1"
     enhancer        42672. .43671
                     /note="gamma 1"
     repeat_region   44207. .44538
                     /note="L1_MM"
                     /rpt_family="LINE/L1"
     repeat_region   44929. .45306
                     /note="MTA"
                     /rpt_family="LTR/MaLR"
     repeat_region   complement(45997. .52297)
                     /note="L1_MM"
                     /rpt_family="LINE/L1"
     repeat_region   complement(52755. .53823)
                     /note="L1_MM"
                     /rpt_family="LINE/L1"
     repeat_region   complement(54403. .56302)
```

```
                     /note="L1_MM"
                     /rpt_family="LINE/L1"
     repeat_region   56304. .56828
                     /note="L1"
                     /rpt_family="LINE/L1"
     repeat_region   57584. .57973
                     /note="MTD"
                     /rpt_family="LTR/MaLR"
     repeat_region   complement(59800. .60109)
                     /note="L1"
                     /rpt_family="LINE/L1"
     repeat_region   complement(61119. .62347)
                     /note="L1_MM"
                     /rpt_family="LINE/L1"
...
Note: remainder of annotations omitted.

  Query Match             6.7%;  Score 20;  DB 28;  Length 105726;
  Best Local Similarity 100.0%;  Pred. No. 3.01e-01;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  42568 TACATATATACATACATACA 42587
          ||||||||||||||||||||
Cp    197 TACATATATACATACATACA 178


RESULT   48
LOCUS       AC005000   105754 bp    DNA              HTG       12-JUN-1998
DEFINITION  *** SEQUENCING IN PROGRESS *** Homo sapiens clone DJ241P17; HTGS
            phase 1, 7 unordered pieces.
ACCESSION   AC005000
NID         g3212991
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 105754)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 105754)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-1998) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     ***                                                          ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                          ***
            * This sequence is unfinished. It consists of 7 contigs for
```



```
            * which the order is not known; their order in this record is
            * arbitrary. In some cases, the exact lengths of the gaps
            * between the contigs are also unknown; these gaps are  presented
            * as runs of N as a convenience only. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
            *       1     1462: contig of 1462 bp in length
            *    1463     1479: gap of unknown length
            *    1480     3457: contig of 1978 bp in length
            *    3458     3474: gap of unknown length
            *    3475     6303: contig of 2829 bp in length
            *    6304     6320: gap of unknown length
            *    6321    19779: contig of 13459 bp in length
            *   19780    19796: gap of unknown length
            *   19797    25173: contig of 5377 bp in length
            *   25174    25190: gap of unknown length
            *   25191    45064: contig of 19874 bp in length
            *   45065    45081: gap of unknown length
            *   45082   105754: contig of 60673 bp in length.
FEATURES             Location/Qualifiers
     source          1. .105754
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="DJ241P17"
BASE COUNT    29659 a  22985 c  23120 g  29888 t    102 others
ORIGIN

  Query Match             6.7%;  Score 20;  DB 17;  Length 105754;
  Best Local Similarity 100.0%;  Pred. No. 3.01e-01;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    114 TGTATGTATGTATATATGTA 133
          ||||||||||||||||||||
Qy    178 TGTATGTATGTATATATGTA 197


RESULT   49
LOCUS       AC004927   112098 bp    DNA              HTG       12-JUN-1998
DEFINITION  *** SEQUENCING IN PROGRESS *** Homo sapiens clone DJ0910I17; HTGS
            phase 1, 2 unordered pieces.
ACCESSION   AC004927
NID         g3213070
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 112098)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
```



```
REFERENCE   2  (bases 1 to 112098)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-1998) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     ***                                                          ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                          ***
            * This sequence is unfinished. It consists of 2 contigs for
            * which the order is not known; their order in this record is
            * arbitrary. In some cases, the exact lengths of the gaps
            * between the contigs are also unknown; these gaps are  presented
            * as runs of N as a convenience only. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
            *       1    18271: contig of 18271 bp in length
            *   18272    18289: gap of unknown length
            *   18290   112098: contig of 93809 bp in length.
FEATURES             Location/Qualifiers
     source          1. .112098
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="DJ0910I17"
BASE COUNT    31285 a  25794 c  25483 g  29415 t    121 others
ORIGIN

  Query Match             6.7%;  Score 20;  DB 17;  Length 112098;
  Best Local Similarity 100.0%;  Pred. No. 3.01e-01;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db 104927 ATATATACATACATACACAC 104946
          ||||||||||||||||||||
Cp    194 ATATATACATACATACACAC 175


RESULT   50
LOCUS       AC005373   112220 bp    DNA              PRI       01-AUG-1998
DEFINITION  Homo sapiens chromosome 5, PAC clone 45L14 (LBNL H91), complete
            sequence.
ACCESSION   AC005373 AC002271 AC002272 AC002273 AC001153 AC001154 AC001587
            AC001588 AC001589 AC001590 AC001591 AC001592 AC001593 AC001594
            AC001595 AC001596 AC001597 AC001598 AC001599 AC001600 AC001601
NID         g3367501
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 112220)
  AUTHORS   Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
```



```
            Kadner,K., Miguel,T., Miller,C., Pitluck,S., Pollard,M.,
            Rojeski,H., Subramanian,S. and Martin,C.H.
  TITLE     Sequencing of human chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 112220)
  AUTHORS   Ricke,D.O.
  TITLE     Large Scale Sequence Analysis and Annotation with the Sequence
            Comparison Analysis (SCAN) System
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 112220)
  AUTHORS   Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
            Davis,C.A., Kadner,K., Miguel,T., Pitluck,S., Pollard,M.,
            Rojeski,H., Subramanian,S. and Martin,C.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-AUG-1998) Human Genome Center, DOE Joint Genome
            Institute, Lawrence Berkeley National Laboratory, MS 74-157,
            Berkeley, CA 94720, U.S.A.
COMMENT     Sequence submitted by:
            DOE Joint Genome Institute.
FEATURES             Location/Qualifiers
     source          1. .112220
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="5q"
                     /clone="45L14"
                     /chromosome="5"
                     /note="LBNL H91"
     repeat_region   1078. .1467
                     /rpt_family="THE1"
     repeat_region   1468. .1542
                     /rpt_family="MSTAR"
     repeat_region   1553. .2593
                     /rpt_family="L1"
     repeat_region   3095. .3149
                     /rpt_family="THE1"
     repeat_region   3613. .3895
                     /rpt_family="Alu"
     repeat_region   complement(4313. .5476)
                     /rpt_family="L1"
     repeat_region   5483. .5913
                     /rpt_family="THE1"
     repeat_region   6848. .6883
                     /note="(AT)18"
                     /rpt_type=tandem
                     /rpt_unit=AT
     misc_feature    complement(9657. .9803)
                     /note="GRAIL 2 excellent exon, frame 2"
     misc_feature    complement(10526. .10664)
                     /note="GRAIL 2 excellent exon, frame 0"
     repeat_region   complement(11478. .11548)
                     /rpt_family="MER7"
```

```
     repeat_region   complement(11584. .11790)
                     /rpt_family="MER7"
     repeat_region   16900. .16923
                     /note="(T)24"
                     /rpt_type=tandem
                     /rpt_unit=T
     repeat_region   complement(16906. .17200)
                     /rpt_family="Alu"
     repeat_region   complement(17801. .18037)
                     /rpt_family="MER20"
     repeat_region   19038. .19176
                     /rpt_family="MIR"
     repeat_region   22442. .22465
                     /note="(AC)12"
                     /rpt_type=tandem
                     /rpt_unit=AC
     repeat_region   25243. .25278
                     /note="(ATCT)9"
                     /rpt_type=tandem
                     /rpt_unit=ATCT
     repeat_region   complement(25301. .25442)
                     /rpt_family="Alu"
     repeat_region   complement(25617. .25696)
                     /rpt_family="Alu"
     repeat_region   complement(25837. .25913)
                     /rpt_family="L1"
     repeat_region   25949. .26237
                     /rpt_family="Alu"
     repeat_region   26631. .26655
                     /note="(T)25"
                     /rpt_type=tandem
                     /rpt_unit=T
     repeat_region   complement(26637. .26908)
                     /rpt_family="Alu"
     repeat_region   27106. .27177
                     /rpt_family="MER1"
     repeat_region   27301. .27373
                     /rpt_family="MER1"
     repeat_region   complement(29966. .30097)
                     /rpt_family="MIR"
     repeat_region   30187. .30208
                     /note="(T)22"
                     /rpt_type=tandem
                     /rpt_unit=T
     repeat_region   complement(32444. .32513)
                     /rpt_family="THE1"
     misc_feature    complement(34698. .34832)
                     /note="GRAIL 2 excellent exon, frame 0"
     repeat_region   complement(36089. .36161)
                     /rpt_family="MIR"
     repeat_region   38186. .38265
```



```
                     /rpt_family="MER5"
     repeat_region   50236. .50502
                     /rpt_family="Alu"
     repeat_region   complement(52674. .52966)
                     /rpt_family="Alu"
     repeat_region   complement(54074. .55783)
                     /rpt_family="MER42"
     repeat_region   54259. .54499
                     /rpt_family="Alu"
     repeat_region   complement(57560. .58042)
                     /rpt_family="LOR1"
     repeat_region   59560. .59583
                     /note="(GT)12"
                     /rpt_type=tandem
                     /rpt_unit=GT
     repeat_region   complement(61313. .64155)
                     /rpt_family="L1"
     repeat_region   64162. .65230
                     /rpt_family="L1"
     repeat_region   66124. .66215
                     /rpt_family="MER4"
     repeat_region   complement(67666. .67749)
                     /rpt_family="L1"
     repeat_region   complement(68817. .69161)
                     /rpt_family="THE1"
     repeat_region   70626. .70657
                     /note="(GT)16"
                     /rpt_type=tandem
                     /rpt_unit=GT
     repeat_region   complement(73588. .73874)
                     /rpt_family="L1"
     repeat_region   76336. .76630
                     /rpt_family="Alu"
     repeat_region   77189. .77402
                     /rpt_family="Alu"
     repeat_region   77914. .78860
                     /rpt_family="L1"
     repeat_region   79121. .79156
                     /note="(AC)18"
                     /rpt_type=tandem
                     /rpt_unit=AC
     repeat_region   complement(81236. .81306)
                     /rpt_family="Alu"
     misc_feature    complement(81467. .81575)
                     /note="GRAIL 2 excellent exon, frame 0"
     misc_feature    86926. .87014
                     /note="GRAIL 2 excellent exon, frame 2"
     repeat_region   complement(88159. .90695)
                     /rpt_family="L1"
     repeat_region   91773. .91971
                     /rpt_family="MLT1"
```



```
     repeat_region   complement(92057. .92344)
                     /rpt_family="Alu"
     repeat_region   92771. .92794
                     /note="(CT)12"
                     /rpt_type=tandem
                     /rpt_unit=CT
     repeat_region   93462. .93771
                     /rpt_family="Alu"
     repeat_region   93752. .93776
                     /note="(A)25"
                     /rpt_type=tandem
                     /rpt_unit=A
     repeat_region   complement(93868. .94039)
                     /rpt_family="MIR"
     repeat_region   105572. .105691
                     /rpt_family="Alu"
     repeat_region   complement(108489. .108731)
                     /rpt_family="Alu"
     repeat_region   110933. .111122
                     /rpt_family="Alu"
     repeat_region   complement(111222. .111467)
                     /rpt_family="Alu"
BASE COUNT    36180 a  20313 c  19896 g  35831 t
ORIGIN

  Query Match           6.7%;  Score 20;  DB 26;  Length 112220;
  Best Local Similarity 100.0%;  Pred. No. 3.01e-01;
  Matches   20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  28512 TGTATGTATGTATATATGTA 28531
          ||||||||||||||||||||
Qy    178 TGTATGTATGTATATATGTA 197


Search completed: Tue Dec 22 20:18:31 1998
Job time : 729 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Tue Dec 22 20:41:06 1998; MasPar time 271.12 Seconds
1149.286 Million cell updates/sec

Tabular output not generated.

Title: >US-08-530-112A-1219
Description: (1-174) from US08530112A.seq
Perfect Score: 174
N.A. Sequence: 1 GATCTGAGCAGGCAAAGCTC.....AAGTGAATTTGCAAGTGAAA 174
Comp: CTAGACTCGTCCGTTTCGAG.....TTCACTTAAACGTTCACTTT

Scoring table: TABLE jmetric
Gap 60

Nmatch STD : Dbase 0; Query 0

Searched: 2275026 seqs, 895388244 bases x 2

Post-processing: Minimum Match 0%
Listing first 1000 summaries

Database: embl-est55
1:em_est1 2:em_gss1 3:em_gss2 4:em_gss3
Database: genbank-est107
5:gb_est1 6:gb_est10 7:gb_est11 8:gb_est12 9:gb_est13
10:gb_est14 11:gb_est15 12:gb_est16 13:gb_est17
14:gb_est18 15:gb_est19 16:gb_est2 17:gb_est20
18:gb_est21 19:gb_est3 20:gb_est4 21:gb_est5 22:gb_est6
23:gb_est7 24:gb_est8 25:gb_est9 26:gb_gss1 27:gb_gss2
28:gb_gss3 29:gb_gss4

Statistics: Mean 8.101; Variance 1.552; scale 5.221

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
c	1	174	100.0	458	13	AA846103	ak83f07.s1 Barstead sp	0.00e+00
c	2	171	98.3	440	15	AA894763	oj55a12.s1 NCI_CGAP_Ki	0.00e+00
	3	167	96.0	174	7	HUMGS01365	Human HL60 3'directed	2.78e-301
c	4	20	11.5	353	25	AA372536	EST84447 Colon adenoca	7.70e-06
	5	19	10.9	402	14	C89667	Dictyostelium discoide	1.77e-04
	6	19	10.9	442	7	C24637	Dictyostelium discoide	1.77e-04
c	7	19	10.9	453	24	AA323500	EST26285 Cerebellum II	1.77e-04
	8	19	10.9	473	14	C90335	Dictyostelium discoide	1.77e-04
	9	18	10.3	157	6	C23819	Dictyostelium discoide	3.66e-03
	10	18	10.3	213	19	N20126	yx29e02.s1 Homo sapien	3.66e-03
	11	18	10.3	213	14	AA957734	UI-R-E1-gc-b-12-0-UI.s	3.66e-03
	12	18	10.3	271	5	T18134	0695c3 Plasmodium falc	3.66e-03
	13	18	10.3	286	15	AA998261	UI-R-C0-ie-g-05-0-UI.s	3.66e-03
	14	18	10.3	308	18	AI144676	UI-R-BT0-pl-h-06-0-UI.	3.66e-03
c	15	18	10.3	374	28	AQ194533	RPCI11-60K17.TK RPCI11	3.66e-03
c	16	18	10.3	386	26	AQ002609	CIT978SK-176G21-1.TV C	3.66e-03
	17	18	10.3	404	28	AQ124657	HS_2267_B1_H04_MF CIT	3.66e-03
c	18	18	10.3	446	27	AQ081794	RPCI11-55K19.TK RPCI11	3.66e-03
	19	18	10.3	453	14	C93203	Dictyostelium discoide	3.66e-03
c	20	18	10.3	480	23	AA138125	mr05g10.r1 Soares mous	3.66e-03
c	21	18	10.3	481	10	AA619999	vl76e01.r1 Knowles Sol	3.66e-03
c	22	18	10.3	487	22	AA020977	ze65d07.s1 Soares reti	3.66e-03
	23	18	10.3	523	17	AI070018	UI-R-C1-ln-a-07-0-UI.s	3.66e-03
c	24	18	10.3	535	23	AA106432	ml94c03.r1 Stratagene	3.66e-03
c	25	18	10.3	560	23	AA204477	mu32e11.r1 Soares 2NbM	3.66e-03
	26	18	10.3	585	17	AI131650	ui41b01.x1 Sugano mous	3.66e-03
c	27	18	10.3	602	21	W73116	zd55d10.r1 Soares feta	3.66e-03
c	28	18	10.3	603	12	AA794906	vr48e06.s1 Knowles Sol	3.66e-03
	29	18	10.3	621	21	W73136	zd55d10.s1 Soares feta	3.66e-03
	30	18	10.3	720	27	AQ041937	CIT-HSP-2335B16.TF CIT	3.66e-03
c	31	17	9.8	187	25	AA337232	EST42297 Endometrial t	6.68e-02
	32	17	9.8	205	15	AA999152	UI-R-E1-fy-h-12-0-UI.s	6.68e-02
c	33	17	9.8	208	7	HUML11312	Human fetal-lung cDNA	6.68e-02
c	34	17	9.8	212	15	C94090	Dictyostelium discoide	6.68e-02
c	35	17	9.8	213	7	HUML1371	Human fetal-lung cDNA	6.68e-02
c	36	17	9.8	242	7	AA476057	vh26f02.r1 Soares mous	6.68e-02
c	37	17	9.8	256	7	HUM420E11B	Human fetal brain cDNA	6.68e-02
	38	17	9.8	273	24	AA327070	EST30320 Colon I Homo	6.68e-02
c	39	17	9.8	274	5	T93193	ye24g08.r1 Homo sapien	6.68e-02
	40	17	9.8	285	21	W40029	T2642 MVAT4 bloodstrea	6.68e-02
	41	17	9.8	287	8	AA281555	zt04a01.s1 NCI_CGAP_GC	6.68e-02
c	42	17	9.8	300	9	C53994	C.elegans cDNA clone y	6.68e-02

c	43	17	9.8	307	15	AA263147	PMY0513 KG1-a Lambda Z	6.68e-02
c	44	17	9.8	308	14	C91163	Dictyostelium discoide	6.68e-02
	45	17	9.8	313	14	AA957958	UI-R-E1-fs-g-04-0-UI.s	6.68e-02
c	46	17	9.8	317	5	T93501	ye10b12.r1 Homo sapien	6.68e-02
	47	17	9.8	318	15	AI011239	EST2056490 Normalized r	6.68e-02
c	48	17	9.8	319	5	T92492	ye21h11.r1 Homo sapien	6.68e-02
	49	17	9.8	323	14	AA956559	UI-R-E1-fk-a-03-0-UI.s	6.68e-02
c	50	17	9.8	325	16	R80040	yi91d06.r1 Homo sapien	6.68e-02
	51	17	9.8	326	13	AA620463	ae58e04.s1 Stratagene	6.68e-02
	52	17	9.8	333	5	T85983	yd62d09.r1 Homo sapien	6.68e-02
	53	17	9.8	336	14	AA986530	ue14e03.x1 Sugano mous	6.68e-02
c	54	17	9.8	341	5	T90412	ye16d02.r1 Homo sapien	6.68e-02
c	55	17	9.8	341	7	HUML11592	Human fetal-lung cDNA	6.68e-02
c	56	17	9.8	343	5	T90582	ye15b06.r1 Homo sapien	6.68e-02
c	57	17	9.8	343	5	T93237	ye28g01.r1 Homo sapien	6.68e-02
c	58	17	9.8	343	6	CELK022DXR	C.elegans cDNA clone y	6.68e-02
c	59	17	9.8	345	15	AA934368	op27f12.s1 Soares_NFL_	6.68e-02
c	60	17	9.8	347	5	T93175	ye24e04.r1 Homo sapien	6.68e-02
c	61	17	9.8	348	16	R44558	yg29e07.s1 Homo sapien	6.68e-02
c	62	17	9.8	363	11	AA673574	vn70h03.r1 Barstead mo	6.68e-02
	63	17	9.8	366	21	SSC25B04	S.acrofa mRNA; express	6.68e-02
c	64	17	9.8	368	5	T94023	ye33b04.r1 Homo sapien	6.68e-02
c	65	17	9.8	369	12	AA141945	CK00046.3prime CK Dros	6.68e-02
c	66	17	9.8	371	5	T87917	ye10b01.r1 Homo sapien	6.68e-02
c	67	17	9.8	372	20	W62560	md73c12.r1 Soares mous	6.68e-02
c	68	17	9.8	373	5	T94463	ye35a07.r1 Homo sapien	6.68e-02
c	69	17	9.8	382	22	AA133803	zo16c09.r1 Stratagene	6.68e-02
c	70	17	9.8	383	28	AQ134132	HS_3057_A2_B01_MR CIT	6.68e-02
c	71	17	9.8	383	13	AA806504	oc23g10.s1 NCI_CGAP_GC	6.68e-02
c	72	17	9.8	383	13	AA814396	nz12b07.s1 NCI_CGAP_GC	6.68e-02
c	73	17	9.8	386	5	T92999	ye23e10.r1 Homo sapien	6.68e-02
	74	17	9.8	388	27	B87274	RPCI11-28G10.TP RPCI11	6.68e-02
c	75	17	9.8	389	5	T93761	ye10g02.r1 Homo sapien	6.68e-02
c	76	17	9.8	389	5	T90401	ye16b06.r1 Homo sapien	6.68e-02
c	77	17	9.8	395	14	AA930156	vz52f02.s1 Soares 2NbM	6.68e-02
c	78	17	9.8	395	28	AQ205617	HS_3235_B1_B02_MR CIT	6.68e-02
c	79	17	9.8	396	5	T91550	ye20g07.r1 Homo sapien	6.68e-02
c	80	17	9.8	397	23	AA166116	ms28f01.r1 Stratagene	6.68e-02
c	81	17	9.8	397	5	T89807	ye11b07.r1 Homo sapien	6.68e-02
c	82	17	9.8	397	5	T94154	ye28e11.r1 Homo sapien	6.68e-02
c	83	17	9.8	397	5	T93050	ye27f09.r1 Homo sapien	6.68e-02
c	84	17	9.8	398	5	T94502	ye36a01.r1 Homo sapien	6.68e-02
c	85	17	9.8	398	5	T90428	ye16g01.r1 Homo sapien	6.68e-02
c	86	17	9.8	400	5	T92820	ye23f04.r1 Homo sapien	6.68e-02
c	87	17	9.8	401	5	T96475	ye34a07.r1 Homo sapien	6.68e-02
c	88	17	9.8	402	26	AQ022969	HS_2183_A2_D09_MR CIT	6.68e-02
c	89	17	9.8	402	5	R10564	yf31g06.r1 Homo sapien	6.68e-02
c	90	17	9.8	403	5	T92667	ye26a06.r1 Homo sapien	6.68e-02
c	91	17	9.8	406	5	T90397	ye16a07.r1 Homo sapien	6.68e-02
	92	17	9.8	407	28	AQ124820	HS_2270_A1_D01_MF CIT	6.68e-02
c	93	17	9.8	409	17	AI106339	ab03c07_p1 ZF adult he	6.68e-02

c	94	17	9.8	412	5	T92395	ye19e04.r1 Homo sapien	6.68e-02
c	95	17	9.8	413	5	T92380	ye19c03.r1 Homo sapien	6.68e-02
c	96	17	9.8	415	11	AA687396	nv62d02.s1 NCI_CGAP_GC	6.68e-02
c	97	17	9.8	415	12	AA765941	oa14f02.s1 NCI_CGAP_GC	6.68e-02
c	98	17	9.8	415	5	T94542	ye36h01.r1 Homo sapien	6.68e-02
c	99	17	9.8	415	5	T92160	ye17g03.r1 Homo sapien	6.68e-02
c	100	17	9.8	416	5	T92290	ye18d02.r1 Homo sapien	6.68e-02
c	101	17	9.8	417	19	N50909	yy93a01.s1 Homo sapien	6.68e-02
c	102	17	9.8	422	23	AA184186	mt32f02.r1 Soares mous	6.68e-02
c	103	17	9.8	422	5	T94035	ye33d05.r1 Homo sapien	6.68e-02
c	104	17	9.8	422	6	CELK007C6R	C.elegans cDNA clone y	6.68e-02
c	105	17	9.8	423	5	T94216	ye33b09.r1 Homo sapien	6.68e-02
c	106	17	9.8	424	5	T94266	ye31f06.r1 Homo sapien	6.68e-02
c	107	17	9.8	424	5	T90507	ye15f08.r1 Homo sapien	6.68e-02
c	108	17	9.8	424	5	T92621	ye22b03.r1 Homo sapien	6.68e-02
c	109	17	9.8	425	5	T94054	ye33g04.r1 Homo sapien	6.68e-02
c	110	17	9.8	428	7	AA449474	zx08f09.s1 Soares tota	6.68e-02
c	111	17	9.8	430	5	T92305	ye18f07.r1 Homo sapien	6.68e-02
c	112	17	9.8	430	28	AQ112709	CIT-HSP-2370A6.TR CIT-	6.68e-02
c	113	17	9.8	431	5	T92866	ye27f06.r1 Homo sapien	6.68e-02
c	114	17	9.8	432	5	T92632	ye22c07.r1 Homo sapien	6.68e-02
	115	17	9.8	434	11	AA120822	zm25a10.s1 Stratagene	6.68e-02
c	116	17	9.8	434	5	T96479	ye34b07.r1 Homo sapien	6.68e-02
c	117	17	9.8	434	5	T92135	ye17b01.r1 Homo sapien	6.68e-02
c	118	17	9.8	434	5	T94491	ye35f02.r1 Homo sapien	6.68e-02
c	119	17	9.8	439	20	W31574	zb93f04.r1 Soares para	6.68e-02
c	120	17	9.8	439	20	W63996	md78g05.r1 Soares mous	6.68e-02
c	121	17	9.8	441	23	AA122809	mr04h10.r1 Soares mous	6.68e-02
c	122	17	9.8	441	29	AQ239742	CIT-HSP-2386G19.TF.1 C	6.68e-02
c	123	17	9.8	442	15	AA996832	UI-R-C0-hg-d-02-0-UI.s	6.68e-02
c	124	17	9.8	445	5	T93009	ye23h09.r1 Homo sapien	6.68e-02
c	125	17	9.8	449	19	N50903	yy92h02.s1 Homo sapien	6.68e-02
	126	17	9.8	452	29	AQ235487	HS_2055_A1_A03_MR CIT	6.68e-02
c	127	17	9.8	455	21	AA050825	mj22e08.r1 Soares mous	6.68e-02
	128	17	9.8	456	14	AA925253	UI-R-A1-ee-e-01-0-UI.s	6.68e-02
	129	17	9.8	459	27	AQ082737	RPCI11-54C16.TJ RPCI11	6.68e-02
c	130	17	9.8	459	21	AA048989	mj50h05.r1 Soares mous	6.68e-02
c	131	17	9.8	463	12	AA797400	vw27e04.r1 Soares mous	6.68e-02
c	132	17	9.8	464	27	B83613	RPCI11-16P19.TVB RPCI1	6.68e-02
	133	17	9.8	467	14	AA829813	od40a01.s1 NCI_CGAP_GC	6.68e-02
	134	17	9.8	467	17	AI075074	ou61h12.x1 NCI_CGAP_Br	6.68e-02
c	135	17	9.8	479	5	T91692	ye25h03.r1 Homo sapien	6.68e-02
c	136	17	9.8	479	29	AQ227388	HS_2017_B1_D07_MR CIT	6.68e-02
c	137	17	9.8	481	19	N71087	za86g12.s1 Homo sapien	6.68e-02
c	138	17	9.8	486	29	AQ235991	HS_2056_A1_F12_T7 CIT	6.68e-02
	139	17	9.8	490	7	AA449637	zx08f09.r1 Soares tota	6.68e-02
c	140	17	9.8	490	6	AA396369	vb40g11.r1 Soares mous	6.68e-02
c	141	17	9.8	492	5	T89221	ye12c08.r1 Homo sapien	6.68e-02
c	142	17	9.8	493	13	AA818428	UI-R-A0-au-g-06-0-UI.s	6.68e-02
	143	17	9.8	494	12	AA819817	UI-R-A0-aq-h-07-0-UI.s	6.68e-02
c	144	17	9.8	497	12	AA776790	ah36g12.s1 Soares test	6.68e-02

c	145	17	9.8	500	5	T89236	ye12f01.r1 Homo sapien	6.68e-02
c	146	17	9.8	505	5	T93539	ye14c02.r1 Homo sapien	6.68e-02
c	147	17	9.8	513	22	AA086024	zl84a03.r1 Stratagene	6.68e-02
	148	17	9.8	527	8	AA549976	1055m3 gmbPfHB3.1, G.	6.68e-02
c	149	17	9.8	530	26	B15706	345013.TV CIT978SKA1 H	6.68e-02
c	150	17	9.8	541	27	B59405	CIT-HSP-2017H13.TR CIT	6.68e-02
c	151	17	9.8	543	5	T89926	ye12b10.r1 Homo sapien	6.68e-02
	152	17	9.8	547	27	AQ037289	CIT-HSP-2335J6.TF CIT-	6.68e-02
	153	17	9.8	574	12	AA817834	UI-R-A0-ae-e-10-0-UI.s	6.68e-02
c	154	17	9.8	577	7	AA537225	vk45h10.r1 Soares mous	6.68e-02
	155	17	9.8	593	28	AQ208256	HS_3219_A1_B07_T7 CIT	6.68e-02
	156	17	9.8	616	17	AI109619	GH08858.5prime GH Dros	6.68e-02
c	157	17	9.8	619	27	FR0032515	Fugu rubripes GSS sequ	6.68e-02
c	158	17	9.8	648	27	B92992	CIT-HSP-2168N19.TR CIT	6.68e-02
c	159	17	9.8	668	8	AA550120	1221m3 gmbPfHB3.1, G.	6.68e-02
	160	17	9.8	670	27	AQ083792	CpG0633A CpIOWAgDNA1 C	6.68e-02
c	161	17	9.8	693	11	C84185	Dictyostelium discoide	6.68e-02
	162	17	9.8	720	27	AQ061777	CIT-HSP-2352L10.TR CIT	6.68e-02
	163	17	9.8	932	15	AA826243	of11b07.s1 NCI_CGAP_Co	6.68e-02
	164	17	9.8	937	28	AF011034	Homo sapiens chromosom	6.68e-02
	165	16	9.2	120	20	DMD11EST	D.melanogaster EST, cl	1.06e+00
c	166	16	9.2	120	7	CELK093G7R	C.elegans cDNA clone y	1.06e+00
	167	16	9.2	146	16	R63093	yi01g02.s1 Homo sapien	1.06e+00
c	168	16	9.2	150	21	AA025557	ze85e06.s1 Soares feta	1.06e+00
c	169	16	9.2	171	7	MUSGS00271	Mouse liver 3'-directe	1.06e+00
c	170	16	9.2	185	18	AI152607	ud89d11.r1 Soares mous	1.06e+00
c	171	16	9.2	187	7	AA448873	zx11e03.s1 Soares tota	1.06e+00
c	172	16	9.2	194	10	AA631750	np76e04.s1 NCI_CGAP_Br	1.06e+00
	173	16	9.2	203	10	AA110260	mp07g10.r1 Life Tech m	1.06e+00
c	174	16	9.2	214	13	AA181321	zp16a03.r1 Stratagene	1.06e+00
c	175	16	9.2	216	26	B53686	CIT-HSP-2012C8.TR CIT-	1.06e+00
	176	16	9.2	219	6	AA036894	zk29c11.r1 Soares preg	1.06e+00
c	177	16	9.2	222	22	N98006	2029C3 czapPFDd2.1, De	1.06e+00
	178	16	9.2	226	5	T02381	wEST03102 Caenorhabdit	1.06e+00
	179	16	9.2	228	18	AI153487	uc50b01.r1 Soares 2NbM	1.06e+00
c	180	16	9.2	229	22	N98102	2181C3 czapPFDd2.1, De	1.06e+00
c	181	16	9.2	231	12	AA812505	ai73d10.s1 Soares test	1.06e+00
c	182	16	9.2	234	26	B05320	cSRL-SD7-u cSRL flow s	1.06e+00
	183	16	9.2	235	7	HUM354D08B	Human aorta cDNA 5'-en	1.06e+00
c	184	16	9.2	241	7	HUM031H02A	Human fetal brain cDNA	1.06e+00
c	185	16	9.2	244	20	HSC05H062	H. sapiens partial cDN	1.06e+00
	186	16	9.2	246	14	AA921070	vy83h10.r1 Stratagene	1.06e+00
	187	16	9.2	247	5	T02382	wEST03103 Caenorhabdit	1.06e+00
c	188	16	9.2	247	24	AA296740	EST11228 Uterus Homo s	1.06e+00
c	189	16	9.2	250	5	T00431	wEST01152 Caenorhabdit	1.06e+00
c	190	16	9.2	253	25	AA358358	EST67282 Fetal lung II	1.06e+00
	191	16	9.2	255	5	T02316	wEST03037 Caenorhabdit	1.06e+00
	192	16	9.2	257	16	T27610	EST101084 Homo sapiens	1.06e+00
	193	16	9.2	258	6	AA010360	zi08f11.s1 Soares feta	1.06e+00
	194	16	9.2	259	27	AQ062240	CIT-HSP-2347N8.TF CIT-	1.06e+00
	195	16	9.2	259	18	AI158117	ud23c10.r1 Soares 2NbM	1.06e+00

c	196	16	9.2	259	21	T16524	NIB1428 Normalized inf	1.06e+00
c	197	16	9.2	261	19	H72427	ys06h08.r1 Homo sapien	1.06e+00
c	198	16	9.2	261	23	AA108731	mp35b07.r1 Barstead MP	1.06e+00
	199	16	9.2	263	22	AA120451	mn47f10.r1 Beddington	1.06e+00
c	200	16	9.2	275	27	AQ002768	CIT-HSP-254P8-1.TV CIT	1.06e+00
c	201	16	9.2	278	5	T01515	wEST02236 Caenorhabdit	1.06e+00
c	202	16	9.2	278	24	HSPD02159	H.sapiens EST sequence	1.06e+00
	203	16	9.2	284	15	AI008970	Schizosaccharomyces po	1.06e+00
	204	16	9.2	284	25	AA372593	EST84589 Colon adenoca	1.06e+00
c	205	16	9.2	284	10	T04420	467 Lambda-PRL2 Arabid	1.06e+00
c	206	16	9.2	285	11	AA688799	vr99a11.r1 Barstead mo	1.06e+00
	207	16	9.2	287	8	AA480876	aa28c02.s1 NCI_CGAP_GC	1.06e+00
	208	16	9.2	289	12	AA814359	nz07h01.s1 NCI_CGAP_GC	1.06e+00
c	209	16	9.2	289	8	AA281264	zs94f09.s1 NCI_CGAP_GC	1.06e+00
	210	16	9.2	294	15	AI030479	UI-R-C0-jc-b-06-0-UI.s	1.06e+00
	211	16	9.2	295	14	AA899768	UI-R-E0-da-d-05-0-UI.s	1.06e+00
c	212	16	9.2	297	12	AA713671	nv69c07.s1 NCI_CGAP_GC	1.06e+00
c	213	16	9.2	300	9	C35028	C.elegans cDNA clone y	1.06e+00
c	214	16	9.2	300	9	C36410	C.elegans cDNA clone y	1.06e+00
c	215	16	9.2	300	6	C08332	C.elegans cDNA clone y	1.06e+00
c	216	16	9.2	300	9	C31012	C.elegans cDNA clone y	1.06e+00
c	217	16	9.2	300	9	C35850	C.elegans cDNA clone y	1.06e+00
c	218	16	9.2	300	9	C32657	C.elegans cDNA clone y	1.06e+00
c	219	16	9.2	300	9	C38533	C.elegans cDNA clone y	1.06e+00
c	220	16	9.2	300	9	C33960	C.elegans cDNA clone y	1.06e+00
	221	16	9.2	307	20	HSB98D121	H. sapiens partial cDN	1.06e+00
c	222	16	9.2	310	9	AA586151	28800 Lambda-PRL2 Arab	1.06e+00
	223	16	9.2	313	15	RICC0895A	Rice cDNA, partial seq	1.06e+00
	224	16	9.2	316	5	T01221	wEST01942 Caenorhabdit	1.06e+00
	225	16	9.2	316	27	AQ089584	HS_3005_A1_G05_MR CIT	1.06e+00
	226	16	9.2	316	23	AA119088	mp65g08.r1 Soares 2NbM	1.06e+00
c	227	16	9.2	316	15	AA897702	of90d06.s1 NCI_CGAP_Li	1.06e+00
	228	16	9.2	317	15	AI019362	ub23a02.r1 Soares 2NbM	1.06e+00
	229	16	9.2	318	18	AI181505	uc57h02.r1 Soares 2NbM	1.06e+00
	230	16	9.2	319	15	AI049412	ub33e02.r1 Soares 2NbM	1.06e+00
c	231	16	9.2	320	5	T93698	ye09b04.r1 Homo sapien	1.06e+00
	232	16	9.2	322	23	AA216795	mu86c02.r1 Soares mous	1.06e+00
c	233	16	9.2	322	8	AA226602	nc18f02.r1 NCI_CGAP_Pr	1.06e+00
c	234	16	9.2	323	21	T16720	NIB1755 Normalized inf	1.06e+00
c	235	16	9.2	324	15	AA938767	on28e01.s1 NCI_CGAP_Lu	1.06e+00
c	236	16	9.2	325	5	T93713	ye09e04.r1 Homo sapien	1.06e+00
c	237	16	9.2	328	14	AA947094	on69d04.s1 Soares_NFL_	1.06e+00
	238	16	9.2	329	8	AA234582	zr75b03.r1 Soares NhHM	1.06e+00
	239	16	9.2	332	23	AA073064	mm78h06.r1 Stratagene	1.06e+00
	240	16	9.2	332	17	AI122465	uc58e11.r1 Soares 2NbM	1.06e+00
c	241	16	9.2	333	16	H18506	yn50e04.s1 Homo sapien	1.06e+00
	242	16	9.2	334	7	CELK112E9R	C.elegans cDNA clone y	1.06e+00
c	243	16	9.2	334	26	FR0019646	F.rubripes GSS sequenc	1.06e+00
c	244	16	9.2	335	14	AA826565	oe64g11.s1 NCI_CGAP_Lu	1.06e+00
	245	16	9.2	337	5	T01749	wEST02470 Caenorhabdit	1.06e+00
	246	16	9.2	338	21	R84764	yt65d12.r1 Soares reti	1.06e+00

c	247	16	9.2	338	23	AA213103	mw82h08.r1 Soares mous	1.06e+00
	248	16	9.2	342	29	AQ152463	HS_3108_A2_H02_MR CIT	1.06e+00
	249	16	9.2	343	19	N53201	yy97f01.r1 Homo sapien	1.06e+00
c	250	16	9.2	344	28	AQ114467	CIT-HSP-2378F16.TF CIT	1.06e+00
c	251	16	9.2	345	26	B16793	343H12.TPB CIT978SKA1	1.06e+00
	252	16	9.2	349	19	ATTS6178	A. thaliana transcribe	1.06e+00
	253	16	9.2	352	10	AA424407	zv82g02.s1 Soares tota	1.06e+00
	254	16	9.2	354	14	AA830284	oc49c04.s1 NCI_CGAP_GC	1.06e+00
	255	16	9.2	355	23	AA119084	mp65e08.r1 Soares 2NbM	1.06e+00
	256	16	9.2	357	15	AI050384	ub30c07.r1 Soares 2NbM	1.06e+00
c	257	16	9.2	357	14	AA825457	oe64f11.s1 NCI_CGAP_Lu	1.06e+00
	258	16	9.2	360	23	AA111062	mp62g12.r1 Soares 2NbM	1.06e+00
	259	16	9.2	360	6	CELK056F7F	C.elegans cDNA clone y	1.06e+00
c	260	16	9.2	360	7	CELK097DXF	C.elegans cDNA clone y	1.06e+00
c	261	16	9.2	360	9	C41530	C.elegans cDNA clone y	1.06e+00
c	262	16	9.2	360	7	CELK082E5P	C.elegans cDNA clone y	1.06e+00
c	263	16	9.2	360	6	C09761	C.elegans cDNA clone y	1.06e+00
	264	16	9.2	365	16	H26953	yo58c01.s1 Homo sapien	1.06e+00
	265	16	9.2	366	6	AA388381	vc95b08.r1 Barstead MP	1.06e+00
	266	16	9.2	367	10	AA429987	zw60f06.r1 Soares tota	1.06e+00
c	267	16	9.2	372	22	AA172678	ms56h05.r1 Life Tech m	1.06e+00
c	268	16	9.2	373	21	AA009183	mg95g06.r1 Soares mous	1.06e+00
	269	16	9.2	374	14	AA914571	vz02d12.r1 Soares mous	1.06e+00
c	270	16	9.2	375	9	C47726	C.elegans cDNA clone y	1.06e+00
c	271	16	9.2	376	28	AQ202949	RPCI11-46H6.TJ RPCI11	1.06e+00
	272	16	9.2	377	12	AA763304	vv89g08.r1 Stratagene	1.06e+00
c	273	16	9.2	377	21	W69753	zd48d05.r1 Soares feta	1.06e+00
	274	16	9.2	380	27	AQ037326	CIT-HSP-2334N19.TF CIT	1.06e+00
c	275	16	9.2	380	8	AA485140	aa40g07.s1 NCI_CGAP_GC	1.06e+00
	276	16	9.2	381	18	AI136513	UI-R-C2p-nq-b-05-0-UI.	1.06e+00
c	277	16	9.2	381	11	AA122111	zm23a05.r1 Stratagene	1.06e+00
	278	16	9.2	383	6	C04247	Human Heart cDNA, clon	1.06e+00
c	279	16	9.2	384	28	AQ101165	HS_3102_B1_C07_T7 CIT	1.06e+00
c	280	16	9.2	385	17	AI126772	qb85h11.x1 Soares_feta	1.06e+00
	281	16	9.2	386	15	AI047806	uh83f07.r1 Soares mous	1.06e+00
c	282	16	9.2	387	19	H83701	yv82f02.s1 Homo sapien	1.06e+00
	283	16	9.2	388	26	AQ014732	HS_2200_A1_MR_D12 CIT	1.06e+00
	284	16	9.2	388	5	T01557	wEST02278 Caenorhabdit	1.06e+00
c	285	16	9.2	388	29	AQ136007	HS_3055_B2_G09_MR CIT	1.06e+00
	286	16	9.2	392	16	R80729	yi93h12.r1 Homo sapien	1.06e+00
	287	16	9.2	394	18	AI140489	ow82f09.s1 Soares_feta	1.06e+00
c	288	16	9.2	395	23	AA184255	mt32b05.r1 Soares mous	1.06e+00
	289	16	9.2	396	28	AQ166951	HS_3148_B2_G05_MR CIT	1.06e+00
	290	16	9.2	397	20	HHEA09Z	H. sapiens partial cDN	1.06e+00
c	291	16	9.2	397	16	H16752	ym34d01.s1 Homo sapien	1.06e+00
	292	16	9.2	399	10	AA604409	np57a05.s1 NCI_CGAP_Br	1.06e+00
	293	16	9.2	399	22	AA052000	SWAMCA1480SK Brugia ma	1.06e+00
	294	16	9.2	402	16	H28364	yl52e03.s1 Homo sapien	1.06e+00
	295	16	9.2	403	6	AA039275	zk38d04.r1 Soares preg	1.06e+00
	296	16	9.2	404	15	AI009213	EST203664 Normalized r	1.06e+00
c	297	16	9.2	404	5	T94231	ye33h09.r1 Homo sapien	1.06e+00

c	298	16	9.2	405	19	H69343	yu19c12.r1 Homo sapien	1.06e+00
c	299	16	9.2	406	6	W90568	zh75g11.r1 Soares feta	1.06e+00
	300	16	9.2	407	25	AA358093	EST66951 Fetal lung II	1.06e+00
c	301	16	9.2	411	23	AA197445	mu18c04.r1 Soares 2NbM	1.06e+00
c	302	16	9.2	411	28	AQ170049	HS_3178_B1_D09_MR CIT	1.06e+00
c	303	16	9.2	413	28	AQ208532	HS_3237_A1_B08_MR CIT	1.06e+00
	304	16	9.2	416	8	AA494648	ne38a04.s1 NCI_CGAP_Co	1.06e+00
c	305	16	9.2	416	28	AQ170066	HS_3178_B1_G04_MR CIT	1.06e+00
	306	16	9.2	420	7	AA464062	zx82e05.r1 Soares ovar	1.06e+00
c	307	16	9.2	421	8	AA262478	zs17g09.s1 NCI_CGAP_GC	1.06e+00
	308	16	9.2	422	20	W05220	za78c02.r1 Soares feta	1.06e+00
	309	16	9.2	423	27	AQ030058	RPCI11-39B14.TV RPCI11	1.06e+00
c	310	16	9.2	424	26	AQ017888	CIT-HSP-2308P21.TF CIT	1.06e+00
	311	16	9.2	425	16	H12810	yj14d01.r1 Homo sapien	1.06e+00
	312	16	9.2	425	5	R54669	yj74d09.r1 Homo sapien	1.06e+00
c	313	16	9.2	427	6	W90031	zh77c02.s1 Soares feta	1.06e+00
	314	16	9.2	428	17	AI081593	ox63h02.s1 Soares_NhHM	1.06e+00
	315	16	9.2	428	26	B52322	CIT-HSP-388P9.TF CIT-H	1.06e+00
c	316	16	9.2	428	19	N22732	yx65a06.s1 Homo sapien	1.06e+00
	317	16	9.2	429	16	R82866	yj25e12.r1 Homo sapien	1.06e+00
c	318	16	9.2	430	23	AA253718	mv76e01.r1 Soares mous	1.06e+00
	319	16	9.2	432	21	R94056	yt73h02.r1 Soares feta	1.06e+00
	320	16	9.2	434	19	H60147	yr37f11.r1 Homo sapien	1.06e+00
	321	16	9.2	434	17	AI070052	UI-R-C1-ln-g-07-0-UI.s	1.06e+00
c	322	16	9.2	434	10	AA424650	zv92a08.s1 Soares NhHM	1.06e+00
c	323	16	9.2	436	11	AA726541	vu41a07.r1 Barstead mo	1.06e+00
	324	16	9.2	440	27	AQ036803	HS_2175_B2_D10_MR CIT	1.06e+00
	325	16	9.2	440	22	W95639	ze06f11.r1 Soares feta	1.06e+00
c	326	16	9.2	440	19	N26614	yx92b07.s1 Homo sapien	1.06e+00
c	327	16	9.2	441	6	AA410488	zv23c03.r1 Soares NhHM	1.06e+00
c	328	16	9.2	442	15	AI013399	EST208074 Normalized r	1.06e+00
	329	16	9.2	443	26	B45994	HS-1062-B2-C03-MF.abi	1.06e+00
c	330	16	9.2	445	20	W49201	mc91f10.r1 Soares mous	1.06e+00
	331	16	9.2	446	16	R44602	yg23h09.s1 Homo sapien	1.06e+00
c	332	16	9.2	447	13	AA630209	ab99g09.s1 Stratagene	1.06e+00
	333	16	9.2	448	24	AA274011	vb70h01.r1 Barstead mo	1.06e+00
	334	16	9.2	449	20	N75255	yz74d10.r1 Homo sapien	1.06e+00
	335	16	9.2	449	28	AQ193985	RPCI11-43P13.TK RPCI11	1.06e+00
	336	16	9.2	450	28	AQ113670	CIT-HSP-2371A6.TF CIT-	1.06e+00
c	337	16	9.2	450	20	AA004024	mg80f06.r1 Soares mous	1.06e+00
c	338	16	9.2	450	20	ATTS0155	A. thaliana transcribe	1.06e+00
	339	16	9.2	451	18	AI144296	qb59h03.x1 NCI_CGAP_Br	1.06e+00
	340	16	9.2	451	18	AI184627	qf46d05.x1 Soares_test	1.06e+00
c	341	16	9.2	451	18	AI170430	EST216357 Normalized r	1.06e+00
	342	16	9.2	452	19	N44690	yy22c08.r1 Homo sapien	1.06e+00
c	343	16	9.2	452	8	AA283134	zt14c04.r1 NCI_CGAP_GC	1.06e+00
c	344	16	9.2	454	22	AA132973	zo22c04.s1 Stratagene	1.06e+00
c	345	16	9.2	457	19	N62071	yz85c12.s1 Homo sapien	1.06e+00
c	346	16	9.2	460	19	N25267	yx74h01.s1 Homo sapien	1.06e+00
c	347	16	9.2	460	27	AQ005234	CIT-HSP-2290P6.TR CIT-	1.06e+00
c	348	16	9.2	461	28	AQ178229	HS_2217_B2_B12_MR CIT	1.06e+00

c	349	16	9.2	461	27	AQ003454	RPCI11-28N10.TPB RPCI1	1.06e+00
c	350	16	9.2	464	7	AA443305	zw94c01.s1 Soares tota	1.06e+00
c	351	16	9.2	465	7	AA447911	zx11b11.r1 Soares tota	1.06e+00
c	352	16	9.2	466	22	AA025513	ze90f04.r1 Soares feta	1.06e+00
c	353	16	9.2	466	24	AA267683	mz99h12.r1 Soares mous	1.06e+00
c	354	16	9.2	467	19	N63466	yy61d05.s1 Homo sapien	1.06e+00
	355	16	9.2	469	6	C16942	Human placenta cDNA 5'	1.06e+00
c	356	16	9.2	470	12	AA797659	vw26f04.r1 Soares mous	1.06e+00
	357	16	9.2	474	21	W73670	zd55f06.r1 Soares feta	1.06e+00
c	358	16	9.2	474	27	AQ078251	CIT-HSP-2382A18.TF CIT	1.06e+00
c	359	16	9.2	474	5	R40991	yf83g11.s1 Homo sapien	1.06e+00
c	360	16	9.2	475	13	AA608528	ae53b10.s1 Stratagene	1.06e+00
c	361	16	9.2	476	8	AA256352	zr80a09.r1 Soares NhHM	1.06e+00
c	362	16	9.2	477	27	AQ005806	CIT-HSP-2292E24.TR CIT	1.06e+00
	363	16	9.2	480	6	AA399858	vd69e09.r1 Knowles Sol	1.06e+00
	364	16	9.2	482	16	H42758	yo71a09.r1 Homo sapien	1.06e+00
c	365	16	9.2	482	15	AI022630	ox05e08.x1 Soares_feta	1.06e+00
c	366	16	9.2	482	18	AI157367	ui45h03.y1 Sugano mous	1.06e+00
c	367	16	9.2	483	28	AQ169241	HS_3171_B1_B06_MR CIT	1.06e+00
c	368	16	9.2	483	13	AA858564	UI-R-E0-bq-e-08-0-UI.s	1.06e+00
	369	16	9.2	484	27	B60256	CIT-HSP-2006F18.TRB CI	1.06e+00
	370	16	9.2	485	23	AA200864	mu09f09.r1 Soares 2NbM	1.06e+00
	371	16	9.2	485	16	H51811	yp85h09.r1 Homo sapien	1.06e+00
	372	16	9.2	486	14	AA891580	EST195383 Normalised r	1.06e+00
c	373	16	9.2	486	6	AA397483	CpEST.273 uniIAPCpIOWA	1.06e+00
	374	16	9.2	487	10	AA631263	nq81d10.s1 NCI_CGAP_Co	1.06e+00
c	375	16	9.2	487	28	AQ207363	HS_3240_B1_G11_T7 CIT	1.06e+00
c	376	16	9.2	489	11	AA688854	vs04c09.r1 Barstead mo	1.06e+00
	377	16	9.2	490	15	AI013354	EST208029 Normalised r	1.06e+00
	378	16	9.2	490	16	H30348	yp42g06.r1 Homo sapien	1.06e+00
	379	16	9.2	491	22	AA189396	mt85h02.r1 Soares mous	1.06e+00
c	380	16	9.2	491	28	AQ134766	HS_3052_B1_H10_MF CIT	1.06e+00
	381	16	9.2	492	19	N33733	yx72b06.r1 Homo sapien	1.06e+00
c	382	16	9.2	494	6	AA148521	zl06h05.s1 Soares preg	1.06e+00
c	383	16	9.2	495	14	AA926408	vr42d08.r1 Soares 2NbM	1.06e+00
c	384	16	9.2	498	9	AA606413	vo07a01.r1 Stratagene	1.06e+00
c	385	16	9.2	498	28	AQ124182	HS_3122_B2_A10_T7 CIT	1.06e+00
c	386	16	9.2	498	18	AI153668	vz89h11.r1 Soares 2NbM	1.06e+00
c	387	16	9.2	501	12	AA748765	ny10e04.s1 NCI_CGAP_GC	1.06e+00
	388	16	9.2	504	13	AA858507	UI-R-E0-bo-d-03-0-UI.s	1.06e+00
c	389	16	9.2	511	6	AA148533	zl06d05.s1 Soares preg	1.06e+00
	390	16	9.2	512	28	AQ206878	HS_3235_A2_F12_T7 CIT	1.06e+00
c	391	16	9.2	514	21	W84211	mf36h02.r1 Soares mous	1.06e+00
	392	16	9.2	515	28	AQ216057	HS_3244_B1_C09_T7 CIT	1.06e+00
	393	16	9.2	518	28	AQ110863	CIT-HSP-2376G15.TF CIT	1.06e+00
c	394	16	9.2	518	27	B59414	CIT-HSP-2017H22.TF CIT	1.06e+00
	395	16	9.2	519	17	AI119712	ue99h09.x1 Sugano mous	1.06e+00
c	396	16	9.2	521	29	AQ153091	HS_2230_A2_H11_MR CIT	1.06e+00
	397	16	9.2	523	15	AI011610	EST206061 Normalized r	1.06e+00
c	398	16	9.2	524	15	C79684	Mus musculus 3.5-dpc b	1.06e+00
	399	16	9.2	527	8	AA524887	nh34f04.s1 NCI_CGAP_Pr	1.06e+00

	400	16	9.2	529	28	AQ208036	HS_3215_A1_D03_T7 CIT	1.06e+00
c	401	16	9.2	530	28	AQ191599	HS_3237_A1_H09_T7 CIT	1.06e+00
c	402	16	9.2	531	15	AA905049	ol13b01.s1 Soares_NFL_	1.06e+00
c	403	16	9.2	532	27	B64973	CIT-HSP-2014N5.TRB CIT	1.06e+00
	404	16	9.2	533	12	AA440796	LD15672.3prime LD Dros	1.06e+00
c	405	16	9.2	534	27	AQ040578	CIT-HSP-2327A24.TF CIT	1.06e+00
c	406	16	9.2	536	6	AA404351	zv63f09.s1 Soares tota	1.06e+00
	407	16	9.2	539	29	AQ144150	HS_3076_B2_B07_MF CIT	1.06e+00
	408	16	9.2	540	8	AA524447	ng45a05.s1 NCI_CGAP_Co	1.06e+00
	409	16	9.2	544	18	AI103447	EST212736 Normalized r	1.06e+00
c	410	16	9.2	552	29	AQ225554	HS_2012_A2_B04_MR CIT	1.06e+00
c	411	16	9.2	555	23	AA110900	mp62b11.r1 Soares 2NbM	1.06e+00
	412	16	9.2	558	18	AI172660	EST137 Manduca sexta m	1.06e+00
c	413	16	9.2	563	27	B69041	CIT-HSP-2025K22.TR CIT	1.06e+00
	414	16	9.2	565	15	AI063109	GH02589.5prime GH Dros	1.06e+00
c	415	16	9.2	565	8	C22998	Japanese flounder liver	1.06e+00
	416	16	9.2	568	13	C85674	Mus musculus fertilize	1.06e+00
c	417	16	9.2	568	8	AA550193	1309m3 gmbPfHB3.1, G.	1.06e+00
	418	16	9.2	569	27	AQ004397	CIT-HSP-2293P22.TF CIT	1.06e+00
c	419	16	9.2	569	23	AA123691	mp92f01.r1 Soares 2NbM	1.06e+00
c	420	16	9.2	574	26	B16768	343G11.TVB CIT978SKA1	1.06e+00
	421	16	9.2	581	19	H91892	ys81e02.r1 Homo sapien	1.06e+00
c	422	16	9.2	581	28	AQ159214	nbxb0013K21f CUGI Rice	1.06e+00
c	423	16	9.2	582	18	AI152506	ud92a03.r1 Soares mous	1.06e+00
	424	16	9.2	585	17	AU018866	Mus musculus 8-cell em	1.06e+00
	425	16	9.2	585	17	AU018563	Mus musculus 8-cell em	1.06e+00
	426	16	9.2	596	15	AA944342	EST199841 Normalized r	1.06e+00
c	427	16	9.2	599	15	AI036291	ud60f05.x1 Sugano mous	1.06e+00
	428	16	9.2	602	18	AI177002	EST220609 Normalized r	1.06e+00
	429	16	9.2	603	26	B88587	CpG0142A CpIOWAgDNA1 C	1.06e+00
c	430	16	9.2	606	10	AA110431	mp33g04.r1 Barstead MP	1.06e+00
c	431	16	9.2	608	28	AQ160486	mgxb0005N24r CUGI Rice	1.06e+00
	432	16	9.2	615	18	AI175728	EST219293 Normalized r	1.06e+00
	433	16	9.2	615	15	AA944164	EST199663 Normalized r	1.06e+00
	434	16	9.2	619	26	FR00044436	F.rubripes GSS sequenc	1.06e+00
c	435	16	9.2	620	14	AA897889	NCC3E6T7 Conidial Neur	1.06e+00
	436	16	9.2	621	7	C25537	Dictyostelium discoide	1.06e+00
	437	16	9.2	630	27	B93873	CIT-HSP-2168H16.TR CIT	1.06e+00
	438	16	9.2	634	26	AQ010970	F23I7TFB IGF Arabidops	1.06e+00
	439	16	9.2	635	11	AA674126	SWOvL3CAN01H02SK Oncho	1.06e+00
c	440	16	9.2	640	28	AQ112391	CIT-HSP-2372A11.TF CIT	1.06e+00
c	441	16	9.2	642	6	AA127743	zk92h07.s1 Soares preg	1.06e+00
c	442	16	9.2	644	26	B72442	RPCI11-80I0.TP RPCI11	1.06e+00
	443	16	9.2	650	28	AQ109663	CIT-HSP-2374N21.TR CIT	1.06e+00
	444	16	9.2	654	26	B30347	F3A3TF IGF Arabidopsis	1.06e+00
c	445	16	9.2	657	26	B19461	T5P1-T7 TAMU Arabidops	1.06e+00
c	446	16	9.2	658	27	AQ029613	RPCI11-40C15.TJ RPCI11	1.06e+00
	447	16	9.2	660	28	AQ197155	CIT-HSP-2385M9.TR CIT-	1.06e+00
	448	16	9.2	661	17	AI110080	GH09538.5prime GH Dros	1.06e+00
	449	16	9.2	669	26	B22644	F20M2TF IGF Arabidopsi	1.06e+00
c	450	16	9.2	669	28	AQ162727	mgxb0016I11r CUGI Rice	1.06e+00

c	451	16	9.2	670	27	AQ013689	RPCI11-25L20.TKBR RPCI	1.06e+00
	452	16	9.2	672	28	AQ195310	RPCI11-61H21.TJ RPCI11	1.06e+00
c	453	16	9.2	692	26	B63320	RPCI11-2L7.TPB RPCI11	1.06e+00
	454	16	9.2	696	28	AG014328	Homo sapiens genomic D	1.06e+00
c	455	16	9.2	696	26	AG006805	Homo sapiens genomic D	1.06e+00
	456	16	9.2	700	26	AG006791	Homo sapiens genomic D	1.06e+00
	457	16	9.2	710	26	AG006790	Homo sapiens genomic D	1.06e+00
c	458	16	9.2	711	26	AG006806	Homo sapiens genomic D	1.06e+00
	459	16	9.2	713	14	C89974	Dictyostelium discoide	1.06e+00
	460	16	9.2	740	8	C23417	Jpanese flounder liver	1.05e+00
c	461	16	9.2	745	27	AQ049896	nbxb0003aA03f CUGI Ric	1.06e+00
c	462	16	9.2	750	7	C23859	Dictyostelium discoide	1.06e+00
	463	16	9.2	767	27	AQ055204	CIT-HSP-2340E20.TR CIT	1.06e+00
c	464	16	9.2	768	18	AI176842	EST220444 Normalized r	1.06e+00
c	465	16	9.2	783	28	AQ162480	mgxb0013K17r CUGI Rice	1.06e+00
	466	16	9.2	800	26	B19729	T3L14-Sp6 TAMU Arabido	1.06e+00
	467	16	9.2	802	17	AI069388	mgae0006d10f Magnapor	1.06e+00
	468	16	9.2	809	28	AF011067	Homo sapiens chromosom	1.06e+00
c	469	16	9.2	813	26	B05678	cSRL-69g2-u cSRL flow	1.06e+00
c	470	16	9.2	820	15	AI035723	ue20c12.x1 Sugano mous	1.06e+00
	471	16	9.2	876	22	AA203518	zx56c12.r1 Soares feta	1.06e+00
c	472	16	9.2	1238	26	B12133	F16D18-Sp6 IGF Arabido	1.06e+00
	473	15	8.6	94	12	AA794132	vr42g12.s1 Knowles Sol	1.46e+01
c	474	15	8.6	103	22	AA122694	mo38a06.r1 Life Tech m	1.46e+01
c	475	15	8.6	120	6	C01043	HUMGS00006885, Human Ge	1.46e+01
	476	15	8.6	122	21	H54035	EST00008 HE6W Homo sap	1.46e+01
c	477	15	8.6	133	7	HUML1727	Human fetal-lung cDNA	1.46e+01
	478	15	8.6	146	17	AU016219	Mus musculus 2-cell em	1.46e+01
c	479	15	8.6	158	22	AA089445	cchn2805.seq.F Fetal h	1.46e+01
	480	15	8.6	160	5	T38564	EST104067 Saccharomyce	1.46e+01
	481	15	8.6	163	11	AA677367	zj60g01.s1 Soares feta	1.46e+01
	482	15	8.6	166	19	N73570	za50g03.s1 Homo sapien	1.46e+01
	483	15	8.6	171	23	AA204600	mu25b06.r1 Soares 2NbM	1.46e+01
	484	15	8.6	173	25	AA338388	EST43378 Fetal brain I	1.46e+01
c	485	15	8.6	174	7	AA508932	MBAPCW1D02T3 Brugia ma	1.46e+01
c	486	15	8.6	182	10	HSCZOE022	H. sapiens partial cDN	1.46e+01
	487	15	8.6	189	12	AA813668	ai69d07.s1 Soares test	1.46e+01
c	488	15	8.6	190	27	AQ043735	CIT-HSP-2329I7.TR CIT-	1.46e+01
	489	15	8.6	193	5	T00753	wEST01474 Caenorhabdit	1.46e+01
c	490	15	8.6	194	17	AA902150	ok77g08.s1 NCI_CGAP_GC	1.46e+01
c	491	15	8.6	197	19	N42360	yy14h10.r1 Homo sapien	1.46e+01
c	492	15	8.6	209	19	N38990	yv21f08.s1 Homo sapien	1.46e+01
	493	15	8.6	215	5	T37935	EST103250 Saccharomyce	1.46e+01
	494	15	8.6	217	18	AI176377	EST219961 Normalized r	1.46e+01
	495	15	8.6	219	8	AA215588	zr95b12.r1 NCI_CGAP_GC	1.46e+01
c	496	15	8.6	219	22	N97436	0767H7 gmbPfHB3.1, G.	1.46e+01
	497	15	8.6	221	8	AA235949	zs40g04.s1 Soares NhHM	1.46e+01
	498	15	8.6	221	20	ATTS3391	A. thaliana transcribe	1.46e+01
	499	15	8.6	222	14	AA924006	UI-R-A1-dq-h-09-0-UI.s	1.46e+01
	500	15	8.6	223	19	HSC1XB032	H. sapiens partial cDN	1.46e+01
	501	15	8.6	229	15	AI026453	TENU0664 T. crusi epim	1.46e+01

	502	15	8.6	230	20	HSC04D062	H. sapiens partial cDN	1.46e+01
	503	15	8.6	233	11	AA693378	ah21g05.s1 Soares para	1.46e+01
	504	15	8.6	233	19	HSC2BH122	H. sapiens partial cDN	1.46e+01
c	505	15	8.6	233	7	HUM328C02B	Human aorta cDNA 5'-en	1.46e+01
c	506	15	8.6	238	19	H50521	yr44e12.s1 Homo sapien	1.46e+01
	507	15	8.6	239	23	AA123752	aq75f09.r1 Stratagene	1.46e+01
c	508	15	8.6	241	5	R07847	yf16a04.s1 Homo sapien	1.46e+01
c	509	15	8.6	243	9	C47209	C.elegans cDNA clone y	1.46e+01
	510	15	8.6	244	12	AA788972	ah29b02.s1 Soares para	1.46e+01
c	511	15	8.6	245	5	R30877	yh59h06.s1 Homo sapien	1.46e+01
	512	15	8.6	246	17	AI055888	ow30c07.s1 Soares_para	1.46e+01
	513	15	8.6	248	14	AA907102	ol03c08.s1 NCI_CGAP_Lu	1.46e+01
	514	15	8.6	248	18	AI079619	ox91d11.s1 Soares_sene	1.46e+01
	515	15	8.6	251	7	HUM322H05B	Human aorta cDNA 5'-en	1.46e+01
	516	15	8.6	256	21	W04583	zb94g01.s1 Soares para	1.46e+01
	517	15	8.6	258	8	AA563217	vl65e01.r1 Knowles Sol	1.46e+01
c	518	15	8.6	258	15	AA996791	UI-R-C0-hh-f-11-0-UI.s	1.46e+01
	519	15	8.6	259	6	W90386	zh75f06.s1 Soares feta	1.46e+01
	520	15	8.6	265	16	H14639	yl26f07.s1 Homo sapien	1.46e+01
c	521	15	8.6	265	17	AI022997	ow57d05.s1 Soares_para	1.46e+01
c	522	15	8.6	265	5	T19620	761F Homo sapiens cDNA	1.46e+01
	523	15	8.6	267	5	R39312	yd01d01.s1 Homo sapien	1.46e+01
	524	15	8.6	267	8	AA233619	zr43d07.s1 Soares NhHM	1.46e+01
c	525	15	8.6	270	19	N38976	yv21d08.s1 Homo sapien	1.46e+01
	526	15	8.6	271	19	H92957	yt94g09.r1 Homo sapien	1.46e+01
	527	15	8.6	271	18	AA878551	oj19f12.s1 NCI_CGAP_Ki	1.46e+01
c	528	15	8.6	272	14	AA837790	oe06g12.s1 NCI_CGAP_Ov	1.46e+01
	529	15	8.6	276	7	HUM170H08B	Human fetal brain cDNA	1.46e+01
	530	15	8.6	276	26	B67488	T25M8TR TAMU Arabidops	1.46e+01
c	531	15	8.6	276	16	H21924	yl29g08.s1 Homo sapien	1.46e+01
	532	15	8.6	277	10	AA659938	JM00M225.QM3 Miracidia	1.46e+01
	533	15	8.6	278	27	AQ046869	RPCI11-33M12.TK RPCI11	1.46e+01
c	534	15	8.6	280	9	AA551418	nj55a10.s1 NCI_CGAP_Pr	1.46e+01
	535	15	8.6	281	15	AA971806	op87b04.s1 Soares_NFL_	1.46e+01
	536	15	8.6	286	16	T30487	EST17476 Homo sapiens	1.46e+01
c	537	15	8.6	286	17	AI122378	uc61d06.r1 Soares mous	1.46e+01
c	538	15	8.6	290	19	HSC2TF012	H. sapiens partial cDN	1.46e+01
c	539	15	8.6	291	23	AA076880	7B04E09 Chromosome 7 P	1.46e+01
	540	15	8.6	292	5	T00425	wEST01146 Caenorhabdit	1.46e+01
c	541	15	8.6	293	29	AQ230101	HS_2034_A2_C03_T7 CIT	1.46e+01
c	542	15	8.6	295	25	AA371433	EST83396 Platelet Homo	1.46e+01
c	543	15	8.6	296	21	AA007220	13cDNA56-3.seq Soares	1.46e+01
c	544	15	8.6	296	11	AA678774	ah04c02.s1 Gessler Wil	1.46e+01
	545	15	8.6	297	5	R51240	yg70b10.s1 Homo sapien	1.46e+01
	546	15	8.6	299	12	AA721428	ns74d05.s1 NCI_CGAP_GC	1.46e+01
	547	15	8.6	299	17	AA918549	ol42e02.s1 Soares_NFL_	1.46e+01
c	548	15	8.6	299	29	AQ228585	HS_2009_A2_D09_T7 CIT	1.46e+01
	549	15	8.6	300	9	C51442	C.elegans cDNA clone y	1.46e+01
	550	15	8.6	300	9	C51968	C.elegans cDNA clone y	1.46e+01
	551	15	8.6	300	9	C52220	C.elegans cDNA clone y	1.46e+01
	552	15	8.6	300	9	C52768	C.elegans cDNA clone y	1.46e+01

	553	15	8.6	300	9	C53184	C.elegans cDNA clone y	1.46e+01
	554	15	8.6	300	9	C55185	C.elegans cDNA clone y	1.46e+01
	555	15	8.6	300	9	C55143	C.elegans cDNA clone y	1.46e+01
	556	15	8.6	300	9	C55973	C.elegans cDNA clone y	1.46e+01
c	557	15	8.6	300	23	AA249346	j1231.seq.F Human feta	1.46e+01
c	558	15	8.6	300	27	AQ066502	HS_2247_A1_H04_MF CIT	1.46e+01
c	559	15	8.6	300	9	C33428	C.elegans cDNA clone y	1.46e+01
	560	15	8.6	301	19	H65561	yz67h01.r1 Homo sapien	1.46e+01
	561	15	8.6	301	5	T23585	seq3909 Homo sapiens c	1.46e+01
	562	15	8.6	302	13	AA885435	aml4g10.s1 Soares NFL	1.46e+01
	563	15	8.6	303	16	M85704	EST02222 Homo sapiens	1.46e+01
	564	15	8.6	304	19	HSC1XC032	H. sapiens partial cDN	1.46e+01
	565	15	8.6	304	14	H33681	EST109908 Rat PC-12 ce	1.46e+01
	566	15	8.6	304	5	M79772	wEST00309 Caenorhabdit	1.46e+01
	567	15	8.6	306	5	M79740	wEST00277 Caenorhabdit	1.46e+01
c	568	15	8.6	306	19	HSC1UG081	H. sapiens partial cDN	1.46e+01
c	569	15	8.6	307	16	T34928	EST76889 Homo sapiens	1.46e+01
	570	15	8.6	308	19	HSU25949	Human mRNA clone hsc25	1.46e+01
c	571	15	8.6	308	5	R38402	yc96h12.s1 Homo sapien	1.46e+01
	572	15	8.6	309	6	C21233	HUMGS0002268, Human Ge	1.46e+01
	573	15	8.6	311	5	R06949	yf12e07.r1 Homo sapien	1.46e+01
c	574	15	8.6	311	20	HSB37H102	H. sapiens partial cDN	1.46e+01
c	575	15	8.6	312	19	H84974	ys89e03.r1 Homo sapien	1.46e+01
c	576	15	8.6	312	16	R78885	yi83d07.s1 Homo sapien	1.46e+01
c	577	15	8.6	316	18	AI137544	UI-R-C2p-oc-g-03-0-UI.	1.46e+01
c	578	15	8.6	317	8	AA549701	vk73a01.s1 Knowles Sol	1.46e+01
	579	15	8.6	319	15	AA896501	UI-R-C0-he-b-04-0-UI.s	1.46e+01
c	580	15	8.6	319	5	T10264	seq950 Homo sapiens cD	1.46e+01
	581	15	8.6	323	28	AQ093613	HS_3022_A1_A08_MR CIT	1.46e+01
c	582	15	8.6	324	28	AQ111334	CIT-HSP-2382A5.TF CIT-	1.46e+01
c	583	15	8.6	325	10	AA584087	no04g11.s1 NCI_CGAP_Ph	1.46e+01
	584	15	8.6	326	10	AA582770	nn39d08.s1 NCI_CGAP_GC	1.46e+01
	585	15	8.6	328	10	AA435497	zt72a12.s1 Soares test	1.46e+01
	586	15	8.6	331	19	H72653	yu05d12.s1 Homo sapien	1.46e+01
c	587	15	8.6	331	12	AA747371	nx76g03.s1 NCI_CGAP_Ew	1.46e+01
	588	15	8.6	333	20	ATTS1855	A. thaliana transcribe	1.46e+01
	589	15	8.6	333	15	AA953291	oo87h09.s1 NCI_CGAP_Ki	1.46e+01
	590	15	8.6	335	14	AA974458	oq37h02.s1 NCI_CGAP_Ki	1.46e+01
c	591	15	8.6	335	16	H53730	yu38g12.r1 Homo sapien	1.46e+01
	592	15	8.6	336	13	AA883171	aml7f11.s1 Soares NFL	1.46e+01
	593	15	8.6	336	10	AA601750	no03e12.s1 NCI_CGAP_Ph	1.46e+01
	594	15	8.6	337	26	FR0013468	F.rubripes GSS sequenc	1.46e+01
c	595	15	8.6	337	19	HSC2CD012	H. sapiens partial cDN	1.46e+01
	596	15	8.6	338	16	H12724	yj13e11.r1 Homo sapien	1.46e+01
c	597	15	8.6	339	7	CELK099F8R	C.elegans cDNA clone y	1.46e+01
	598	15	8.6	340	5	T71280	yd15d08.s1 Homo sapien	1.46e+01
	599	15	8.6	341	16	H00448	yj24c01.s1 Homo sapien	1.46e+01
	600	15	8.6	341	18	AI122977	ok32f01.s1 Soares_NSF_	1.46e+01
c	601	15	8.6	342	14	AA919732	vy16d03.r1 Stratagene	1.46e+01
	602	15	8.6	344	6	AA438156	ve64e08.r1 Beddington	1.46e+01
	603	15	8.6	345	19	HSC2UH032	H. sapiens partial cDN	1.46e+01

c	604	15	8.6	348	14	AA887354	L30-326T3 Ice plant La	1.46e+01
	605	15	8.6	350	17	AI001751	ot38g01.s1 Soares_test	1.46e+01
	606	15	8.6	350	17	AA992175	ot80c03.s1 Soares_tota	1.46e+01
	607	15	8.6	351	24	AA298703	EST114310 Synovial sar	1.46e+01
	608	15	8.6	354	19	H60003	yz27d01.s1 Homo sapien	1.46e+01
c	609	15	8.6	355	15	AA931341	oo06d12.s1 Soares_NFL_	1.46e+01
	610	15	8.6	357	20	T15396	IB1080 Infant brain, B	1.46e+01
c	611	15	8.6	357	16	H37965	yp47f03.r1 Homo sapien	1.46e+01
	612	15	8.6	358	15	AI044650	UI-R-C1-jx-c-12-0-UI.s	1.46e+01
c	613	15	8.6	358	15	AA969616	op38f10.s1 Soares_NFL_	1.46e+01
c	614	15	8.6	360	9	C61354	C.elegans cDNA clone y	1.46e+01
c	615	15	8.6	360	12	AA812492	ai73c08.s1 Soares test	1.46e+01
c	616	15	8.6	360	9	C40051	C.elegans cDNA clone y	1.46e+01
	617	15	8.6	362	17	AI022605	ow49h10.s1 Soares_para	1.46e+01
	618	15	8.6	363	16	R59762	yh07c03.r1 Homo sapien	1.46e+01
c	619	15	8.6	365	18	AI174478	an42b01.s1 Gessler Wil	1.46e+01
c	620	15	8.6	366	16	H54095	yq90a01.r1 Homo sapien	1.46e+01
c	621	15	8.6	367	14	AA889508	aj95c09.s1 Soares_para	1.46e+01
	622	15	8.6	368	12	AA757066	ah55c01.s1 Soares test	1.46e+01
	623	15	8.6	368	16	R67589	yi34e05.s1 Homo sapien	1.46e+01
c	624	15	8.6	368	5	T93043	ye27c12.r1 Homo sapien	1.46e+01
	625	15	8.6	369	29	AQ242473	HS_2057_B1_D04_MR CIT	1.46e+01
c	626	15	8.6	369	8	AA560575	vl20d02.r1 Stratagene	1.46e+01
	627	15	8.6	370	5	R00076	ye73b05.s1 Homo sapien	1.46e+01
	628	15	8.6	370	15	AA997502	UI-R-C0-hw-f-03-0-UI.s	1.46e+01
c	629	15	8.6	371	13	AA626888	ag30g07.s1 Jia bone ma	1.46e+01
	630	15	8.6	372	19	N49095	yy79e07.s1 Homo sapien	1.46e+01
	631	15	8.6	372	12	AA790054	vt78h02.r1 Barstead mo	1.46e+01
c	632	15	8.6	372	21	AA063698	T3322 MVAT4 bloodstrea	1.46e+01
	633	15	8.6	373	28	AQ093946	HS_2199_A2_A03_MR CIT	1.46e+01
c	634	15	8.6	373	26	B47627	CIT-HSP-340G23.TVB CIT	1.46e+01
c	635	15	8.6	375	27	AQ035565	CIT-HSP-2319L5.TF CIT-	1.46e+01
c	636	15	8.6	376	9	C61264	C.elegans cDNA clone y	1.46e+01
c	637	15	8.6	376	5	T98233	ye30f12.r1 Homo sapien	1.46e+01
c	638	15	8.6	377	27	AQ001603	CIT-HSP-2284M22.TR CIT	1.46e+01
	639	15	8.6	378	23	T02909	FB15C1 Fetal brain, St	1.46e+01
c	640	15	8.6	378	27	B69578	CIT-HSP-2054O14.TF CIT	1.46e+01
c	641	15	8.6	378	19	H85300	yv86g05.r1 Homo sapien	1.46e+01
c	642	15	8.6	379	5	R13363	yf76d07.r1 Homo sapien	1.46e+01
	643	15	8.6	380	8	AA398441	zt62e05.s1 Soares test	1.46e+01
	644	15	8.6	380	27	AQ031184	HS_2198_A2_G01_MR CIT	1.46e+01
c	645	15	8.6	382	17	AI040415	ow10d12.x1 Soares_para	1.46e+01
	646	15	8.6	383	7	HUM303A02B	Human aorta cDNA 5'-en	1.46e+01
	647	15	8.6	383	17	T76254	11032 Lambda-PRL2 Arab	1.46e+01
c	648	15	8.6	383	8	AA503289	ng21a07.s1 NCI_CGAP_Ov	1.46e+01
	649	15	8.6	384	15	AI045492	UI-R-C1-kj-c-12-0-UI.s	1.46e+01
c	650	15	8.6	384	28	AQ207022	HS_3242_A2_D07_MR CIT	1.46e+01
c	651	15	8.6	385	19	N75601	za85h06.s1 Homo sapien	1.46e+01
c	652	15	8.6	385	16	R67351	yi32h02.s1 Homo sapien	1.46e+01
	653	15	8.6	388	17	AI117828	uc41a05.r1 Soares mous	1.46e+01
	654	15	8.6	389	11	AA689802	vt61e11.r1 Barstead mo	1.46e+01

c	655	15	8.6	389	17	AI025889	ov95a03.s1 Soares_test	1.46e+01
	656	15	8.6	390	27	B70218	CIT-HSP-2028G11.TR CIT	1.46e+01
c	657	15	8.6	390	16	M85548	EST02064 Homo sapiens	1.46e+01
	658	15	8.6	391	15	AI061330	an36e11.x1 Gessler Wil	1.46e+01
	659	15	8.6	391	16	R78137	yi80g09.r1 Homo sapien	1.46e+01
	660	15	8.6	391	29	AQ151884	HS_3096_B2_H01_MR CIT	1.46e+01
	661	15	8.6	394	10	AA427162	ve79c03.r1 Soares mous	1.46e+01
c	662	15	8.6	394	29	AQ147044	HS_2269_A1_G06_MF CIT	1.46e+01
c	663	15	8.6	395	13	AA823224	vw42b10.r1 Soares mous	1.46e+01
	664	15	8.6	396	8	AA465861	ve88f04.s1 Knowles Sol	1.46e+01
c	665	15	8.6	397	28	AQ190811	HS_3231_B2_A02_MR CIT	1.46e+01
	666	15	8.6	398	21	T19455	c03016t Testis 1 Homo	1.46e+01
c	667	15	8.6	398	20	W85710	zh55c06.s1 Soares feta	1.46e+01
c	668	15	8.6	398	28	AQ123521	HS_3101_B1_F09_MR CIT	1.46e+01
c	669	15	8.6	399	14	AA888277	nw78a09.s1 NCI_CGAP_Pr	1.46e+01
	670	15	8.6	400	13	AA860906	ak22b08.s1 Soares test	1.46e+01
	671	15	8.6	400	16	H41859	yo07g06.r1 Homo sapien	1.46e+01
	672	15	8.6	401	29	AQ241882	CITBI-E1-2501021.TR.1	1.46e+01
	673	15	8.6	401	12	AA778820	zj38b02.s1 Soares feta	1.46e+01
	674	15	8.6	401	21	W49803	zc44f06.s1 Soares sene	1.46e+01
c	675	15	8.6	401	27	AQ045667	RPCI11-34L22.TJ RPCI11	1.46e+01
c	676	15	8.6	401	10	AA602142	no98d01.s1 NCI_CGAP_Pr	1.46e+01
	677	15	8.6	402	26	B33806	HS-1023-A2-D08-MF.abi	1.46e+01
	678	15	8.6	402	11	AA671412	vn81f02.r1 Knowles Sol	1.46e+01
c	679	15	8.6	402	19	H44834	yy39e09.r1 Homo sapien	1.46e+01
c	680	15	8.6	403	27	AQ066827	HS_2209_B1_G07_MR CIT	1.46e+01
	681	15	8.6	404	26	AQ016811	CIT-HSP-2308F9.TF CIT-	1.46e+01
	682	15	8.6	404	8	AA290772	zt23g12.s1 Soares ovar	1.46e+01
	683	15	8.6	404	13	C86089	Mus musculus fertilize	1.46e+01
c	684	15	8.6	404	15	AA948286	oq34c08.s1 NCI_CGAP_GC	1.46e+01
	685	15	8.6	405	17	AI033057	ow93e02.s1 Soares_feta	1.46e+01
c	686	15	8.6	406	8	AA410860	zt36e02.r1 Soares ovar	1.46e+01
	687	15	8.6	407	17	AI038307	ox48b10.x1 Soares_tota	1.46e+01
	688	15	8.6	407	10	AA417922	zv94b04.s1 Soares NhHM	1.46e+01
	689	15	8.6	408	13	AA875622	UI-R-E0-ctr-e-03-0-UI.s	1.46e+01
c	690	15	8.6	408	28	AQ188376	HS_3228_A2_E04_MR CIT	1.46e+01
c	691	15	8.6	409	25	AA351683	EST59784 Infant brain	1.46e+01
	692	15	8.6	410	27	AQ066836	HS_2209_B1_H07_MR CIT	1.46e+01
c	693	15	8.6	410	8	AA555958	vi64g11.r1 Beddington	1.46e+01
	694	15	8.6	411	22	AA184879	mu46a01.r1 Soares mous	1.46e+01
	695	15	8.6	411	26	AQ016712	CIT-HSP-2309L10.TR CIT	1.46e+01
c	696	15	8.6	411	28	AQ093473	HS_3018_B1_D10_T7 CIT	1.46e+01
	697	15	8.6	412	8	AA226493	nc20e02.r1 NCI_CGAP_Pr	1.46e+01
	698	15	8.6	412	13	AA861304	ak14g01.s1 Soares para	1.46e+01
c	699	15	8.6	412	28	AQ123987	HS_3107_B2_C05_MR CIT	1.46e+01
c	700	15	8.6	412	23	AA124914	mp76c02.r1 Soares 2NbM	1.46e+01
	701	15	8.6	413	15	AI001998	ot42e05.s1 Soares_test	1.46e+01
	702	15	8.6	413	28	AQ195193	RPCI11-48C5.TK RPCI11	1.46e+01
	703	15	8.6	413	28	AQ131942	HS_2255_A2_C10_MF CIT	1.46e+01
c	704	15	8.6	413	13	AA598913	ae41c06.s1 Gessler Wil	1.46e+01
	705	15	8.6	414	16	H19414	ym46b10.r1 Homo sapien	1.46e+01

c	706	15	8.6	415	27	AQ043736	CIT-HSP-2327P2.TR CIT-	1.46e+01
c	707	15	8.6	415	9	AA531420	nj62d09.s1 NCI_CGAP_Pr	1.46e+01
c	708	15	8.6	416	11	AA728479	33658 CD4-6 Arabidopsi	1.46e+01
c	709	15	8.6	417	10	AA160890	zo79b08.s1 Stratagene	1.46e+01
	710	15	8.6	418	6	CELK013G5R	C.elegans cDNA clone y	1.46e+01
	711	15	8.6	418	10	AA418692	zv98a05.s1 Soares NhHM	1.46e+01
c	712	15	8.6	418	12	AA740724	ny98c07.s1 NCI_CGAP_GC	1.46e+01
	713	15	8.6	419	5	T71109	yc50f10.r1 Homo sapien	1.46e+01
	714	15	8.6	419	12	AA789404	vv93h06.r1 Stratagene	1.46e+01
c	715	15	8.6	419	11	AA668692	ac47d12.s1 Stratagene	1.46e+01
c	716	15	8.6	420	5	T84104	yd46f07.s1 Homo sapien	1.46e+01
c	717	15	8.6	421	26	AQ011740	HS_2260_A1_MF_C07 CIT	1.46e+01
	718	15	8.6	422	18	AI087318	ox45f04.x1 Soares_NhHM	1.46e+01
	719	15	8.6	424	8	AA099306	z179a08.s1 Stratagene	1.46e+01
	720	15	8.6	425	28	AQ175306	HS_3212_B2_A11_MR CIT	1.46e+01
c	721	15	8.6	426	19	N23333	yx71a09.s1 Homo sapien	1.46e+01
	722	15	8.6	427	6	AA149166	z142a09.r1 Soares preg	1.46e+01
	723	15	8.6	427	27	AQ035426	CIT-HSP-2331G11.TR CIT	1.46e+01
	724	15	8.6	427	18	AI128743	qa63b09.s1 Soares_feta	1.46e+01
c	725	15	8.6	427	7	AA455972	aa02a08.s1 Soares NhHM	1.46e+01
c	726	15	8.6	427	11	AA707271	ah36b12.s1 Soares test	1.46e+01
	727	15	8.6	428	6	AA398477	zt63a09.s1 Soares test	1.46e+01
	728	15	8.6	429	15	AI011339	EST205790 Normalized r	1.46e+01
	729	15	8.6	429	29	AQ145616	HS_2199_B2_H08_MR CIT	1.46e+01
	730	15	8.6	430	20	N81035	zb04e01.s1 Homo sapien	1.46e+01
	731	15	8.6	430	6	AA436239	zv23h10.s1 Soares NhHM	1.46e+01
c	732	15	8.6	430	14	AA902130	ok94f04.s1 NCI_CGAP_Lu	1.46e+01
	733	15	8.6	431	29	AQ250457	T6E14-T7 TAMU Arabidop	1.46e+01
	734	15	8.6	432	11	AA553625	nk82b08.s1 NCI_CGAP_Sc	1.46e+01
	735	15	8.6	432	16	R84407	yp89g09.s1 Homo sapien	1.46e+01
	736	15	8.6	432	29	AQ230973	HS_2037_B2_C06_MR CIT	1.46e+01
	737	15	8.6	433	27	AQ024394	HS_2184_A1_A09_MR CIT	1.46e+01
c	738	15	8.6	433	20	WO0438	yz99c05.r1 Homo sapien	1.46e+01
	739	15	8.6	434	15	AA998223	UI-R-C0-ib-d-08-0-UI.s	1.46e+01
	740	15	8.6	434	24	AA268690	va45d06.r1 Soares mous	1.46e+01
c	741	15	8.6	434	6	AA001615	zh82e08.s1 Soares feta	1.46e+01
	742	15	8.6	436	12	AA738066	nx09h10.s1 NCI_CGAP_GC	1.46e+01
	743	15	8.6	436	7	HUM088A04B	Human fetal brain cDNA	1.46e+01
	744	15	8.6	436	29	AQ146482	HS_2229_A1_B05_MF CIT	1.46e+01
c	745	15	8.6	436	6	AA037841	zk38d12.s1 Soares preg	1.46e+01
c	746	15	8.6	436	27	AQ056388	CIT-HSP-2341L9.TR CIT-	1.46e+01
	747	15	8.6	437	16	R45589	yg44g07.s1 Homo sapien	1.46e+01
	748	15	8.6	437	13	AA859691	UI-R-E0-bx-e-12-0-UI.s	1.46e+01
	749	15	8.6	437	16	H18102	yn47d05.s1 Homo sapien	1.46e+01
	750	15	8.6	437	12	AA713499	nv67c07.s1 NCI_CGAP_GC	1.46e+01
	751	15	8.6	437	26	B78209	T31K23TF TAMU Arabidop	1.46e+01
	752	15	8.6	438	10	T43220	6483 Lambda-PRL2 Arabi	1.46e+01
	753	15	8.6	438	8	AA543161	vj81g02.r1 Soares mous	1.46e+01
c	754	15	8.6	438	15	AA856729	od54c06.s1 NCI_CGAP_GC	1.46e+01
c	755	15	8.6	439	28	AQ109736	CIT-HSP-2372O19.TR CIT	1.46e+01
	756	15	8.6	440	17	AI071211	UI-R-C2-my-d-12-0-UI.s	1.46e+01

c	757	15	8.6	440	26	B39843	HS-1050-B1-D08-MF.abi 1.46e+01
c	758	15	8.6	440	28	AQ128542	HS_2252_A1_E11_MR CIT 1.46e+01
	759	15	8.6	441	26	B56797	CIT-HSP-2007N23.TF CIT 1.46e+01
	760	15	8.6	441	13	AA854231	aj72c06.s1 Soares para 1.46e+01
c	761	15	8.6	441	13	AA823234	vw42c10.r1 Soares mous 1.46e+01
	762	15	8.6	444	15	C82923	Oryctolagus cuniculus 1.46e+01
c	763	15	8.6	444	15	C83779	Oryctolagus cuniculus 1.46e+01
	764	15	8.6	446	28	AQ129680	HS_2252_A1_F04_MR CIT 1.46e+01
	765	15	8.6	446	5	R53937	yg83b04.s1 Homo sapien 1.46e+01
	766	15	8.6	447	17	AI097131	qb60g08.x1 NCI_CGAP_Br 1.46e+01
	767	15	8.6	448	20	W47810	mc90c02.r1 Soares mous 1.46e+01
c	768	15	8.6	448	5	R05549	yf87c01.r1 Homo sapien 1.46e+01
	769	15	8.6	449	6	AA001918	zh83b07.s1 Soares feta 1.46e+01
c	770	15	8.6	450	19	N63950	yz81a08.s1 Homo sapien 1.46e+01
c	771	15	8.6	450	6	AA115264	sl08c12.s1 Soares preg 1.46e+01
	772	15	8.6	451	8	AA281660	zt06g01.s1 NCI_CGAP_GC 1.46e+01
	773	15	8.6	451	18	AI038752	ox38g05.s1 Soares_tota 1.46e+01
c	774	15	8.6	451	5	R13132	yf72b12.r1 Homo sapien 1.46e+01
c	775	15	8.6	451	15	AA997997	UI-R-C0-hu-g-05-0-UI.s 1.46e+01
c	776	15	8.6	451	16	H30222	yp41h12.s1 Homo sapien 1.46e+01
c	777	15	8.6	451	18	AI104810	EST214099 Normalized r 1.46e+01
	778	15	8.6	452	8	AA236093	za05c05.r1 NCI_CGAP_GC 1.46e+01
c	779	15	8.6	452	10	AA634558	zu82e01.s1 Soares test 1.46e+01
c	780	15	8.6	452	17	AI021920	ow96d04.x1 Soares_feta 1.46e+01
c	781	15	8.6	453	15	AI011961	EST206412 Normalized r 1.46e+01
	782	15	8.6	454	11	AA070327	zm68e03.r1 Stratagene 1.46e+01
	783	15	8.6	454	22	AA022723	ze73e07.s1 Soares feta 1.46e+01
c	784	15	8.6	454	7	AA455338	aa02g08.s1 Soares NhHM 1.46e+01
c	785	15	8.6	454	12	AA775271	ad18d08.s1 Soares NbHF 1.46e+01
c	786	15	8.6	455	24	AA266381	mz67c05.r1 Soares mous 1.46e+01
	787	15	8.6	456	22	AA024488	ze76b01.s1 Soares feta 1.46e+01
c	788	15	8.6	456	8	AA455696	aa22c06.r1 NCI_CGAP_GC 1.46e+01
c	789	15	8.6	457	17	AI127704	qc26f04.x1 Soares_preg 1.46e+01
c	790	15	8.6	457	17	AI127702	qc26f02.x1 Soares_preg 1.46e+01
	791	15	8.6	458	11	AA076014	zm89g10.s1 Stratagene 1.46e+01
	792	15	8.6	459	18	AI156679	ue53f10.r1 Soares mous 1.46e+01
	793	15	8.6	459	16	H46955	yo19d11.s1 Homo sapien 1.46e+01
c	794	15	8.6	459	12	AA743637	ny24f09.s1 NCI_CGAP_GC 1.46e+01
c	795	15	8.6	459	11	AA657542	nt66g04.s1 NCI_CGAP_Pr 1.46e+01
c	796	15	8.6	460	19	N70270	za57h12.s1 Homo sapien 1.46e+01
	797	15	8.6	461	18	AI138306	qd85e01.x1 Soares_test 1.46e+01
c	798	15	8.6	462	12	AA793151	vp17h11.r1 Soares mous 1.46e+01
	799	15	8.6	463	12	AA201216	LD04144.5prime LD Dros 1.46e+01
c	800	15	8.6	463	19	N25825	yz22c04.r1 Homo sapien 1.46e+01
c	801	15	8.6	463	17	AI126367	qc51c05.x1 Soares_preg 1.46e+01
c	802	15	8.6	463	18	AI105118	EST214407 Normalized r 1.46e+01
	803	15	8.6	464	17	AI026997	ow41d05.s1 Soares_para 1.46e+01
	804	15	8.6	465	27	AQ042468	CIT-HSP-2317P21.TF CIT 1.46e+01
	805	15	8.6	465	29	AQ227726	HS_2012_B2_B12_MR CIT 1.46e+01
c	806	15	8.6	465	15	AA913508	ol37h04.s1 Soares_NFL_ 1.46e+01
	807	15	8.6	467	19	H57562	yr05g11.r1 Homo sapien 1.46e+01

c	808	15	8.6	467	8	AA278557	zs76d04.r1 NCI_CGAP_GC 1.46e+01
c	809	15	8.6	467	5	T77406	yc94c01.r1 Homo sapien 1.46e+01
	810	15	8.6	468	22	CHESTM018	C.hircus expressed seq 1.46e+01
	811	15	8.6	468	15	AI044859	UI-R-C1-kd-c-11-0-UI.s 1.46e+01
	812	15	8.6	469	15	AA996459	UI-R-C0-hi-f-08-0-UI.s 1.46e+01
c	813	15	8.6	470	21	W49485	zc42d11.s1 Soares sene 1.46e+01
	814	15	8.6	471	12	AA391934	LD10857.5prime LD Dros 1.46e+01
	815	15	8.6	472	13	AA830650	oc52g02.s1 NCI_CGAP_GC 1.46e+01
	816	15	8.6	472	16	H06306	y179a01.r1 Homo sapien 1.46e+01
	817	15	8.6	472	21	N93650	zb50b03.s1 Soares feta 1.46e+01
	818	15	8.6	473	12	AA812755	aj43g06.s1 Soares test 1.46e+01
	819	15	8.6	473	15	AA936412	on65a09.s1 Soares_NFL_ 1.46e+01
	820	15	8.6	473	14	AA920349	vx94g06.r1 Stratagene 1.46e+01
c	821	15	8.6	473	11	AA700893	zj44a01.s1 Soares feta 1.46e+01
c	822	15	8.6	473	10	AA420220	wf52e05.r1 Soares mous 1.46e+01
	823	15	8.6	474	17	AI080596	oz67g07.x1 Soares_sene 1.46e+01
	824	15	8.6	475	8	AA207144	zr87h04.r1 NCI_CGAP_GC 1.46e+01
	825	15	8.6	475	16	R55730	yg89a07.r1 Homo sapien 1.46e+01
	826	15	8.6	476	11	AA682337	ae91f02.s1 Stratagene 1.46e+01
	827	15	8.6	477	22	AA009489	ze83e01.s1 Soares feta 1.46e+01
	828	15	8.6	477	18	AI140674	qe06a07.x1 Soares_test 1.46e+01
c	829	15	8.6	477	19	H57439	yr05a06.r1 Homo sapien 1.46e+01
c	830	15	8.6	477	19	H82820	yq48f06.s1 Homo sapien 1.46e+01
c	831	15	8.6	478	17	AI071668	UI-R-C1-ku-d-11-0-UI.s 1.46e+01
	832	15	8.6	479	8	AA258052	zs76b03.s1 NCI_CGAP_GC 1.46e+01
	833	15	8.6	483	21	R85851	yq27e11.s1 Soares feta 1.46e+01
c	834	15	8.6	483	19	N70455	za62g10.s1 Homo sapien 1.46e+01
c	835	15	8.6	484	19	H72652	yu05d12.r1 Homo sapien 1.46e+01
c	836	15	8.6	484	29	AQ140352	HS_3108_A2_F02_T7 CIT 1.46e+01
c	837	15	8.6	485	15	AA894210	EST198013 Normalized r 1.46e+01
c	838	15	8.6	486	11	AA693510	ah33e02.s1 Soares test 1.46e+01
	839	15	8.6	487	15	AI008100	EST202551 Normalized r 1.46e+01
c	840	15	8.6	487	17	AA992561	ot89g08.s1 Soares_tota 1.46e+01
	841	15	8.6	488	28	AQ212512	HS_3108_B2_D05_T7 CIT 1.46e+01
	842	15	8.6	489	19	N72892	za04e05.s1 Homo sapien 1.46e+01
	843	15	8.6	490	12	AA818707	UI-R-A0-az-d-03-0-UI.s 1.46e+01
c	844	15	8.6	491	19	H68828	yr83g07.s1 Homo sapien 1.46e+01
c	845	15	8.6	491	10	AA629395	zu78d05.s1 Soares test 1.46e+01
	846	15	8.6	493	28	AQ209796	HS_3244_A1_D09_T7 CIT 1.46e+01
c	847	15	8.6	493	29	AQ144271	HS_3085_A2_H06_MF CIT 1.46e+01
c	848	15	8.6	493	24	AA268443	va98g10.r1 Soares mous 1.46e+01
c	849	15	8.6	496	13	AA843181	aj15f11.s1 Soares para 1.46e+01
c	850	15	8.6	497	26	B35476	HS-1029-A1-F07-MF.abi 1.46e+01
	851	15	8.6	498	10	AA593858	nn19f01.s1 NCI_CGAP_Co 1.46e+01
c	852	15	8.6	498	7	AA462472	vg67h11.r1 Soares mous 1.46e+01
c	853	15	8.6	498	13	AA158104	zo57g12.r1 Stratagene 1.46e+01
	854	15	8.6	499	12	AA745246	ny63f06.s1 NCI_CGAP_GC 1.46e+01
c	855	15	8.6	499	26	AQ020930	CIT-HSP-2309M12.TR CIT 1.46e+01
c	856	15	8.6	499	20	AA000249	mg33h10.r1 Soares mous 1.46e+01
	857	15	8.6	500	17	AI110149	GH09635.3prime GH Dros 1.46e+01
	858	15	8.6	500	19	H72385	ys06b02.r1 Homo sapien 1.46e+01

	859	15	8.6	501	21	AA050501	mj20g01.r1 Soares mous 1.46e+01
	860	15	8.6	501	13	AA813939	oc02h04.s1 NCI_CGAP_GC 1.46e+01
	861	15	8.6	502	29	AQ237533	RPCI11-66P15.TR RPCI11 1.46e+01
c	862	15	8.6	502	6	AA131567	zl33a04.s1 Soares preg 1.46e+01
	863	15	8.6	503	18	AI167297	ox65c06.s1 Soares_NhHM 1.46e+01
	864	15	8.6	503	28	AQ188013	HS_3148_B2_B02_T7 CIT 1.46e+01
c	865	15	8.6	503	28	AQ208517	HS_3233_A2_H12_MR CIT 1.46e+01
c	866	15	8.6	503	28	AQ187393	HS_3128_B1_C04_T7 CIT 1.46e+01
	867	15	8.6	504	7	AA459946	zx66a10.s1 Soares tota 1.46e+01
	868	15	8.6	504	28	AQ205909	HS_3239_A1_H05_T7 CIT 1.46e+01
	869	15	8.6	504	28	AQ194301	RPCI11-62L17.TK RPCI11 1.46e+01
c	870	15	8.6	504	12	AA725029	ai06e03.s1 Soares para 1.46e+01
	871	15	8.6	505	19	N35669	yx62f07.r1 Homo sapien 1.46e+01
c	872	15	8.6	505	29	AQ143897	HS_3074_A2_B02_MF CIT 1.46e+01
c	873	15	8.6	506	11	AA693173	vr62c05.s1 Knowles Sol 1.46e+01
	874	15	8.6	507	12	AA141164	CK01292.3prime CK Dros 1.46e+01
c	875	15	8.6	508	29	AQ147930	HS_3115_B1_G03_T7 CIT 1.46e+01
	876	15	8.6	509	13	AA881219	vx10c04.r1 Soares 2NbM 1.46e+01
	877	15	8.6	509	17	AI111861	UI-R-Y0-mp-e-06-0-UI.s 1.46e+01
c	878	15	8.6	509	10	AA647759	vp04b04.r1 Soares mous 1.46e+01
c	879	15	8.6	509	29	AQ149205	HS_3164_B1_G04_T7 CIT 1.46e+01
	880	15	8.6	512	28	AQ190018	HS_3219_B2_B04_T7 CIT 1.46e+01
	881	15	8.6	512	5	T86840	yd81d03.s1 Homo sapien 1.46e+01
	882	15	8.6	514	5	T86955	yd88f02.s1 Homo sapien 1.46e+01
	883	15	8.6	515	15	AA923168	on18h07.s1 NCI_CGAP_Lu 1.46e+01
	884	15	8.6	517	28	AQ204528	HS_3231_A1_E05_T7 CIT 1.46e+01
c	885	15	8.6	517	28	AQ205340	HS_3235_B1_F04_T7 CIT 1.46e+01
c	886	15	8.6	517	8	AA547462	vk69g11.s1 Knowles Sol 1.46e+01
c	887	15	8.6	518	14	C92177	Dictyostelium discoide 1.46e+01
c	888	15	8.6	518	20	W89012	zh72b09.s1 Soares feta 1.46e+01
c	889	15	8.6	519	28	AQ204357	HS_3218_A2_B10_T7 CIT 1.46e+01
c	890	15	8.6	519	21	AA025819	ze85g10.s1 Soares feta 1.46e+01
c	891	15	8.6	520	6	AA412216	zu10d04.s1 Soares test 1.46e+01
c	892	15	8.6	520	14	C90059	Dictyostelium discoide 1.46e+01
c	893	15	8.6	520	13	AA877451	ob31f09.s1 NCI_CGAP_Ki 1.46e+01
c	894	15	8.6	521	29	AQ235761	HS_2015_B2_H04_T7 CIT 1.46e+01
	895	15	8.6	522	13	AA811831	ob39d03.s1 NCI_CGAP_GC 1.46e+01
	896	15	8.6	523	23	AA142491	mq56e09.r1 Soares 2NbM 1.46e+01
	897	15	8.6	525	13	AA158105	zo57g12.s1 Stratagene 1.46e+01
	898	15	8.6	526	14	AA939098	oq22a09.s1 NCI_CGAP_GC 1.46e+01
	899	15	8.6	526	28	AQ213126	HS_3251_B1_D09_T7 CIT 1.46e+01
	900	15	8.6	527	11	N96115	21428 CD4-15 Arabidops 1.46e+01
	901	15	8.6	529	15	AI009730	EST204181 Normalized r 1.46e+01
	902	15	8.6	529	12	AA740908	nz02g05.s1 NCI_CGAP_GC 1.46e+01
c	903	15	8.6	532	28	AQ172865	HS_3185_B2_B07_T7 CIT 1.46e+01
	904	15	8.6	533	26	AQ010548	F24F17TF8 IGF Arabidops 1.46e+01
	905	15	8.6	533	28	AQ193353	HS_3221_A1_D05_T7 CIT 1.46e+01
	906	15	8.6	534	9	AA533062	nj60d04.s1 NCI_CGAP_Pr 1.46e+01
	907	15	8.6	535	14	C93368	Dictyostelium discoide 1.46e+01
	908	15	8.6	536	17	AI027435	ow49f11.s1 Soares_para 1.46e+01
	909	15	8.6	538	28	AQ176577	HS_3213_B1_F06_T7 CIT 1.46e+01

c	910	15	8.6	542	6	AA292153	zt46b07.r1 Soares ovar 1.46e+01
	911	15	8.6	543	11	AA495700	pat.pk0005.f2 chicken 1.46e+01
	912	15	8.6	545	15	AA834173	of14f02.s1 NCI_CGAP_Ki 1.46e+01
	913	15	8.6	547	27	B83672	RPCI11-17K10.TVB RPCI1 1.46e+01
c	914	15	8.6	548	7	AA476175	yh22g12.r1 Soares mous 1.46e+01
c	915	15	8.6	553	12	AA802495	GM05702.5prime GM Dros 1.46e+01
	916	15	8.6	555	10	AA614322	no77h04.s1 NCI_CGAP_AA 1.46e+01
c	917	15	8.6	555	10	AA422423	vf45h03.r1 Soares mous 1.46e+01
	918	15	8.6	556	5	T54914	yb45c02.s1 Homo sapien 1.46e+01
	919	15	8.6	556	13	C85487	Mus musculus fertilize 1.46e+01
c	920	15	8.6	556	18	AI151637	SWOvAPCAP06E01SK Oncho 1.46e+01
c	921	15	8.6	557	15	AA928777	on98e01.s1 Soares_NFL_ 1.46e+01
	922	15	8.6	558	7	AA529758	vi42h02.r1 Beddington 1.46e+01
c	923	15	8.6	558	26	FR0009567	F.rubripes GSS sequenc 1.46e+01
	924	15	8.6	560	15	AI026209	MEEG0445.MLF Schistoso 1.46e+01
c	925	15	8.6	560	15	C80444	Mus musculus 3.5-dpc b 1.46e+01
	926	15	8.6	562	27	B66402	CIT-HSP-2009M19.TRB CI 1.46e+01
	927	15	8.6	564	8	AA549857	0930m3 gmbPfHB3.1, G. 1.46e+01
c	928	15	8.6	564	14	AA851801	EST194569 Normalized r 1.46e+01
	929	15	8.6	565	14	C91307	Dictyostelium discoide 1.46e+01
	930	15	8.6	566	26	B72708	RPCI11-10B15.TP RPCI11 1.46e+01
	931	15	8.6	568	27	B92947	CIT-HSP-216785.TF CIT- 1.46e+01
	932	15	8.6	569	19	H97627	yv93d02.s1 Homo sapien 1.46e+01
	933	15	8.6	573	22	AA089313	mj51g06.r1 Soares mous 1.46e+01
c	934	15	8.6	573	27	AQ041253	CIT-HSP-2328F13.TR CIT 1.46e+01
	935	15	8.6	576	22	AA001188	ze48g05.r1 Soares reti 1.46e+01
c	936	15	8.6	576	8	AA514405	nf57c12.s1 NCI_CGAP_Co 1.46e+01
c	937	15	8.6	581	7	AA446540	zw67h05.r1 Soares test 1.46e+01
c	938	15	8.6	582	27	B86736	RPCI11-24N19.TV RPCI11 1.46e+01
	939	15	8.6	583	14	C92355	Dictyostelium discoide 1.46e+01
	940	15	8.6	584	29	AQ251218	F22G11-SP6 IGF Arabido 1.46e+01
	941	15	8.6	588	29	AQ239261	RPCI11-70M15.TJ RPCI11 1.46e+01
c	942	15	8.6	588	29	AQ231346	HS_2053_A1_F06_MR CIT 1.46e+01
c	943	15	8.6	595	19	N62894	yy67a11.s1 Homo sapien 1.46e+01
	944	15	8.6	596	28	AQ251193	RPCI11-45G20.TJ RPCI11 1.46e+01
	945	15	8.6	603	11	AA704886	zj52d01.s1 Soares feta 1.46e+01
	946	15	8.6	605	28	AQ110884	CIT-HSP-2376G21.TR CIT 1.46e+01
c	947	15	8.6	608	26	B56982	CIT-HSP-2008P24.TF CIT 1.46e+01
c	948	15	8.6	608	26	FR0008899	F.rubripes GSS sequenc 1.46e+01
	949	15	8.6	610	8	AA543982	vj83f02.r1 Soares mous 1.46e+01
	950	15	8.6	611	26	FR0009903	F.rubripes GSS sequenc 1.46e+01
	951	15	8.6	615	8	AA545633	vj65h02.r1 Knowles Sol 1.46e+01
c	952	15	8.6	618	26	FR0017042	F.rubripes GSS sequenc 1.46e+01
	953	15	8.6	619	27	FR0032642	Fugu rubripes GSS sequ 1.46e+01
c	954	15	8.6	622	26	B76099	RPCI11-13J17.TV RPCI11 1.46e+01
	955	15	8.6	625	27	B82579	RPCI11-16P12.TP RPCI11 1.46e+01
	956	15	8.6	626	15	AI010275	EST204726 Normalized r 1.46e+01
	957	15	8.6	628	14	C91010	Dictyostelium discoide 1.46e+01
c	958	15	8.6	629	14	C92854	Dictyostelium discoide 1.46e+01
	959	15	8.6	636	18	AI082640	oy56g07.s1 NCI_CGAP_Br 1.46e+01
c	960	15	8.6	639	26	AQ018913	CIT-HSP-2307J22.TF CIT 1.46e+01

	961	15	8.6	641	7	D86654	Rat ovary mRNA, mRNA s	1.46e+01
c	962	15	8.6	641	7	AA519035	vi26d03.r1 Barstead mo	1.46e+01
c	963	15	8.6	643	19	N66622	yy70d04.s1 Homo sapien	1.46e+01
	964	15	8.6	647	18	AI185631	qe65a03.x1 Soares_feta	1.46e+01
	965	15	8.6	648	26	CI22203	Ciona intestinalis gen	1.46e+01
c	966	15	8.6	648	8	AA088430	zl89e04.s1 Stratagene	1.46e+01
c	967	15	8.6	651	27	AQ048775	cLM-9h6-u cLM Giardia	1.46e+01
	968	15	8.6	654	18	AI103913	EST213202 Normalized r	1.46e+01
	969	15	8.6	654	28	AQ194054	RPCI11-43B20.TJ RPCI11	1.46e+01
c	970	15	8.6	654	22	AA155972	zo48a02.s1 Stratagene	1.46e+01
	971	15	8.6	655	29	AQ239573	CIT-HSP-2387I3.TF.1 CI	1.46e+01
	972	15	8.6	668	14	C91527	Dictyostelium discoide	1.46e+01
c	973	15	8.6	668	26	B68086	CIT978SK-A-476C8.TP CI	1.46e+01
c	974	15	8.6	678	14	C91194	Dictyostelium discoide	1.46e+01
c	975	15	8.6	682	28	AQ195953	RPCI11-58E13.TJ RPCI11	1.46e+01
	976	15	8.6	685	14	C92444	Dictyostelium discoide	1.46e+01
c	977	15	8.6	686	27	AQ079078	CIT-HSP-2356E10.TF CIT	1.46e+01
c	978	15	8.6	691	27	AQ049991	nbxb0003aF06r CUGI Ric	1.46e+01
	979	15	8.6	698	12	AA141060	CK01129.5prime.contig	1.46e+01
	980	15	8.6	711	23	AA105949	ml84a09.r1 Stratagene	1.46e+01
	981	15	8.6	716	8	C22997	Jpanese flounder liver	1.46e+01
	982	15	8.6	725	17	AI130084	SWOvL3CAN53G04SK Oncho	1.46e+01
c	983	15	8.6	727	26	B67673	T23N16TR TAMU Arabidop	1.46e+01
	984	15	8.6	730	18	AI170341	EST216267 Normalized r	1.46e+01
c	985	15	8.6	746	26	B76166	RPCI11-13N15.TV RPCI11	1.46e+01
c	986	15	8.6	750	17	AI075195	oy96c07.x1 Soares_feta	1.46e+01
c	987	15	8.6	756	18	AI142100	ow61f03.x1 Soares_NSF_	1.46e+01
c	988	15	8.6	760	18	AI088191	oz97g11.x1 Soares_para	1.46e+01
	989	15	8.6	766	17	AI129633	qc49a12.x1 Soares_preg	1.46e+01
c	990	15	8.6	774	29	AQ254560	CpG0697B CpIOWAgDNA1 C	1.46e+01
c	991	15	8.6	786	15	C93845	Dictyostelium discoide	1.46e+01
	992	15	8.6	805	26	AB005953	Mouse genomic DNA, chr	1.46e+01
	993	15	8.6	816	12	AA817246	LD22445.5prime LD Dros	1.46e+01
	994	15	8.6	867	26	B08768	T7L4-T7 TAMU Arabidops	1.46e+01
c	995	15	8.6	894	27	AQ026858	CIT-HSP-2322N15.TF CIT	1.46e+01
c	996	15	8.6	899	27	AQ027025	CIT-HSP-2323O10.TF CIT	1.46e+01
c	997	15	8.6	916	11	C84167	Dictyostelium discoide	1.46e+01
	998	15	8.6	1078	24	C22923	Dictyostelium discoide	1.46e+01
c	999	15	8.6	1196	26	B09810	F26P15-Sp6 IGF Arabido	1.46e+01
c	1000	15	8.6	1994	10	AA415086	Mg0020 RCW Lambda Zap	1.46e+01

ALIGNMENTS

```
RESULT    1
LOCUS       AA846103       458 bp    mRNA            EST       04-MAR-1998
DEFINITION  ak83f07.s1 Barstead spleen HPLRB2 Homo sapiens cDNA clone
            IMAGE:1414501 3', mRNA sequence.
ACCESSION   AA846103
NID         g2932243
KEYWORDS    EST.
```



```
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 458)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
            Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
            Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40m13 fwd. ET from Amersham.
FEATURES              Location/Qualifiers
     source           1. .458
                      /organism="Homo sapiens"
                      /note="Organ: spleen; Vector: pT7T3D-Pac (Pharmacia) with
                      a modified polylinker; Site_1: EcoRI; Site_2: NotI; 1st
                      strand cDNA was primed with a Not I - oligo(dT) primer [5'
                      TGTTACGAATCTGAAGTGGGAGCGGCCGCCCCTTTTTTTTTTTTTTTTTTTTTTTT
                      3']; double-stranded cDNA was ligated to Eco RI adaptors
                      [AATTCGGATCCTTG], digested with Not I and cloned into the
                      Not I and Eco RI sites of the modified pT7T3 vector.
                      Library constructed by Bob Barstead."
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1414501"
                      /clone_lib="Barstead spleen HPLRB2"
                      /sex="male"
                      /dev_stage="adult, 17 years"
                      /lab_host="DH10B"
BASE COUNT      112 a    107 c     56 g    183 t
ORIGIN

  Query Match          100.0%;  Score 174;  DB 13;  Length 458;
  Best Local Similarity 100.0%;  Pred. No. 0.00e+00;
  Matches  174;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      2 TTTCACTTGCAAATTCACTTGTTTTATTTTTATCTGGCATATAAATTGCATATTTTTATG 61
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp    174 TTTCACTTGCAAATTCACTTGTTTTATTTTTATCTGGCATATAAATTGCATATTTTTATG 115

Db     62 TGAAATGTGGTTGTATACATTAGAAATAAGATTTACACATTTCAAAGCACACTACTGCAA 121
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```



```
Cp    114 TGAAATGTGGTTGTATACATTAGAAATAAGATTTACACATTTCAAAGCACACTACTGCAA 55

Db    122 AAATAAATTATTTTTAACCTCCAAACTCTCTTTTGAGCTTTGCCTGCTCAGATC 175
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     54 AAATAAATTATTTTTAACCTCCAAACTCTCTTTTGAGCTTTGCCTGCTCAGATC 1


RESULT    2
LOCUS       AA894763       440 bp    mRNA            EST       09-JUN-1998
DEFINITION  oj55a12.s1 NCI_CGAP_Kid3 Homo sapiens cDNA clone IMAGE:1502206 3',
            mRNA sequence.
ACCESSION   AA894763
NID         g3031164
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 440)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 915   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 429.
FEATURES              Location/Qualifiers
     source           1. .440
                      /organism="Homo sapiens"
                      /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                      a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                      strand cDNA was primed with a Not I - oligo(dT) primer,
                      double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Not I and cloned into the Not
                      I and Eco RI sites of the modified pT7T3 vector. mRNA
                      source: 2 pooled kidneys. Library went through one round
                      of normalization. Library constructed by Bento Soares and
                      M. Fatima Bonaldo."
```



```
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1502206"
                      /clone_lib="NCI_CGAP_Kid3"
                      /lab_host="DH10B"
BASE COUNT      109 a    105 c     53 g    173 t
ORIGIN

  Query Match          98.3%;  Score 171;  DB 15;  Length 440;
  Best Local Similarity 100.0%;  Pred. No. 0.00e+00;
  Matches  171;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 CACTTGCAAATTCACTTGTTTTATTTTTATCTGGCATATAAATTGCATATTTTTATGTGA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp    171 CACTTGCAAATTCACTTGTTTTATTTTTATCTGGCATATAAATTGCATATTTTTATGTGA 112

Db     61 AATGTGGTTGTATACATTAGAAATAAGATTTACACATTTCAAAGCACACTACTGCAAAAA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp    111 AATGTGGTTGTATACATTAGAAATAAGATTTACACATTTCAAAGCACACTACTGCAAAAA 52

Db    121 TAAATTATTTTTAACCTCCAAACTCTCTTTTGAGCTTTGCCTGCTCAGATC 171
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     51 TAAATTATTTTTAACCTCCAAACTCTCTTTTGAGCTTTGCCTGCTCAGATC 1


RESULT    3
LOCUS       HUMGS01365    174 bp    mRNA            EST       17-JUN-1996
DEFINITION  Human HL60 3'directed MboI cDNA, HUMGS01365, clone pm1864, mRNA
            sequence.
ACCESSION   D20391
NID         g501487
KEYWORDS    EST; EST(expressed sequence tag); HL60; gene signature(GS);
            granulocyte; macrophage; monocyte; promyelocyte.
SOURCE      Homo sapiens adult female promyelocyte cell_line:HL60 cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (sites)
  AUTHORS   Murakawa,K., Matsubara,K., Fukushima,A., Yoshii,J. and Okubo,K.
  TITLE     Chromosomal assignments of 3'-directed partial cDNA sequences
            representing novel genes expressed in granulocytoid cells
  JOURNAL   Genomics 23 (2), 379-389 (1994)
  MEDLINE   95137584
REFERENCE   2  (sites)
  AUTHORS   Okubo,K., Itoh,K., Fukushima,A., Yoshii,J. and Matsubara,K.
  TITLE     Monitoring cell physiology by expression profiles and discovering
            cell type-specific genes by compiled expression profiles
  JOURNAL   Genomics 30 (2), 178-186 (1995)
  MEDLINE   96163870
REFERENCE   3  (bases 1 to 174)
  AUTHORS   Okubo,K., Fukushima,A., Yoshii,J., Niiyama,T., Kojima,Y.,
```

```
            Yoshinari,H., Arimoto,J. and Matsubara,K.
  TITLE     Gene expression of human promyelocytic cell line HL60before and
            after induction of differentiation. A new application of
            3'directedcDNA sequencing
  JOURNAL   Unpublished (1993)
REFERENCE   4  (bases 1 to 174)
  AUTHORS   Okubo,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-SEP-1993) to the DDBJ/EMBL/GenBank databases. Kousaku
            Okubo, Osaka University, Institute for Molecular and Cellular Bio;
            1-3,Yamada-oka, Suita, Osaka Pref. 565, Japan
            (E-mail:kousaku@imcb.osaka-u.ac.jp, Tel:06-877-5111(ex.3315),
            Fax:06-877-1922)
COMMENT     Submitted (10-Sep-1993) to DDBJ by:
            Kousaku Okubo
            Institute for Molecular and Cellular Biology
            Osaka University
            3-1, Yamadaoka
            Suita, Osaka, 565
            Japan
            Phone: 06-877-5111
            Fax  : 06-875-1922.
FEATURES              Location/Qualifiers
     source           1. .174
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /cell_line="HL60"
                      /cell_type="promyelocyte"
                      /dev_stage="adult"
                      /sex="female"
BASE COUNT       68 a     18 c     30 g     51 t      7 others
ORIGIN

  Query Match          96.0%;  Score 167;  DB 7;  Length 174;
  Best Local Similarity  96.0%;  Pred. No. 2.78e-301;
  Matches  167;  Conservative    0;  Mismatches   7;  Indels   0;  Gaps   0;

Db      1 GATCTGAGCAGGCAAAGCTCAAAAGAGAGTTTGGAGGNTAAAAATAATTTATTTTTGCAG 60
          ||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Qy      1 GATCTGAGCAGGCAAAGCTCAAAAGAGAGTTTGGAGGTTAAAAATAATTTATTTTTGCAG 60

Db     61 TAGTGTGCTTTGAAATGTGTAAATNTTATTNNTAATGTATACAACCACATTTCACATAAA 120
          |||||||||||||||||||||||| |||||  ||||||||||||||||||||||||||||
Qy     61 TAGTGTGCTTTGAAATGTGTAAATCTTATTTCTAATGTATACAACCACATTTCACATAAA 120

Db    121 ANTATGCAATTTATATGCCAGATAAAANTAAAACAAGTGAATTTGCAAGNGAAA 174
          | ||||||||||||||||||||||||| ||||||||||||||||||||| ||||
Qy    121 AATATGCAATTTATATGCCAGATAAAAATAAAACAAGTGAATTTGCAAGTGAAA 174


RESULT    4
```



```
LOCUS       AA372536      353 bp    mRNA            EST       21-APR-1997
DEFINITION  EST84447 Colon adenocarcinoma IV Homo sapiens cDNA 5' end, mRNA
            sequence.
ACCESSION   AA372536
NID         g2024854
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 353)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
            Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
            White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
            Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
            Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
            Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
            Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
            He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
            Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H.,
            Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
            Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
            Fraser,C.M. and Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   96026280
COMMENT
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES              Location/Qualifiers
     source           1. .353
                      /organism="Homo sapiens"
                      /note="Organ: colon; Vector: pBluescript SK-; Site_1:
                      EcoRI; Site_2: XhoI"
                      /db_xref="ATCC (inhost):176893"
```



```
                      /db_xref="taxon:9606"
                      /clone_lib="Colon adenocarcinoma IV"
                      /dev_stage="adult"
     mRNA             <1. .>353
BASE COUNT       99 a     82 c     47 g    123 t      2 others
ORIGIN

  Query Match          11.5%;  Score 20;  DB 25;  Length 353;
  Best Local Similarity 100.0%;  Pred. No. 7.70e-06;
  Matches   20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    190 TTTTTAACCTCCAAACTCTC 209
          ||||||||||||||||||||
Cp     44 TTTTTAACCTCCAAACTCTC 25


RESULT    5
LOCUS       C89667        402 bp    mRNA            EST       15-APR-1998
DEFINITION  Dictyostelium discoideum slug cDNA, clone SSA560, mRNA sequence.
ACCESSION   C89667
NID         g3059287
KEYWORDS    EST; EST(expressed sequence tag).
SOURCE      Dictyostelium discoideum (strain:AX4) slug cDNA to mRNA,
            clone_lib:SS.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 402)
  AUTHORS   Urushihara,H.
  TITLE     Developmental cDNA in Dictyostelium discoideum
  JOURNAL   Published Only in DataBase (1998) In press
REFERENCE   2  (bases 1 to 402)
  AUTHORS   Urushihara,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAR-1998) to the DDBJ/EMBL/GenBank databases. Hideko
            Urushihara, University of Tsukuba, Institute of Biological
            Sciences; 1-1-1 Ten-noudai, Tsukuba-shi 305-8572, Japan
            (E-mail:qxurushi@bank.dna.affrc.go.jp, Tel:+81-0298-53-4664,
            Fax:+81-0298-53-6614)
COMMENT     PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES              Location/Qualifiers
     source           1. .402
                      /organism="Dictyostelium discoideum"
                      /strain="AX4"
                      /db_xref="taxon:44689"
                      /clone_lib="SS"
                      /dev_stage="slug"
BASE COUNT      137 a     36 c     81 g    138 t     10 others
ORIGIN

  Query Match          10.9%;  Score 19;  DB 14;  Length 402;
  Best Local Similarity 100.0%;  Pred. No. 1.77e-04;
```



```
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    321 TTAAAAATAATTTATTTTT 339
          ||||||||||||||||||||
Qy     38 TTAAAAATAATTTATTTTT 56


RESULT    6
LOCUS       C24637        442 bp    mRNA            EST       24-JUL-1997
DEFINITION  Dictyostelium discoideum slug cDNA, clone SL-X036, mRNA sequence.
ACCESSION   C24637
NID         g2243058
KEYWORDS    EST; EST(expressed sequence tag).
SOURCE      Dictyostelium discoideum (strain:AX4) slug cDNA to mRNA,
            clone_lib:SS.
  ORGANISM  Dictyostelium discoideum
            Eukaryotae; mitochondrial eukaryotes; Dictyosteliida;
            Dictyostelium.
REFERENCE   1  (bases 1 to 442)
  AUTHORS   Urushihara,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUN-1997) to the DDBJ/EMBL/GenBank databases. Hideko
            Urushihara, University of Tsukuba, Institute of Biological
            Sciences; Ten-noudai 1-1-1, Tsukuba-shi, Ibaraki-ken 305, Japan
            (E-mail:i4402hu@sakura.cc.tsukuba.ac.jp, Tel:0298-53-4664,
            Fax:0298-53-6614)
REFERENCE   2  (sites)
  AUTHORS   Morio,T. and Tanaka,Y.
  TITLE     Developmental cDNA in Dictyostelium discoideum(970624)
  JOURNAL   Unpublished (1997)
COMMENT     PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES              Location/Qualifiers
     source           1. .442
                      /organism="Dictyostelium discoideum"
                      /strain="AX4"
                      /db_xref="taxon:44689"
                      /clone_lib="SS"
                      /dev_stage="slug"
BASE COUNT      155 a     51 c     83 g    153 t
ORIGIN

  Query Match          10.9%;  Score 19;  DB 7;  Length 442;
  Best Local Similarity 100.0%;  Pred. No. 1.77e-04;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    422 TTAAAAATAATTTATTTTT 440
          ||||||||||||||||||||
Qy     38 TTAAAAATAATTTATTTTT 56


RESULT    7
```

```
LOCUS       AA323500      453 bp    mRNA            EST       20-APR-1997
DEFINITION  EST26285 Cerebellum II Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION   AA323500
NID         g1975825
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 453)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
            Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
            White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
            Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
            Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
            Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
            Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
            He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
            Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H.,
            Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
            Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
            Fraser,C.M. and Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   96026280
COMMENT     Other_ESTs: THC175584
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES              Location/Qualifiers
     source           1. .453
                      /organism="Homo sapiens"
                      /note="Organ: brain; Vector: pBluescript SK-; Site_1:
                      EcoRI; Site_2: XhoI"
                      /db_xref="ATCC (inhost):124038"
                      /db_xref="taxon:9606"
```



```
                      /clone_lib="Cerebellum II"
                      /tissue_type="cerebellum"
                      /dev_stage="adult"
     mRNA             <1. .>453
BASE COUNT      136 a      76 c      95 g     143 t       3 others
ORIGIN

  Query Match           10.9%;  Score 19;  DB 24;  Length 453;
  Best Local Similarity 100.0%;  Pred. No. 1.77e-04;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    129 TGCATATTTTTATGTGAAA 147
          ||||||||||||||||||
Cp    128 TGCATATTTTTATGTGAAA 110


RESULT    8
LOCUS       C90335        473 bp    mRNA            EST       15-APR-1998
DEFINITION  Dictyostelium discoideum slug cDNA, clone SSI219, mRNA sequence.
ACCESSION   C90335
NID         g3059955
KEYWORDS    EST; EST(expressed sequence tag).
SOURCE      Dictyostelium discoideum (strain:AX4) slug cDNA to mRNA,
            clone_lib:SS.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Dictyostelida; Dictyostelium.
REFERENCE   1  (bases 1 to 473)
  AUTHORS   Urushihara,H.
  TITLE     Developmental cDNA in Dictyostelium discoideum
  JOURNAL   Published Only in DataBase (1998) In press
REFERENCE   2  (bases 1 to 473)
  AUTHORS   Urushihara,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAR-1998) to the DDBJ/EMBL/GenBank databases. Hideko
            Urushihara, University of Tsukuba, Institute of Biological
            Sciences; 1-1-1 Ten-noudai, Tsukuba-shi 305-8572, Japan
            (E-mail:qxurushi@bank.dna.affrc.go.jp, Tel:+81-0298-53-4664,
            Fax:+81-0298-53-6614)
COMMENT     PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES              Location/Qualifiers
     source           1. .473
                      /organism="Dictyostelium discoideum"
                      /strain="AX4"
                      /db_xref="taxon:44689"
                      /clone_lib="SS"
                      /dev_stage="slug"
BASE COUNT      165 a      54 c      87 g     167 t
ORIGIN

  Query Match           10.9%;  Score 19;  DB 14;  Length 473;
  Best Local Similarity 100.0%;  Pred. No. 1.77e-04;
```



```
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    412 TTAAAAATAATTTATTTTT 430
          ||||||||||||||||||||
Qy     38 TTAAAAATAATTTATTTTT 56


RESULT    9
LOCUS       C23819        157 bp    mRNA            EST       22-APR-1997
DEFINITION  Dictyostelium discoideum gamete cDNA, clone FCL-AB12, mRNA
            sequence.
ACCESSION   C23819
NID         g2073331
KEYWORDS    EST; EST(expressed sequence tag).
SOURCE      Dictyostelium discoideum (strain:KAX3) Sexually competent gamete
            cDNA to mRNA, clone_lib:FCL.
  ORGANISM  Dictyostelium discoideum
            Eukaryotae; mitochondrial eukaryotes; Dictyosteliida;
            Dictyostelium.
REFERENCE   1  (bases 1 to 157)
  AUTHORS   Urushihara,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-APR-1997) to the DDBJ/EMBL/GenBank databases. Hideko
            Urushihara, University of Tsukuba, Institute of Biological
            Sciences; 1-1-1 Tennohdai, Tsukuba, Ibaraki 305, Japan
            (E-mail:d402hu@sakura.cc.tsukuba.ac.jp, Tel:0298-53-4664,
            Fax:0298-53-6614)
REFERENCE   2  (sites)
  AUTHORS   Kobayashi,A., Shimizu,H. and Urushihara,H.
  TITLE     Sexual cDNA in CSM
  JOURNAL   Unpublished (1997)
COMMENT     PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES              Location/Qualifiers
     source           1. .157
                      /organism="Dictyostelium discoideum"
                      /strain="KAX3"
                      /db_xref="taxon:44689"
                      /cell_type="gamete"
                      /clone_lib="FCL"
                      /dev_stage="Sexually competent"
BASE COUNT       97 a       5 c       2 g      53 t
ORIGIN

  Query Match           10.3%;  Score 18;  DB 6;  Length 157;
  Best Local Similarity 100.0%;  Pred. No. 3.66e-03;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    127 TAAAAATAATTTATTTTT 144
          ||||||||||||||||||
Qy     39 TAAAAATAATTTATTTTT 56
```



```
RESULT   10
LOCUS       N20126        213 bp    mRNA            EST       15-DEC-1995
DEFINITION  yx29e02.s1 Homo sapiens cDNA clone 263162 3'.
ACCESSION   N20126
NID         g1124793
KEYWORDS    EST.
SOURCE      human clone=263162 primer=m13 -40 forward library=Soares melanocyte
            2NbHM vector=pT7T3D (Pharmacia) with a modified polylinker
            host=DH10B (ampicillin resistant) Rsite1=Not I Rsite2=Eco RI Male.
            1st strand cDNA was primed with a Not I - oligo(dT) primer
            [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCAGTTTTTTTTTTTTTTTTTT-3'],
            double-stranded cDNA was size selected, ligated to Eco RI adapters
            (Pharmacia), digested with Not I and cloned into the Not I and Eco
            RI sites of a modified pT7T3 vector (Pharmacia). Library
            constructed by Bento Soares and M.Fatima Bonaldo. RNA from normal
            foreskin melanocytes (FS374) was kindly provided by Dr. Anthony P.
            Albino.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 213)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 105
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES              Location/Qualifiers
     source           1. .213
                      /organism="Homo sapiens"
                      /clone="263162"
     mRNA             <1. .>213
BASE COUNT       53 a       42 c       35 g      76 t       7 others
ORIGIN
```

```
  Query Match          10.3%;  Score 18;  DB 19;  Length 213;
  Best Local Similarity 100.0%;  Pred. No. 3.66e-03;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      51 ATTTCACATAAAAATATG 68
           ||||||||||||||||||
Qy     109 ATTTCACATAAAAATATG 126


RESULT   11
LOCUS       AA957734       213 bp    mRNA            EST       07-MAY-1998
DEFINITION  UI-R-E1-gc-b-12-0-UI.s1 UI-R-E1 Rattus norvegicus cDNA clone
            UI-R-E1-gc-b-12-0-UI 3', mRNA sequence.
ACCESSION   AA957734
NID         g31211429
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 213)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT
            Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence tag present in the cDNA between the NotI site and the
            oligo-dT track (not shown) served to identify it as a clone from
            the normalized adult 12-Day-Embryo library. cDNA Library
            Preparation: M. Fatima Bonaldo, Ph.D. Clone distribution: clones
            will be available through Research Genetics
            Seq primer: M13 Forward.
FEATURES              Location/Qualifiers
     source           1. .213
                      /organism="Rattus norvegicus"
                      /strain="Sprague-Dawley"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-E1
                      library is a subtracted library derived from the UI-R-E0
                      library.  The UI-R-E0 library  consisted of a mixture of
                      individually  tagged normalized libraries constructed from
                      8, 12 and 18-day embryo.  The tag is a string of 3-5
                      nucleotides present between the Not I site  and the
```



```
                      oligo-dT track which allows identification of the library
                      of  origin of a clone within the mixture.  The subtracted
                      library (UI-R-E1) was constructed as follows: PCR
                      amplified cDNA inserts from a pool of  UI-R-E0 clones from
                      which 3' ESTs had been derived was used as a driver in a
                      hybridization with the UI-R-E0 library in the form of
                      single-stranded circles. The remaining single-stranded
                      circles (subtracted library) was purified by
                      hydroxyapatite column chromatography, converted to
                      double-stranded circles and electroporated into DH10B
                      bacteria (Life Technologies) to generate the UI-R-E1
                      library. This procedure has been previously described
                      (Bonaldo, Lennon and Soares, Genome Research 6: 791-806,
                      1996)"
                      /db_xref="taxon:10116"
                      /clone="UI-R-E1-gc-b-12-0-UI"
                      /clone_lib="UI-R-E1"
                      /dev_stage="adult"
                      /lab_host="DH10B (Life Technologies)"
BASE COUNT       63 a      41 c      55 g      54 t
ORIGIN

  Query Match          10.3%;  Score 18;  DB 14;  Length 213;
  Best Local Similarity 100.0%;  Pred. No. 3.66e-03;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db       9 GTTAAAAATAATTTATTT 26
           ||||||||||||||||||
Qy      37 GTTAAAAATAATTTATTT 54


RESULT   12
LOCUS       T18134         271 bp    mRNA            EST       30-AUG-1994
DEFINITION  0695c3 Plasmodium falciparum cDNA clone 0695c 5'.
ACCESSION   T18134
NID         g462920
KEYWORDS    EST.
SOURCE      malaria parasite clone=0695c library=czapPFDd2.1, Debopam
            Chakrabarti strain=Dd2 vector=Lambda ZAP II host=E. coli XL-1 blue
            primer=T3 Rsite1=EcoR I Rsite2=Xho I PolyA+ RNA, from asynchronous
            blood stage parasites of the Dd2 isolate cultured in vitro, was
            reverse transcribed using an oligo dT-Xho I primer.  Second strand
            was prepared using RNase H and DNA polymerase I. EcoR I adapters
            were ligated to the cDNA, and it was digested with Xho I. Prepared
            fragments were ligated into EcoR I + Xho I digested lambda ZAP II
            vector.
  ORGANISM  Plasmodium falciparum
            Eucaryotae; Protozoa; Apicomplexa; Sporozoa; Coccidia;
            Eucoccidiida; Haemosporina; Plasmodium.
REFERENCE   1  (bases 1 to 271)
  AUTHORS   Chakrabarti,D., Reddy,G.R., Dame,J.B., Almira,E.C., Laipis,P.J.,
```



```
            Ferl,R.J., Yang,T.P., Rowe,T.C. and Schuster,S.M.
  TITLE     Analysis of Expressed Sequence Tags from Plasmodium Falciparum
  JOURNAL   Mol. Biochem. Parasitol. 66, 97-104 (1994)
  MEDLINE   95075403
COMMENT
            Contact: Debopam Chakrabarti
            Biotechnology & Infectious Diseases
            University of Florida
            Box 110880, Gainesville, Fl 32611
            Tel: 9043924700 ext. 5817
            Fax: 9043929704
            Email: dchak@icbr.ifas.ufl.edu.
FEATURES              Location/Qualifiers
     source           1. .271
                      /organism="Plasmodium falciparum"
                      /clone="0695c"
                      /strain="Dd2"
BASE COUNT      116 a      19 c      15 g     115 t      6 others
ORIGIN

  Query Match          10.3%;  Score 18;  DB 5;  Length 271;
  Best Local Similarity 100.0%;  Pred. No. 3.66e-03;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     157 ATAAAAATAAAACAAGTG 174
           ||||||||||||||||||
Qy     142 ATAAAAATAAAACAAGTG 159


RESULT   13
LOCUS       AA998261       286 bp    mRNA            EST       05-JUN-1998
DEFINITION  UI-R-C0-ie-g-05-0-UI.s1 UI-R-C0 Rattus norvegicus cDNA clone
            UI-R-C0-ie-g-05-0-UI 3', mRNA sequence.
ACCESSION   AA998261
NID         g3188912
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 286)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT
            Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
```



```
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence tag present in the cDNA between the NotI site and the
            oligo-dT track (not shown) served to identify it as a clone from
            the normalized adult 12-Day-Embryo library. cDNA Library
            Preparation: M. Fatima Bonaldo, Ph.D. Clone distribution: clones
            will be available through Research Genetics
            Seq primer: M13 Forward.
FEATURES              Location/Qualifiers
     source           1. .286
                      /organism="Rattus norvegicus"
                      /strain="Sprague-Dawley"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-C0
                      library is a subtracted library derived from the UI-R-A1
                      and UI-R-E1 libraries. The UI-R-A1 library consisted of a
                      mixture of individually tagged normalized libraries
                      constructed from rat  placenta, adult lung, brain, liver,
                      kidney, heart, spleen, ovary,  and muscle. The UI-R-E1
                      library  consisted of a mixture of individually tagged
                      normalized libraries constructed from 8, 12 and 18-day
                      embryo. The tag is a string of 3-5 nucleotides present
                      between the Not I  site and the oligo-dT track which
                      allows identification of the library of origin of a clone
                      within the mixture. The subtracted library (UI-R-C0) was
                      constructed as follows: PCR amplified cDNA inserts from a
                      pool of UI-R-A1 and UI-R-E1 clones from which 3' ESTs had
                      been derived was used as a driver in a hybridization with
                      the pooled UI-R-A1 and UI-R-E1 library in the form of
                      single-stranded circles.  The remaining single-stranded
                      circles (subtracted library) was purified by
                      hydroxyapatite column chromatography, converted to
                      double-stranded circles and electroporated into DH10B
                      bacteria (Life Technologies) to generate the UI-R-C0
                      library. This procedure has been previously described
                      (Bonaldo, Lennon and Soares, Genome Research 6: 791-806,
                      1996)"
                      /db_xref="taxon:10116"
                      /clone="UI-R-C0-ie-g-05-0-UI"
                      /clone_lib="UI-R-C0"
                      /dev_stage="adult"
                      /lab_host="DH10B (Life Technologies)"
BASE COUNT       86 a      55 c      87 g      58 t
ORIGIN

  Query Match          10.3%;  Score 18;  DB 15;  Length 286;
  Best Local Similarity 100.0%;  Pred. No. 3.66e-03;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db       1 GTTAAAAATAATTTATTT 18
```

```
                    ||||||||||||||||||
Qy      37 GTTAAAAATAATTTATTT 54


RESULT   14
LOCUS       AI144676      308 bp    mRNA            EST       28-SEP-1998
DEFINITION  UI-R-BT0-pl-h-06-0-UI.s1 UI-R-BT0 Rattus norvegicus cDNA clone
            UI-R-BT0-pl-h-06-0-UI 3', mRNA sequence.
ACCESSION   AI144676
NID         g3666475
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 308)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT
            Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence tag present in the cDNA between the NotI site and the
            oligo-dT track served to identify it as a clone from the normalized
            adult NRBcorpus-striatum library. cDNA Library Preparation: M.
            Fatima Bonaldo, Ph.D. Clone distribution: clones will be available
            through Research Genetics
            Seq primer: M13 Forward.
FEATURES             Location/Qualifiers
     source          1. .308
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; This library
                     (UI-R-BT0) consists of a mixture of individually tagged
                     normalized libraries constructed from rat hippocampus,
                     thalamus, mid-brain, medulla, corpus striatum, cerebral
                     cortex and testis. The tag used to  identify the source
                     tissue is a string of 3-6 nucleotides present  between the
                     Not I site and the oligo-dT track which allows
                     identification  of the library of origin of a clone within
                     the mixture. This library was then subtracted using a
                     driver consisting of a mixture of all clones from UI-R-A0,
                     UI-R-A1, UI-R-E0, UI-R-E1, UI-R-C0, UI-R-C1, UI-R-C2 and
```



```
                     UI-R-C2p."
                     /db_xref="taxon:10116"
                     /clone="UI-R-BT0-pl-h-06-0-UI"
                     /clone_lib="UI-R-BT0"
                     /dev_stage="adult"
                     /lab_host="DH10B (Life Technologies)"
BASE COUNT       84 a      57 c      89 g      78 t
ORIGIN

  Query Match           10.3%;  Score 18;  DB 18;  Length 308;
  Best Local Similarity 100.0%;  Pred. No. 3.66e-03;
  Matches    18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      26 GTTAAAAATAATTTATTT 43
           ||||||||||||||||||
Qy      37 GTTAAAAATAATTTATTT 54


RESULT   15
LOCUS       AQ194533      374 bp    DNA             GSS       15-SEP-1998
DEFINITION  RPCI11-60K17.TK RPCI11 Homo sapiens genomic clone R-60K17, genomic
            survey sequence.
ACCESSION   AQ194533
NID         g3606145
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 374)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
            Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
  TITLE     Use of human BAC End Sequences for Sequence-Ready Map Building
  JOURNAL   Unpublished (1998)
COMMENT     Other_GSSs: RPCI11-60K17.TJ
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            For clone availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .374
                     /organism="Homo sapiens"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
```



```
                     RPCI11 Human Male BAC Library"
                     /db_xref="taxon:9606"
                     /clone="R-60K17"
                     /clone_lib="RPCI11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
BASE COUNT      117 a      53 c      53 g     151 t
ORIGIN

  Query Match           10.3%;  Score 18;  DB 28;  Length 374;
  Best Local Similarity 100.0%;  Pred. No. 3.66e-03;
  Matches    18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     326 AAAAATAAATTATTTTTA 343
           ||||||||||||||||||
Cp      56 AAAAATAAATTATTTTTA 39


RESULT   16
LOCUS       AQ002609      386 bp    DNA             GSS       17-JUN-1998
DEFINITION  CIT978SK-176G21-1.TV CIT978SK Homo sapiens genomic clone 176G21,
            genomic survey sequence.
ACCESSION   AQ002609
NID         g3029813
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 386)
  AUTHORS   Kim,U.-J., Adams,M.D. and Simon,M.I.
  TITLE     Determination of clone end sequences of human Bacterial Artificial
            Chromosomes
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Ung-Jin Kim
            CalTech Genome Research Lab
            California Institute of Technology
            Division of Biology, MS 147-75, Pasadena, CA 91125, USA
            Tel: 626 796 7066
            Fax: 626 395 4901
            Email: ung@ash.tree.caltech.edu
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .386
                     /organism="Homo sapiens"
                     /note="Vector: pBAC108L; Site_1: HindIII; Site_2: HindIII;
```



```
                     CalTech Human BAC Library A"
                     /db_xref="taxon:9606"
                     /clone="176G21"
                     /clone_lib="CIT978SK"
                     /sex="Female"
                     /cell_type="Fibroblast"
BASE COUNT      123 a      78 c      78 g     107 t
ORIGIN

  Query Match           10.3%;  Score 18;  DB 26;  Length 386;
  Best Local Similarity 100.0%;  Pred. No. 3.66e-03;
  Matches    18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     194 CAAAAATAAATTATTTTT 211
           ||||||||||||||||||
Cp      57 CAAAAATAAATTATTTTT 40


RESULT   17
LOCUS       AQ124657      404 bp    DNA             GSS       21-SEP-1998
DEFINITION  HS_2267_B1_H04_MF CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=2267 Col=7 Row=P, genomic survey
            sequence.
ACCESSION   AQ124657
NID         g3501823
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 404)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Furlong,J.,
            Shaker,R., Schmidt,S., Traicoff,R. and Hood,L.E.
  TITLE     Construction of a Characterized Clone Resource for Genomic
            Sequencing
  JOURNAL   Unpublished (1998)
COMMENT
            Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 2267  row: P  column: 7
            Class: BAC ends
            High quality sequence stop: 404.
FEATURES             Location/Qualifiers
     source          1. .404
                     /organism="Homo sapiens"
```

```
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
                     /db_xref="taxon:9606"
                     /clone="Plate=2267 Col=7 Row=P"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
BASE COUNT      116 a     73 c     88 g    124 t      3 others
ORIGIN

  Query Match           10.3%;  Score 18;  DB 28;  Length 404;
  Best Local Similarity  94.7%;  Pred. No. 3.66e-03;
  Matches    18;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db    286 AAAAATATGCAATTTANAT 304
          |||||||||||||||| ||
Qy    118 AAAAATATGCAATTTATAT 136


RESULT   18
LOCUS       AQ081794      446 bp    DNA             GSS       19-AUG-1998
DEFINITION  RPCI11-55K19.TK RPCI11 Homo sapiens genomic clone R-55K19, genomic
            survey sequence.
ACCESSION   AQ081794
NID         g3442978
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 446)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
            Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
  TITLE     Use of human BAC End Sequences for Sequence-Ready Map Building
  JOURNAL   Unpublished (1998)
COMMENT     Other_GSSs: RPCI11-55K19.TJ
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            For clone availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .446
                     /organism="Homo sapiens"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
```



```
                     RPCI11 Human Male BAC Library"
                     /db_xref="taxon:9606"
                     /clone="R-55K19"
                     /clone_lib="RPCI11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
BASE COUNT      140 a     61 c     72 g    173 t
ORIGIN

  Query Match           10.3%;  Score 18;  DB 27;  Length 446;
  Best Local Similarity 100.0%;  Pred. No. 3.66e-03;
  Matches    18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    353 AAAAATAAATTATTTTTA 370
          ||||||||||||||||||
Cp     56 AAAAATAAATTATTTTTA 39


RESULT   19
LOCUS       C93203        453 bp    mRNA            EST       03-JUN-1998
DEFINITION  Dictyostelium discoideum slug cDNA, clone SSM769, mRNA sequence.
ACCESSION   C93203
NID         g3075079
KEYWORDS    EST; EST(expressed sequence tag).
SOURCE      Dictyostelium discoideum (strain:AX4) slug cDNA to mRNA,
            clone_lib:SS.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Dictyostelida; Dictyostelium.
REFERENCE   1  (bases 1 to 453)
  AUTHORS   Urushihara,H.
  TITLE     Developmental cDNA in Dictyostelium discoideum
  JOURNAL   Published Only in DataBase (1998) In press
REFERENCE   2  (bases 1 to 453)
  AUTHORS   Urushihara,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-APR-1998) to the DDBJ/EMBL/GenBank databases. Hideko
            Urushihara, University of Tsukuba, Institute of Biological
            Sciences; 1-1-1 Ten-noudai, Tsukuba-shi 305-8572, Japan
            (E-mail:gxurushi@bank.dna.affrc.go.jp, Tel:+81-0298-53-4664,
            Fax:+81-0298-53-6614)
COMMENT     PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES             Location/Qualifiers
     source          1. .453
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone_lib="SS"
                     /dev_stage="slug"
BASE COUNT      222 a     44 c     53 g    134 t
ORIGIN
```



```
  Query Match           10.3%;  Score 18;  DB 14;  Length 453;
  Best Local Similarity 100.0%;  Pred. No. 3.66e-03;
  Matches    18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    289 TTAAAAATAATTTATTTT 306
          ||||||||||||||||||
Qy     38 TTAAAAATAATTTATTTT 55


RESULT   20
LOCUS       AA138125      480 bp    mRNA            EST       16-FEB-1997
DEFINITION  mr05g10.r1 Soares mouse 3NbMS Mus musculus cDNA clone 596610 5',
            mRNA sequence.
ACCESSION   AA138125
NID         g1700530
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Mus.
REFERENCE   1  (bases 1 to 480)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:362042
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 475.
FEATURES             Location/Qualifiers
     source          1. .480
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGCTGTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
```



```
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. RNA
                     provided by Dr. Bertrand Jordan. Library went through
                     three rounds of normalization, and was constructed by
                     Bento Soares and M.Fatima Bonaldo."
                     /db_xref="taxon:10090"
                     /clone="596610"
                     /clone_lib="Soares mouse 3NbMS"
                     /sex="male"
                     /tissue_type="Spleen"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
     mRNA            <1. .>480
BASE COUNT      167 a     65 c     62 g    186 t
ORIGIN

  Query Match           10.3%;  Score 18;  DB 23;  Length 480;
  Best Local Similarity 100.0%;  Pred. No. 3.66e-03;
  Matches    18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    280 TGTTTTATTTTTATCTGG 297
          ||||||||||||||||||
Cp    155 TGTTTTATTTTTATCTGG 138


RESULT   21
LOCUS       AA619999      481 bp    mRNA            EST       09-OCT-1997
DEFINITION  vl76e01.r1 Knowles Solter mouse blastocyst B1 Mus musculus cDNA
            clone 978168 5', mRNA sequence.
ACCESSION   AA619999
NID         g2523875
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 481)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
```

```
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:558896
            High quality sequence stop: 459.
FEATURES              Location/Qualifiers
     source           1. .481
                      /organism="Mus musculus"
                      /strain="B6D2 F1/J"
                      /note="Organ: embryo; Vector: pSPORT; Site_1: NotI;
                      Site_2: SalI; Cloned unidirectionally from mRNA prepared
                      from 800 blastocysts.  Primer: SalI(dT):
                      5'-CGGTCGACCGTCGACCGTTTTTTTTTTTTTT-3'.  cDNAs were
                      cloned into the NotI/SalI sites of a pSPORT vector (Life
                      Technologies).  Two different size selections:  B1 (larger
                       inserts) and B3."
                      /db_xref="taxon:10090"
                      /clone="978168"
                      /clone_lib="Knowles Solter mouse blastocyst B1"
                      /tissue_type="blastocyst"
                      /dev_stage="embryo (pre-implantation)"
                      /lab_host="DH10B"
BASE COUNT      151 a    104 c    102 g    124 t
ORIGIN

  Query Match           10.3%;  Score 18;  DB 10;  Length 481;
  Best Local Similarity 100.0%;  Pred. No. 3.66e-03;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    443 AAAATAAATTATTTTTAA 460
          ||||||||||||||||||
Cp     55 AAAATAAATTATTTTTAA 38


RESULT   22
LOCUS       AA020977      487 bp    mRNA            EST       30-JAN-1997
DEFINITION  zs55d07.s1 Soares retina N2b4HR Homo sapiens cDNA clone 363853 3',
            mRNA sequence.
ACCESSION   AA020977
NID         g1484730
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 487)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
            Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     WashU-Merck EST Project
```



```
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1848  Std Error: 0.00
            Seq primer: -40M13 fwd. from Amersham
            High quality sequence stop: 271.
FEATURES              Location/Qualifiers
     source           1. .487
                      /organism="Homo sapiens"
                      /note="Organ: eye; Vector: pT7T3D (Pharmacia) with a
                      modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                      strand cDNA was primed with a Not I - oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCGCTTTTTTTTTTTTTTTTT 3'],
                      double-stranded cDNA was size selected, ligated to Eco RI
                      adapters (Pharmacia), digested with Not I and cloned into
                      the Not I and Eco RI sites of a modified pT7T3 vector
                      (Pharmacia). The retinas were obtained from a 55 year old
                      Caucasian  and total cellular poly(A)+ RNA was extracted 6
                      hrs after their removal. The retina RNA  was kindly
                      provided by Roderick R. McInnes M.D. Ph.D.  from the
                      University  of Toronto. Library constructed by Bento
                      Soares and M.Fatima Bonaldo."
                      /db_xref="taxon:9606"
                      /clone="363853"
                      /clone_lib="Soares retina N2b4HR"
                      /sex="male"
                      /tissue_type="retina"
                      /dev_stage="55 year old"
                      /lab_host="DH10B (ampicillin resistant)"
     mRNA             complement(<1. .>487)
BASE COUNT      109 a     72 c     82 g    219 t      5 others
ORIGIN

  Query Match           10.3%;  Score 18;  DB 22;  Length 487;
  Best Local Similarity 100.0%;  Pred. No. 3.66e-03;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     23 ACTTGTTTTATTTTTATC 40
          ||||||||||||||||||
Cp    158 ACTTGTTTTATTTTTATC 141


RESULT   23
```



```
LOCUS       AI070018      523 bp    mRNA            EST       05-AUG-1998
DEFINITION  UI-R-C1-ln-a-07-0-UI.s1 UI-R-C1 Rattus norvegicus cDNA clone
            UI-R-C1-ln-a-07-0-UI 3', mRNA sequence.
ACCESSION   AI070018
NID         g3396269
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 523)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT
            Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence tag present in the cDNA between the NotI site and the
            oligo-dT track served to identify it as a clone from the normalized
            adult Muscle library.  cDNA Library Preparation: M. Fatima Bonaldo,
            Ph.D. Clone distribution: clones will be available through Research
            Genetics
            Seq primer: M13 Forward.
FEATURES              Location/Qualifiers
     source           1. .523
                      /organism="Rattus norvegicus"
                      /strain="Sprague-Dawley"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-C1
                      library is a subtracted library derived from the UI-R-C0
                      library, which is a subtracted library derived from the
                      UI-R-A1 and UI-R-E1 libraries. The UI-R-A1 library
                      consisted of a mixture of  individually tagged normalized
                      libraries constructed from rat placenta,  adult lung,
                      brain, liver, kidney, heart, spleen, ovary, and muscle.
                      The UI-R-E1 library  consisted of a mixture of
                      individually tagged normalized libraries constructed from
                      8, 12 and 18-day embryo. The tag is a string of 3-5
                      nucleotides present between the Not I site and the
                      oligo-dT track which allows identification of the library
                      of origin of a clone within the mixture. The subtracted
                      library (UI-R-C1) was constructed as follows: PCR
                      amplified cDNA inserts from  UI-R-C0 clones from which 3'
                      ESTs had been derived was used as a driver in a
```



```
                      hybridization with the UI-R-C0 library in the form of
                      single-stranded circles. The remaining single-stranded
                      circles (subtracted library) was purified by
                      hydroxyapatite column chromatography, converted to
                      double-stranded circles and electroporated into DH10B
                      bacteria (Life Technologies) to generate the UI-R-C1
                      library. This procedure has been previously described
                      (Bonaldo, Lennon and Soares, Genome Research 6: 791-806,
                      1996)"
                      /db_xref="taxon:10116"
                      /clone="UI-R-C1-ln-a-07-0-UI"
                      /clone_lib="UI-R-C1"
                      /dev_stage="Adult"
                      /lab_host="DH10B (Life Technologies)"
BASE COUNT      173 a     94 c    120 g    136 t
ORIGIN

  Query Match           10.3%;  Score 18;  DB 17;  Length 523;
  Best Local Similarity 100.0%;  Pred. No. 3.66e-03;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     16 TTAAAAATAATTTATTTT 33
          ||||||||||||||||||
Qy     38 TTAAAAATAATTTATTTT 55


RESULT   24
LOCUS       AA106432      535 bp    mRNA            EST       04-FEB-1997
DEFINITION  m194c03.r1 Stratagene mouse kidney (#937315) Mus musculus cDNA
            clone 519652 5', mRNA sequence.
ACCESSION   AA106432
NID         g1656896
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Mus.
REFERENCE   1  (bases 1 to 535)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
```

```
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:313500
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 444.
FEATURES            Location/Qualifiers
     source         1. .535
                    /organism="Mus musculus"
                    /strain="C57/B16"
                    /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                    XhoI; Cloned unidirectionally.  Primer: Oligo dT. Average
                    insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor
                    sequence: 5' GAATTCGGCACGAG 3' -3' adaptor sequence: 5'
                    CTCGAGTTTTTTTTTTTTTTTTT 3'"
                    /db_xref="taxon:10090"
                    /clone="519652"
                    /clone_lib="Stratagene mouse kidney (#937315)"
                    /sex="females"
                    /dev_stage="4 weeks"
                    /lab_host="SOLR (kanamycin resistant)"
     mRNA           <1. .>535
BASE COUNT      189 a     90 c     78 g    178 t
ORIGIN

  Query Match           10.3%;  Score 18;  DB 23;  Length 535;
  Best Local Similarity 100.0%;  Pred. No. 3.66e-03;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    411 TGTTTTATTTTTATCTGG 428
          ||||||||||||||||||
Cp    155 TGTTTTATTTTTATCTGG 138


RESULT   25
LOCUS       AA204477      560 bp    mRNA            EST       19-FEB-1997
DEFINITION  mu32e11.r1 Soares 2NbMT Mus musculus cDNA clone 641132 5', mRNA
            sequence.
ACCESSION   AA204477
NID         g1801038
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Mus.
REFERENCE   1  (bases 1 to 560)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
```



```
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:393124
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 495.
FEATURES            Location/Qualifiers
     source         1. .560
                    /organism="Mus musculus"
                    /strain="C57BL/6J"
                    /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                    polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                    was primed with a Not I - oligo(dT) primer [5'
                    TGTTACCAATCTGAAGTGGGAGCGGCCGCGTTTTTTTTTTTTTTTTTTTTTTTT
                    3']; double-stranded cDNA was ligated to Eco RI adaptors
                    (Pharmacia), digested with Not I and cloned into the Not I
                    and Eco RI sites of the modified pT7T3 vector. RNA
                    provided by Dr. Bertrand Jordan. Library went through two
                    rounds of normalization, and was constructed by Bento
                    Soares and M.Fatima Bonaldo."
                    /db_xref="taxon:10090"
                    /clone="641132"
                    /clone_lib="Soares 2NbMT"
                    /sex="male"
                    /tissue_type="Thymus"
                    /dev_stage="4 weeks"
                    /lab_host="DH10B"
     mRNA           <1. .>560
BASE COUNT      205 a     81 c     73 g    201 t
ORIGIN

  Query Match           10.3%;  Score 18;  DB 23;  Length 560;
  Best Local Similarity 100.0%;  Pred. No. 3.66e-03;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    341 TGTTTTATTTTTATCTGG 358
          ||||||||||||||||||
Cp    155 TGTTTTATTTTTATCTGG 138
```



```
RESULT   26
LOCUS       AI131650      586 bp    mRNA            EST       14-SEP-1998
DEFINITION  ui41b01.x1 Sugano mouse embryo mewa Mus musculus cDNA clone 1884937
            3', mRNA sequence.
ACCESSION   AI131650
NID         g1601666
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 586)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:969261
            Seq primer: custom primer used
            High quality sequence stop: 500.
FEATURES            Location/Qualifiers
     source         1. .586
                    /organism="Mus musculus"
                    /strain="C57BL"
                    /note="Vector: pME18S-FL3; Site_1: DraIII (CACTGTGTG);
                    Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed
                    with an oligo(dT) primer  [ATGTGGCCTTTTTTTTTTTTTTTT];
                    double-stranded cDNA was ligated to a  DraIII adaptor
                    [TGTTGGCCTACTGG], digested and cloned into distinct DraIII
                    sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site
                    CACCATGTG). XhoI should be used to isolate the cDNA
                    insert.  Size selection was performed to exclude fragments
                    <1.5kb.  Library constructed by Dr. Sumio Sugano
                    (University of Tokyo Institute of Medical Science).
                    Custom primers  for sequencing: 5' end primer
                    CTTCTGCTCTAAAAGCTGCG and 3' end primer
                    CGACCTGCAGCTCGAGCACA."
                    /db_xref="taxon:10090"
```



```
                    /clone="1884937"
                    /clone_lib="Sugano mouse embryo mewa"
                    /dev_stage="embryo, 14 dpc"
                    /lab_host="DH10B"
BASE COUNT      224 a     84 c     77 g    201 t
ORIGIN

  Query Match           10.3%;  Score 18;  DB 17;  Length 586;
  Best Local Similarity 100.0%;  Pred. No. 3.66e-03;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    536 CCAGATAAAAATAAAACA 553
          ||||||||||||||||||
Qy    138 CCAGATAAAAATAAAACA 155


RESULT   27
LOCUS       W73116        602 bp    mRNA            EST       16-OCT-1996
DEFINITION  zd55d10.r1 Soares fetal heart NbHH19W Homo sapiens cDNA clone
            344563 5', mRNA sequence.
ACCESSION   W73116
NID         g1383318
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 602)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Putative full length read
            Insert Length: 689  Std Error: 0.00
            Seq primer: ETPrimer
            High quality sequence stop: 454.
FEATURES            Location/Qualifiers
     source         1. .602
```

```
                     /organism="Homo sapiens"
                     /note="Organ: heart; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATCTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by
                     M.Fatima Bonaldo. This library was constructed from the
                     same fetus as the fetal lung library, Soares fetal lung
                     NbHL19W."
                     /db_xref="taxon:9606"
                     /clone="344563"
                     /clone_lib="Soares fetal heart NbHH19W"
                     /sex="unknown"
                     /dev_stage="19 weeks"
                     /lab_host="DH10B (ampicillin resistant)"
     mRNA            <1. .>602
BASE COUNT      175 a     91 c    129 g    207 t
ORIGIN

  Query Match           10.3%;  Score 18;  DB 21;  Length 602;
  Best Local Similarity 100.0%;  Pred. No. 3.66e-03;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    385 AAAAATAAATTATTTTTA 402
          ||||||||||||||||||
Cp     56 AAAAATAAATTATTTTTA 39


RESULT   28
LOCUS       AA794906      603 bp     mRNA            EST       09-FEB-1998
DEFINITION  vr48e06.s1 Knowles Solter mouse 2 cell Mus musculus cDNA clone
            1123906 5', mRNA sequence.
ACCESSION   AA794906
NID         g2857861
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 603)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
```



```
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:613242
            High quality sequence stop: 479.
FEATURES              Location/Qualifiers
     source           1. .603
                      /organism="Mus musculus"
                      /strain="B6D2 F1/J"
                      /note="Organ: embryo; Vector: pBluescribe (modified);
                      Site_1: MluI; Site_2: SalI; Cloned unidirectionally from
                      mRNA prepared from 13,500 2-cell stage  embryos.  Primer:
                      SalI(dT): 5'-CGGTCGACCGTCGACCGTTTTTTTTTTTTTTT-3'.   cDNAs
                      were cloned into the MluI/SalI sites of a modified
                      pBluescribe  vector using commercial linkers (NEB).
                      Average insert size:  1.2 kb."
                      /db_xref="taxon:10090"
                      /clone="1123906"
                      /clone_lib="Knowles Solter mouse 2 cell"
                      /tissue_type="embryo"
                      /dev_stage="2-cell"
                      /lab_host="DH10B"
BASE COUNT      203 a     84 c     88 g    227 t      1 others
ORIGIN

  Query Match           10.3%;  Score 18;  DB 12;  Length 603;
  Best Local Similarity 100.0%;  Pred. No. 3.66e-03;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    215 TGTTTTATTTTTATCTGG 232
          ||||||||||||||||||
Cp    155 TGTTTTATTTTTATCTGG 138


RESULT   29
LOCUS       W73136        621 bp     mRNA            EST       16-OCT-1996
DEFINITION  zd55d10.s1 Soares fetal heart NbHH19W Homo sapiens cDNA clone
            344563 3' similar to contains Alu repetitive element;, mRNA
            sequence.
ACCESSION   W73136
NID         g1383271
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
```



```
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 621)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 689   Std Error: 0.00
            Seq primer: ETPrimer
            High quality sequence stop: 405.
FEATURES              Location/Qualifiers
     source           1. .621
                      /organism="Homo sapiens"
                      /note="Organ: heart; Vector: pT7T3D (Pharmacia) with a
                      modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                      strand cDNA was primed with a Not I - oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCATCTTTTTTTTTTTTTTTT 3'],
                      double-stranded cDNA was size selected, ligated to Eco RI
                      adapters (Pharmacia), digested with Not I and cloned into
                      the Not I and Eco RI sites of a modified pT7T3 vector
                      (Pharmacia). Library went through one round of
                      normalization to a Cot = 5. Library constructed by
                      M.Fatima Bonaldo. This library was constructed from the
                      same fetus as the fetal lung library, Soares fetal lung
                      NbHL19W."
                      /db_xref="taxon:9606"
                      /clone="344563"
                      /clone_lib="Soares fetal heart NbHH19W"
                      /sex="unknown"
                      /dev_stage="19 weeks"
                      /lab_host="DH10B (ampicillin resistant)"
     mRNA             complement(<1. .>621)
BASE COUNT      207 a    130 c     95 g    186 t      3 others
ORIGIN

  Query Match           10.3%;  Score 18;  DB 21;  Length 621;
  Best Local Similarity 100.0%;  Pred. No. 3.66e-03;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```



```
Db    219 TAAAAATAATTTATTTTT 236
          ||||||||||||||||||
Qy     39 TAAAAATAATTTATTTTT 56


RESULT   30
LOCUS       AQ041937      720 bp     DNA             GSS       13-JUL-1998
DEFINITION  CIT-HSP-2335B16.TF CIT-HSP Homo sapiens genomic clone 2335B16,
            genomic survey sequence.
ACCESSION   AQ041937
NID         g3311131
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 720)
  AUTHORS   Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K.,
            Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
            Simon,M. and Venter,J.C.
  TITLE     Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: CIT-HSP-2335B16.TR
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13-21
            Class: BAC ends.
FEATURES              Location/Qualifiers
     source           1. .720
                      /organism="Homo sapiens"
                      /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                      HindIII"
                      /db_xref="taxon:9606"
                      /clone="2335B16"
                      /clone_lib="CIT-HSP"
                      /sex="Male"
                      /cell_type="Sperm"
BASE COUNT      252 a    112 c     85 g    271 t
ORIGIN

  Query Match           10.3%;  Score 18;  DB 27;  Length 720;
```

```
  Best Local Similarity 100.0%;  Pred. No. 3.66e-03;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    333 TAAAAATAATTTATTTTT 350
          ||||||||||||||||||
Qy     39 TAAAAATAATTTATTTTT 56


RESULT   31
LOCUS       AA337232      187 bp    mRNA            EST       21-APR-1997
DEFINITION  EST42297 Endometrial tumor Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION   AA337232
NID         g1989790
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 187)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
            Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
            White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
            Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
            Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
            Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
            Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
            He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
            Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H.,
            Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
            Dillon,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
            Fraser,C.M. and Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   96026280
COMMENT     Other_ESTs: THC140337
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
```



```
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES              Location/Qualifiers
     source           1. .187
                      /organism="Homo sapiens"
                      /note="Organ: endometrium; Vector: pBluescript SK-;
                      Site_1: EcoRI; Site_2: XhoI"
                      /db_xref="ATCC (inhost):139303"
                      /db_xref="taxon:9606"
                      /clone_lib="Endometrial tumor"
                      /sex="female"
                      /dev_stage="adult"
     mRNA             <1. .>187
BASE COUNT      54 a     30 c     27 g     75 t      1 others
ORIGIN

  Query Match           9.8%;  Score 17;  DB 25;  Length 187;
  Best Local Similarity 100.0%;  Pred. No. 6.68e-02;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    143 CTTGTTTTATTTTTATC 159
          |||||||||||||||||
Cp    157 CTTGTTTTATTTTTATC 141


RESULT   32
LOCUS       AA999152      205 bp    mRNA            EST       05-JUN-1998
DEFINITION  UI-R-E1-fy-h-12-0-UI.s2 UI-R-E1 Rattus norvegicus cDNA clone
            UI-R-E1-fy-h-12-0-UI 3', mRNA sequence.
ACCESSION   AA999152
NID         g3189743
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 205)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT
            Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence tag present in the cDNA between the NotI site and the
```



```
            oligo-dT track (not shown) served to identify it as a clone from
            the normalized adult 18-Day-Embryo library. cDNA Library
            Preparation: M. Fatima Bonaldo, Ph.D. Clone distribution: clones
            will be available through Research Genetics
            Seq primer: M13 Forward.
FEATURES              Location/Qualifiers
     source           1. .205
                      /organism="Rattus norvegicus"
                      /strain="Sprague-Dawley"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-E1
                      library is a subtracted library derived from the UI-R-E0
                      library.  The UI-R-E0 library  consisted of a mixture of
                      individually  tagged normalized libraries constructed from
                      8, 12 and 18-day embryo.   The tag is a string of 3-5
                      nucleotides present between the Not I site  and the
                      oligo-dT track which allows identification of the library
                      of  origin of a clone within the mixture.  The subtracted
                      library (UI-R-E1) was constructed as follows: PCR
                      amplified cDNA inserts from a pool of  UI-R-E0 clones from
                      which 3' ESTs had been derived was used as a driver in a
                      hybridization with the UI-R-E0 library in the form of
                      single-stranded circles. The remaining single-stranded
                      circles (subtracted library) was purified by
                      hydroxyapatite column chromatography, converted to
                      double-stranded circles and electroporated into DH10B
                      bacteria (Life Technologies) to generate the UI-R-E1
                      library. This procedure has been previously described
                      (Bonaldo, Lennon and Soares, Genome Research 6: 791-806,
                      1996)"
                      /db_xref="taxon:10116"
                      /clone="UI-R-E1-fy-h-12-0-UI"
                      /clone_lib="UI-R-E1"
                      /dev_stage="adult"
                      /lab_host="DH10B (Life Technologies)"
BASE COUNT      80 a     35 c     36 g     54 t
ORIGIN

  Query Match           9.8%;  Score 17;  DB 15;  Length 205;
  Best Local Similarity 100.0%;  Pred. No. 6.68e-02;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     64 AAAATAAAACAAGTGAA 80
          |||||||||||||||||
Qy    145 AAAATAAAACAAGTGAA 161


RESULT   33
LOCUS       HUML11312     208 bp    mRNA            EST       08-FEB-1995
DEFINITION  Human fetal-lung cDNA 5'-end sequence, mRNA sequence.
ACCESSION   D30880
```



```
NID         g643760
KEYWORDS    EST; EST(expressed sequence tag).
SOURCE      Homo sapiens Fetal Lung cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 208)
  AUTHORS   Sudo,K., Chinen,K. and Nakamura,Y.
  TITLE     2058 expressed sequence tags (ESTs) from a human fetal lung cDNA
            library
  JOURNAL   Genomics 24 (2), 276-279 (1994)
  MEDLINE   95213017
REFERENCE   2  (bases 1 to 208)
  AUTHORS   Nakamura,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAY-1994) to the DDBJ/EMBL/GenBank databases. Yusuke
            Nakamura, Cancer Institute, Department of Biochemistry; 1-37-1
            Kami-Ikebukuro, Toshima-ku, Tokyo 170, Japan
            (E-mail:nakamura@ganvx1.jfcr.or.jp, Tel:03-3918-0111(ex.4501),
            Fax:03-3918-0342)
COMMENT     Submitted (01-May-1994) to DDBJ by:
            Yusuke Nakamura
            Cancer Institute
            Department of Biochemistry
            1-37-1 Kami-Ikebukuro
            Toshima-ku, Tokyo
            170 Japan
            Phone: 03-5394-3866
            Fax  : 03-3918-0342.
FEATURES              Location/Qualifiers
     source           1. .208
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /dev_stage="Fetal"
                      /tissue_type="Lung"
BASE COUNT      69 a     28 c     38 g     67 t      6 others
ORIGIN

  Query Match           9.8%;  Score 17;  DB 7;  Length 208;
  Best Local Similarity 100.0%;  Pred. No. 6.68e-02;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    130 TACATTAGAAATAAGAT 146
          |||||||||||||||||
Cp     99 TACATTAGAAATAAGAT 83


RESULT   34
LOCUS       C94090        212 bp    mRNA            EST       11-JUN-1998
DEFINITION  Dictyostelium discoideum slug cDNA, clone SSG392, mRNA sequence.
```

```
ACCESSION   C94090
NID         g3218705
KEYWORDS    EST; EST(expressed sequence tag).
SOURCE      Dictyostelium discoideum (strain:AX4) slug cDNA to mRNA,
            clone_lib:SS.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Dictyostelida; Dictyostelium.
REFERENCE   1  (bases 1 to 212)
  AUTHORS   Urushihara,H.
  TITLE     Developmental cDNA in Dictyostelium discoideum
  JOURNAL   Published Only in DataBase (1998) In press
REFERENCE   2  (bases 1 to 212)
  AUTHORS   Urushihara,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-MAY-1998) to the DDBJ/EMBL/GenBank databases. Hideko
            Urushihara, University of Tsukuba, Institute of Biological
            Sciences; 1-1-1 Ten-noudai, Tsukuba-shi 305-8572, Japan
            (E-mail:gkurushi@ebank.dna.affrc.go.jp, Tel:+81-298-53-4664,
            Fax:+81-298-53-6614)
COMMENT     PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES             Location/Qualifiers
     source          1. .212
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone_lib="SS"
                     /dev_stage="slug"
BASE COUNT       91 a      7 c      8 g     105 t      1 others
ORIGIN

  Query Match           9.8%;  Score 17;  DB 15;  Length 212;
  Best Local Similarity  94.4%;  Pred. No. 6.68e-02;
  Matches    17;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db      31 AAATAAATTATTTTTANC 48
           ||||||||||||||||  |
Cp      54 AAATAAATTATTTTTAAC 37


RESULT   35
LOCUS       HUML1371      213 bp     mRNA            EST       08-FEB-1995
DEFINITION  Human fetal-lung cDNA 5'-end sequence, mRNA sequence.
ACCESSION   D31403
NID         g644283
KEYWORDS    EST; EST(expressed sequence tag).
SOURCE      Homo sapiens Fetal Lung cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 213)
```



```
  AUTHORS   Sudo,K., Chinen,K. and Nakamura,Y.
  TITLE     2058 expressed sequence tags (ESTs) from a human fetal lung cDNA
            library
  JOURNAL   Genomics 24 (2), 276-279 (1994)
  MEDLINE   95213017
REFERENCE   2  (bases 1 to 213)
  AUTHORS   Nakamura,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAY-1994) to the DDBJ/EMBL/GenBank databases. Yusuke
            Nakamura, Cancer Institute, Department of Biochemistry; 1-37-1
            Kami-Ikebukuro, Toshima-ku, Tokyo 170, Japan
            (E-mail:nakamura@ganvx1.jfcr.or.jp, Tel:03-3918-0111(ex.4501),
            Fax:03-3918-0342)
COMMENT     Submitted (01-May-1994) to DDBJ by:
            Yusuke Nakamura
            Cancer Institute
            Department of Biochemistry
            1-37-1 Kami-Ikebukuro
            Toshima-ku, Tokyo
            170 Japan
            Phone: 03-5394-3866
            Fax  : 03-3918-0342.
FEATURES             Location/Qualifiers
     source          1. .213
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /dev_stage="Fetal"
                     /tissue_type="Lung"
BASE COUNT       77 a      27 c      34 g      72 t       3 others
ORIGIN

  Query Match           9.8%;  Score 17;  DB 7;  Length 213;
  Best Local Similarity 100.0%;  Pred. No. 6.68e-02;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     127 TACATTAGAAATAAGAT 143
           |||||||||||||||||
Cp      99 TACATTAGAAATAAGAT 83


RESULT   36
LOCUS       AA476057      242 bp     mRNA            EST       18-JUN-1997
DEFINITION  vh26f02.r1 Soares mouse mammary gland NbMMG Mus musculus cDNA clone
            876603 5', mRNA sequence.
ACCESSION   AA476057
NID         g2203908
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae;
```



```
            Murinae; Mus.
REFERENCE   1  (bases 1 to 242)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:516083
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 207.
FEATURES             Location/Qualifiers
     source          1. .242
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATGGTTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
                     constructed and normalized by Bento Soares and M.Fatima
                     Bonaldo."
                     /db_xref="taxon:10090"
                     /clone="876603"
                     /clone_lib="Soares mouse mammary gland NbMMG"
                     /sex="male"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
     mRNA            <1. .>242
BASE COUNT       45 a      62 c      63 g      72 t
ORIGIN

  Query Match           9.8%;  Score 17;  DB 7;  Length 242;
  Best Local Similarity 100.0%;  Pred. No. 6.68e-02;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     104 GAGCTTTGCCTGCTCAG 120
```



```
           |||||||||||||||||
Cp      20 GAGCTTTGCCTGCTCAG 4


RESULT   37
LOCUS       HUM420E11B    256 bp     mRNA            EST       21-MAY-1996
DEFINITION  Human fetal brain cDNA 5'-end GEN-420E11, mRNA sequence.
ACCESSION   D61584
NID         g970750
KEYWORDS    EST; EST(expressed sequence tag); Human fetal brain; similar to
            none(May 29,1995).
SOURCE      Homo sapiens cDNA to mRNA, clone_lib:Clontech human fetal brain
            polyA+ mRNA (#6535).
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 256)
  AUTHORS   Fujiwara,T., Hirano,H., Katagiri,T., Kawai,A., Kuga,Y., Nagata,M.,
            Okuno,S., Ozaki,K., Shimizu,F., Shimada,Y., Shinomiya,H.,
            Takaichi,A., Takeda,S., Watanabe,T., Takahashi,E., Hirai,Y.,
            Maekawa,H., Shin,S. and Nakamura,Y.
  TITLE     Unpublished(202)
  JOURNAL   Unpublished (1995)
REFERENCE   2  (bases 1 to 256)
  AUTHORS   Fujiwara,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAY-1995) to the DDBJ/EMBL/GenBank databases. Tsutomu
            Fujiwara, Otsuka GEN Research Institute,Otsuka Pharmaceutical
            Co.,Ltd; 463-10 Kagasuno Kawauchi-cho, Tokushima, Tokushima 771-01,
            Japan (Tel:0886-65-2888, Fax:0886-37-1035)
COMMENT     Submitted (30-May-1995) to DDBJ by:
            Tsutomu Fujiwara
            Otsuka GEN Research Institute
            Otsuka Pharmaceutical Co.,Ltd
            463-10 kagasuno Kawauchi-cho
            Tokushima, Tokushima
            771-01
            Japan
            Phone: 0886-65-2888
            Fax  : 0886-37-1035.
FEATURES             Location/Qualifiers
     source          1. .256
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="Clontech human fetal brain polyA+ mRNA
                     (#6535)"
BASE COUNT       81 a      42 c      34 g      99 t
ORIGIN

  Query Match           9.8%;  Score 17;  DB 7;  Length 256;
```

```
  Best Local Similarity 100.0%;  Pred. No. 6.68e-02;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     217 CTTGTTTTATTTTTATC 233
           |||||||||||||||||
Cp     157 CTTGTTTTATTTTTATC 141


RESULT   38
LOCUS       AA327070       273 bp    mRNA            EST       20-APR-1997
DEFINITION  EST30320 Colon I Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION   AA327070
NID         g1979317
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 273)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
            Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
            White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
            Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
            Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
            Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
            Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
            He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
            Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H.,
            Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
            Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
            Fraser,C.M. and Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   96026280
COMMENT
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
```



```
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES              Location/Qualifiers
     source           1. .273
                      /organism="Homo sapiens"
                      /note="Organ: colon; Vector: pBluescript SK-; Site_1:
                      EcoRI; Site_2: XhoI"
                      /db_xref="ATCC (inhost):127730"
                      /db_xref="taxon:9606"
                      /clone_lib="Colon I"
                      /dev_stage="adult"
     mRNA             <1. .>273
BASE COUNT       91 a     49 c     47 g     85 t      1 others
ORIGIN

  Query Match           9.8%;  Score 17;  DB 24;  Length 273;
  Best Local Similarity 100.0%;  Pred. No. 6.68e-02;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      74 ATAAAACAAGTGAATTT 90
           |||||||||||||||||
Qy     148 ATAAAACAAGTGAATTT 164


RESULT   39
LOCUS       T93193         274 bp    mRNA            EST       22-MAR-1995
DEFINITION  ye24g08.r1 Homo sapiens cDNA clone 118718 5' similar to contains
            Alu repetitive element;.
ACCESSION   T93193
NID         g725106
KEYWORDS    EST.
SOURCE      human clone=118718 library=Stratagene lung (#937210)
            vector=pBluescript SK- host=SOLR cells (kanamycin resistant)
            primer=M13RP1 RsiteI=EcoRI RsiteJ=XhoI Normal lung tissue from a 72
            year old male. Cloned unidirectionally. Primer: Oligo dT. Average
            insert size: 1.0 kb; Uni-ZAP XR Vector; 5' adaptor sequence:
            5'-GAATTCGGCACGAG-3'; 3' adaptor sequence:
            5'-CTCGAGTTTTTTTTTTTTTTTT-3'.
  ORGANISM  Homo sapiens
            Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 274)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
            Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project.
```



```
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 153
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES              Location/Qualifiers
     source           1. .274
                      /organism="Homo sapiens"
                      /clone="118718"
BASE COUNT       64 a     52 c     55 g    102 t      1 others
ORIGIN

  Query Match           9.8%;  Score 17;  DB 5;  Length 274;
  Best Local Similarity 100.0%;  Pred. No. 6.68e-02;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     192 TTGTTTTATTTTTATCT 208
           |||||||||||||||||
Cp     156 TTGTTTTATTTTTATCT 140


RESULT   40
LOCUS       W40029         285 bp    mRNA            EST       12-AUG-1996
DEFINITION  T2642 MVAT4 bloodstream form of serodeme WRATat1.1 Trypanosoma
            brucei rhodesiense cDNA 5', mRNA sequence.
ACCESSION   W40029
NID         g1324050
KEYWORDS    EST.
SOURCE      Trypanosoma brucei rhodesiense.
  ORGANISM  Trypanosoma brucei rhodesiense
            Eukaryotae; mitochondrial eukaryotes; Euglenozoa; Kinetoplastida;
            Trypanosomatidae; Trypanosoma.
REFERENCE   1  (bases 1 to 285)
  AUTHORS   Djikeng,A., Donelson,J.E. and Majiwa,P.A.O.
  TITLE     Generation of expressed sequence tags as physical landmarks in the
            genome of Trypanosoma brucei
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Majiwa PAO
            Molecular Biology Unit
            International Livestock Research Institute
            P.O. Box 30709, Nairobi, Kenya
            Tel: 254-2 630743
            Fax: 254-2 631499
            Email: p.majiwa@cgnet.com
            Seq primer: T3 primer.
FEATURES              Location/Qualifiers
```



```
     source           1. .285
                      /organism="Trypanosoma brucei rhodesiense"
                      /note="Vector: Lambda ZAP II (Stratagene); Site_1: EcorI;
                      Site_2: XhoI; The mRNA was purified from a cloned
                      population of bloodstream trypanosomes reexpressing the
                      MVAT4 metacyclic variant surface glycoprotein (VSG). A
                      unidirectional oligo dT-primed EcoRI/XhoI cDNA library was
                      constructed in lambda ZAP II (Stratagene)."
                      /db_xref="taxon:31286"
                      /clone_lib="MVAT4 bloodstream form of serodeme WRATat1.1"
     mRNA             <1. .>285
BASE COUNT      105 a     51 c     62 g     67 t
ORIGIN

  Query Match           9.8%;  Score 17;  DB 21;  Length 285;
  Best Local Similarity 100.0%;  Pred. No. 6.68e-02;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     230 ATAAAAATAAAACAAGT 246
           |||||||||||||||||
Qy     142 ATAAAAATAAAACAAGT 158


RESULT   41
LOCUS       AA281555       287 bp    mRNA            EST       14-AUG-1997
DEFINITION  st04a01.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:712104 3'
            similar to contains Alu repetitive element;, mRNA sequence.
ACCESSION   AA281555
NID         g1924233
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 287)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 801    Std Error: 0.00
            Seq primer: -41m13 fwd. ET from Amersham
            High quality sequence stop: 231.
FEATURES              Location/Qualifiers
     source           1. .287
```

```
                     /organism="Homo sapiens"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from human tonsillar cells enriched for
                     germinal center B cells by flow sorting (CD20+, IgD-),
                     provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                     (NCI) and Dr. Gerald Marti (CBER).  cDNA synthesis was
                     primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCTGAAGTCGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTT-
                     3']. Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo."
                     /db_xref="taxon:9606"
                     /clone="IMAGE:712104"
                     /clone_lib="NCI_CGAP_GCB1"
                     /tissue_type="germinal center B cell"
                     /lab_host="DH10B"
          mRNA       complement(<1. .>287)
BASE COUNT       66 a      54 c      42 g     124 t       1 others
ORIGIN

  Query Match             9.8%;  Score 17;  DB 8;  Length 287;
  Best Local Similarity 100.0%;  Pred. No. 6.68e-02;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    182 AAAATAATTTATTTTTG 198
          |||||||||||||||||
Qy     41 AAAATAATTTATTTTTG 57


RESULT   42
LOCUS       C53994         300 bp    mRNA            EST       29-AUG-1997
DEFINITION  C.elegans cDNA clone yk352a2 : 3' end, single read, mRNA sequence.
ACCESSION   C53994
NID         g2391751
KEYWORDS    EST; EST(expressed sequence tag).
SOURCE      Caenorhabditis elegans (strain:CB1489 him-8(e1489)) varied
            hermaphrodite, male whole animal cDNA to mRNA, clone_lib:Yuji
            Kohara unpublished cDNA library clone:yk352a2.
  ORGANISM  Caenorhabditis elegans
            Eukaryotae; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (sites)
  AUTHORS   Kohara,Y., Motohashi,T., Tabara,H., Shin-i,T., Watanabe,H.,
            Sano,M., Miyata,A., Ohba,T., Mitani,Y., Uesugi,H., Sugiura,I.,
            Obara,M., Sugimoto,A., Iida,K. and Nishigaki,A.
  TITLE     Expression map of the C.elegans genome
  JOURNAL   Unpublished (1997)
REFERENCE   2  (bases 1 to 300)
```



```
  AUTHORS   Kohara,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1997) to the DDBJ/EMBL/GenBank databases. Yuji
            Kohara, National Institute of Genetics, Gene Network Lab; Yata
            1111, Mishima, Shizuoka 411, Japan (E-mail:ykohara@lab.nig.ac.jp,
            Tel:+81-559-81-6854, Fax:+81-559-81-6855)
FEATURES              Location/Qualifiers
     source           1. .300
                      /organism="Caenorhabditis elegans"
                      /strain="CB1489 him-8(e1489)"
                      /db_xref="taxon:6239"
                      /clone="yk352a2"
                      /clone_lib="Yuji Kohara unpublished cDNA library"
                      /dev_stage="varied"
                      /sex="hermaphrodite, male"
                      /tissue_type="whole animal"
BASE COUNT       78 a      68 c      55 g      91 t       8 others
ORIGIN

  Query Match             9.8%;  Score 17;  DB 9;  Length 300;
  Best Local Similarity 100.0%;  Pred. No. 6.68e-02;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     51 TAGAAATAAGATTTACA 67
          |||||||||||||||||
Cp     94 TAGAAATAAGATTTACA 78


RESULT   43
LOCUS       AA263147       307 bp    mRNA            EST       02-JUL-1998
DEFINITION  PMY0513 KG1-a Lambda Zap Express cDNA library Homo sapiens cDNA 5',
            mRNA sequence.
ACCESSION   AA263147
NID         g1898953
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 307)
  AUTHORS   Claudio,J.O., Liew,C.C., Dempsey,A.A., Cukerman,E., Stewart,A.K.,
            Na,E., Atkins,H.I., Iscove,N.N. and Hawley,R.G.
  TITLE     Identification of sequence-tagged transcripts differentially
            expressed within the human hematopoietic hierarchy
  JOURNAL   Genomics 50 (1), 44-52 (1998)
  MEDLINE   98292493
COMMENT
            Contact: Hawley RG
            Oncology Research Laboratories
            The Toronto Hospital
            CRCS-424, 67 College St., Toronto, Ontario M5G 2M1, Canada
```



```
            Tel: 416 3403834
            Fax: 416 3403453
            Email: r.hawley@utoronto.ca
            Clone was randomly picked from KG1a primary library.
            Seq primer: 5' GAAATTAACCCTCACTAAAGGG 3'
            High quality sequence stop: 307.
FEATURES              Location/Qualifiers
     source           1. .307
                      /organism="Homo sapiens"
                      /note="Vector: Lambda Zap Express (Stratagene); Site_1:
                      EcoRI; Site_2: XhoI; Unidirectional cloning sites:
                      EcoR1-Xhol. mRNA was purified from KG1-a cell line, cDNA
                      was synthesized using an Xhol-OligodT linker primer. EcoR1
                      adaptors were ligated, followed by digestion with Xhol for
                      directional cloning into predigested Lambda Zap Express"
                      /db_xref="taxon:9606"
                      /clone_lib="KG1-a Lambda Zap Express cDNA library"
                      /cell_type="promyeloblast"
                      /cell_line="KG1-a"
BASE COUNT      109 a      40 c      58 g     100 t
ORIGIN

  Query Match             9.8%;  Score 17;  DB 15;  Length 307;
  Best Local Similarity 100.0%;  Pred. No. 6.68e-02;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    243 TACATTAGAAATAAGAT 259
          |||||||||||||||||
Cp     99 TACATTAGAAATAAGAT 83


RESULT   44
LOCUS       C91163         308 bp    mRNA            EST       15-APR-1998
DEFINITION  Dictyostelium discoideum slug cDNA, clone SSJ756, mRNA sequence.
ACCESSION   C91163
NID         g3060529
KEYWORDS    EST; EST(expressed sequence tag).
SOURCE      Dictyostelium discoideum (strain:AX4) slug cDNA to mRNA,
            clone_lib:SS.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Dictyostelida; Dictyostelium.
REFERENCE   1  (bases 1 to 308)
  AUTHORS   Urushihara,H.
  TITLE     Developmental cDNA in Dictyostelium discoideum
  JOURNAL   Published Only in DataBase (1998) In press
REFERENCE   2  (bases 1 to 308)
  AUTHORS   Urushihara,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAR-1998) to the DDBJ/EMBL/GenBank databases. Hideko
            Urushihara, University of Tsukuba, Institute of Biological
            Sciences; 1-1-1 Ten-noudai, Tsukuba-shi 305-8572, Japan
```



```
            (E-mail:qxurushi@bank.dna.affrc.go.jp, Tel:+81-0298-53-4664,
            Fax:+81-0298-53-6614)
COMMENT     PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES              Location/Qualifiers
     source           1. .308
                      /organism="Dictyostelium discoideum"
                      /strain="AX4"
                      /db_xref="taxon:44689"
                      /clone_lib="SS"
                      /dev_stage="slug"
BASE COUNT      123 a      40 c      31 g     114 t
ORIGIN

  Query Match             9.8%;  Score 17;  DB 14;  Length 308;
  Best Local Similarity 100.0%;  Pred. No. 6.68e-02;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    262 AAAAATAAATTATTTTT 278
          |||||||||||||||||
Cp     56 AAAAATAAATTATTTTT 40


RESULT   45
LOCUS       AA957958       313 bp    mRNA            EST       07-MAY-1998
DEFINITION  UI-R-E1-fz-g-04-0-UI.s1 UI-R-E1 Rattus norvegicus cDNA clone
            UI-R-E1-fz-g-04-0-UI 3', mRNA sequence.
ACCESSION   AA957958
NID         g3121653
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 313)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT
            Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence tag present in the cDNA between the NotI site and the
            oligo-dT track (not shown) served to identify it as a clone from
            the normalized adult 18-Day-Embryo library. cDNA Library
            Preparation: M. Fatima Bonaldo, Ph.D. Clone distribution: clones
```

```
            will be available through Research Genetics
            Seq primer: M13 Forward.
FEATURES              Location/Qualifiers
     source           1. .313
                      /organism="Rattus norvegicus"
                      /strain="Sprague-Dawley"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-E1
                      library is a subtracted library derived from the UI-R-E0
                      library.  The UI-R-E0 library  consisted of a mixture of
                      individually  tagged normalized libraries constructed from
                      8, 12 and 18-day embryo.   The tag is a string of 3-5
                      nucleotides present between the Not I site  and the
                      oligo-dT track which allows identification of the library
                      of  origin of a clone within the mixture.  The subtracted
                      library (UI-R-E1) was constructed as follows: PCR
                      amplified cDNA inserts from a pool of  UI-R-E0 clones from
                      which 3' ESTs had been derived was used as a driver in a
                      hybridization with the UI-R-E0 library in the form of
                      single-stranded circles. The remaining single-stranded
                      circles (subtracted library) was purified by
                      hydroxyapatite column chromatography, converted to
                      double-stranded circles and electroporated into DH10B
                      bacteria (Life Technologies) to generate the UI-R-E1
                      library. This procedure has been previously described
                      (Bonaldo, Lennon and Soares, Genome Research 6: 791-806,
                      1996)"
                      /db_xref="taxon:10116"
                      /clone="UI-R-E1-fz-g-04-0-UI"
                      /clone_lib="UI-R-E1"
                      /dev_stage="adult"
                      /lab_host="DH10B (Life Technologies)"
BASE COUNT      122 a      56 c      65 g      70 t
ORIGIN

  Query Match           9.8%;  Score 17;  DB 14;  Length 313;
  Best Local Similarity 100.0%;  Pred. No. 6.68e-02;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     64 AAAATAAAACAAGTGAA 80
          |||||||||||||||||
Qy    145 AAAATAAAACAAGTGAA 161


RESULT   46
LOCUS       T93501        317 bp    mRNA            EST       23-MAR-1995
DEFINITION  ye10b12.r1 Homo sapiens cDNA clone 117311 5' similar to contains
            Alu repetitive element;.
ACCESSION   T93501
NID         g726674
KEYWORDS    EST.
```



```
SOURCE      human clone=117311 library=Stratagene lung (#937210)
            vector=pBluescript SK- host=SOLR cells (kanamycin resistant)
            primer=M13RP1 Rsite1=EcoRI Rsite2=XhoI Normal lung tissue from a 72
            year old male. Cloned unidirectionally. Primer: Oligo dT. Average
            insert size: 1.0 kb; Uni-ZAP XR Vector; 5' adaptor sequence:
            5'-GAATTCGGCACGAG-3'; 3' adaptor sequence:
            5'-CTCGAGTTTTTTTTTTTTTTTTT-3'.
  ORGANISM  Homo sapiens
            Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 317)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
            Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 202
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES              Location/Qualifiers
     source           1. .317
                      /organism="Homo sapiens"
                      /clone="117311"
BASE COUNT       67 a      71 c      66 g     110 t       3 others
ORIGIN

  Query Match           9.8%;  Score 17;  DB 5;  Length 317;
  Best Local Similarity 100.0%;  Pred. No. 6.68e-02;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    245 TTGTTTTATTTTTATCT 261
          |||||||||||||||||
Cp    156 TTGTTTTATTTTTATCT 140


RESULT   47
LOCUS       AI011239      318 bp    mRNA            EST       15-JUN-1998
DEFINITION  EST205690 Normalized rat ovary, Bento Soares Rattus sp. cDNA clone
            ROVAS28 3' end, mRNA sequence.
ACCESSION   AI011239
NID         g3225071
```



```
KEYWORDS    EST.
SOURCE      Rattus sp.
  ORGANISM  Rattus sp.
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 318)
  AUTHORS   Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
            Kerlavage,A.R. and Adams,M.D.
  TITLE     Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
            Gene Index
  JOURNAL   Unpublished (1998)
COMMENT
            Contact: Lee, NH
            ATCC
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
            Seq primer: M13-21.
FEATURES              Location/Qualifiers
     source           1. .318
                      /organism="Rattus sp."
                      /note="Organ: ovary; Vector: pT7T3Pac; Site_1: EcoRI;
                      Site_2: NotI"
                      /db_xref="taxon:10118"
                      /clone="ROVAS28"
                      /clone_lib="Normalized rat ovary, Bento Soares"
BASE COUNT      122 a      61 c      69 g      66 t
ORIGIN

  Query Match           9.8%;  Score 17;  DB 15;  Length 318;
  Best Local Similarity 100.0%;  Pred. No. 6.68e-02;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     41 AAAATAAAACAAGTGAA 57
          |||||||||||||||||
Qy    145 AAAATAAAACAAGTGAA 161


RESULT   48
LOCUS       T92492        319 bp    mRNA            EST       22-MAR-1995
DEFINITION  ye21h11.r1 Homo sapiens cDNA clone 118437 5' similar to contains
            Alu repetitive element;.
ACCESSION   T92492
NID         g724405
KEYWORDS    EST.
SOURCE      human clone=118437 library=Stratagene lung (#937210)
            vector=pBluescript SK- host=SOLR cells (kanamycin resistant)
            primer=M13RP1 Rsite1=EcoRI Rsite2=XhoI Normal lung tissue from a 72
            year old male. Cloned unidirectionally. Primer: Oligo dT. Average
```



```
            insert size: 1.0 kb; Uni-ZAP XR Vector; 5' adaptor sequence:
            5'-GAATTCGGCACGAG-3'; 3' adaptor sequence:
            5'-CTCGAGTTTTTTTTTTTTTTTTT-3'.
  ORGANISM  Homo sapiens
            Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 319)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
            Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 161
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES              Location/Qualifiers
     source           1. .319
                      /organism="Homo sapiens"
                      /clone="118437"
BASE COUNT       72 a      52 c      69 g     121 t       5 others
ORIGIN

  Query Match           9.8%;  Score 17;  DB 5;  Length 319;
  Best Local Similarity 100.0%;  Pred. No. 6.68e-02;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    187 TTGTTTTATTTTTATCT 203
          |||||||||||||||||
Cp    156 TTGTTTTATTTTTATCT 140


RESULT   49
LOCUS       AA956559      323 bp    mRNA            EST       07-MAY-1998
DEFINITION  UI-R-E1-fk-a-03-0-UI.s1 UI-R-E1 Rattus norvegicus cDNA clone
            UI-R-E1-fk-a-03-0-UI 3' similar to gi|3005586|gb|AF048977|AF048977
            Homo sapiens Ser/Arg-related nuclear matrix protein (SRM160) mRNA,
            complete cds, mRNA sequence.
ACCESSION   AA956559
NID         g3120254
KEYWORDS    EST.
SOURCE      Norway rat.
```

```
   ORGANISM  Rattus norvegicus
             Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
             Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
  REFERENCE  1  (bases 1 to 323)
    AUTHORS  Bonaldo,M.F., Lennon,G. and Soares,M.B.
      TITLE  Normalization and subtraction: two approaches to facilitate gene
             discovery
    JOURNAL  Genome Res. 6 (9), 791-806 (1996)
     MEDLINE 97044477
  COMMENT
             Contact: Soares, MB
             Program for Rat Gene Discovery and Mapping
             University of Iowa
             451 Eckstein Medical Research Building Iowa City, IA 52242, USA
             Tel: 319 335 8250
             Fax: 319 335 9565
             Email: msoares@blue.weeg.uiowa.edu
             The sequence tag present in the cDNA between the NotI site and the
             oligo-dT track (not shown) served to identify it as a clone from
             the normalized adult 18-Day-Embryo library. cDNA Library
             Preparation: M. Fatima Bonaldo, Ph.D. Clone distribution: clones
             will be available through Research Genetics
             Seq primer: M13 Forward.
  FEATURES             Location/Qualifiers
       source          1..323
                       /organism="Rattus norvegicus"
                       /strain="Sprague-Dawley"
                       /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                       polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-E1
                       library is a subtracted library derived from the UI-R-E0
                       library.  The UI-R-E0 library  consisted of a mixture of
                       individually  tagged normalized libraries constructed from
                       8, 12 and 18-day embryo.   The tag is a string of 3-5
                       nucleotides present between the Not I site  and the
                       oligo-dT track which allows identification of the library
                       of  origin of a clone within the mixture.  The subtracted
                       library (UI-R-E1) was constructed as follows: PCR
                       amplified cDNA inserts from a pool of  UI-R-E0 clones from
                       which 3' ESTs had been derived was used as a driver in a
                       hybridization with the UI-R-E0 library in the form of
                       single-stranded circles. The remaining single-stranded
                       circles (subtracted library) was purified by
                       hydroxyapatite column chromatography, converted to
                       double-stranded circles and electroporated into DH10B
                       bacteria (Life Technologies) to generate the UI-R-E1
                       library. This procedure has been previously described
                       (Bonaldo, Lennon and Soares, Genome Research 6: 791-806,
                       1996)"
                       /db_xref="taxon:10116"
                       /clone="UI-R-E1-fk-a-03-0-UI"
                       /clone_lib="UI-R-E1"
```



```
                       /dev_stage="adult"
                       /lab_host="DH10B (Life Technologies)"
  BASE COUNT       124 a      58 c      66 g      74 t       1 others
  ORIGIN

    Query Match             9.8%;  Score 17;  DB 14;  Length 323;
    Best Local Similarity 100.0%;  Pred. No. 6.68e-02;
    Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

  Db      64 AAAATAAAACAAGTGAA 80
             |||||||||||||||||
  Qy     145 AAAATAAAACAAGTGAA 161


  RESULT   50
  LOCUS       R80040         325 bp    mRNA             EST       09-JUN-1995
  DEFINITION  yi91d06.r1 Homo sapiens cDNA clone 146603 5' similar to gb:D14043
              PUTATIVE MUCIN CORE PROTEIN PRECURSOR 24 (HUMAN);.
  ACCESSION   R80040
  NID         g856321
  KEYWORDS    EST.
  SOURCE      human clone=146603 library=Soares placenta Nb2HP vector=pT7T3D
              (Pharmacia) with a modified polylinker host=DH10B (ampicillin
              resistant) primer=M13RP1 Rsite1=Not I Rsite2=Eco RI Female placenta
              obtained at birth (full term). 1st strand cDNA was primed with a
              Not I - oligo(dT) primer [5'
              AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTTT 3'], double-stranded
              cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not
              I and cloned into the Not I and Eco RI sites of the modified pT7T3
              vector. Library went through one round of normalization. Library
              constructed by Bento Soares and M.Fatima Bonaldo.
    ORGANISM  Homo sapiens
              Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
              Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
              Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
              Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE   1  (bases 1 to 325)
    AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
              Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
              Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
              Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
              Wilson,R.
      TITLE   The WashU-Merck EST Project
    JOURNAL   Unpublished (1995)
  COMMENT
              Contact: Wilson RK
              WashU-Merck EST Project
              Washington University School of Medicine
              4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
              Tel: 314 286 1800
              Fax: 314 286 1810
```



```
              Email: est@watson.wustl.edu
              High quality sequence stops: 258
              Source: IMAGE Consortium, LLNL
              This clone is available royalty-free through LLNL ; contact the
              IMAGE Consortium (info@image.llnl.gov) for further information.
  FEATURES             Location/Qualifiers
       source          1..325
                       /organism="Homo sapiens"
                       /clone="146603"
  BASE COUNT       115 a      41 c      67 g      102 t
  ORIGIN

    Query Match             9.8%;  Score 17;  DB 16;  Length 325;
    Best Local Similarity 100.0%;  Pred. No. 6.68e-02;
    Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

  Db     152 TACATTAGAAATAAGAT 168
             |||||||||||||||||
  Cp      99 TACATTAGAAATAAGAT 83


  Search completed: Tue Dec 22 20:47:27 1998
  Job time : 381 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Tue Dec 22 20:47:48 1998; MasPar time 39.54 Seconds
598.707 Million cell updates/sec

Tabular output not generated.

Title: >US-08-530-112A-1219
Description: (1-174) from US08530112A.seq
Perfect Score: 174
N.A. Sequence: 1 GATCTGAGCAGGCAAAGCTC.....AAGTGAATTTGCAAGTGAAA 174
Comp: CTAGACTCGTCCGTTTCGAG.....TTCACTTAAACGTTCACTTT

Scoring table: TABLE jmetric
Gap 60

Nmatch STD : Dbase 0; Query 0

Searched: 188442 seqs, 58026449 bases x 2

Post-processing: Minimum Match 0%
Listing first 1000 summaries

Database: n-geneseq32
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39 40:part40

Statistics: Mean 6.322; Variance 3.119; scale 2.027

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
	1	174	100.0	174	19	T20219	Human gene signature	1.11e-126
	2	20	11.5	20	23	T41371	Human gene signature	4.30e-02
c	3	20	11.5	20	23	T41372	Human gene signature	4.30e-02
c	4	17	9.8	168	22	T24247	Human gene signature	2.95e+00
c	5	17	9.8	702	1	N90852	Intron 3 from human f	2.95e+00
c	6	17	9.8	2207	1	N80748	Sequence of a polygal	2.95e+00
	7	17	9.8	3957	18	T14077	Human dihydropyrimidi	2.95e+00
	8	17	9.8	4368	18	T03133	Human dihydropyrimidi	2.95e+00
	9	17	9.8	4369	18	T03143	Human dihydropyrimidi	2.95e+00
c	10	16	9.2	609	2	N70099	Sequence of DNA fragm	1.13e+01
c	11	16	9.2	668	3	N40108	Sequence of the cDNA	1.13e+01
c	12	16	9.2	668	4	N30101	Sequence on recombina	1.13e+01
c	13	16	9.2	761	4	N20031	DNA sequence encoding	1.13e+01
c	14	16	9.2	777	4	N10009	Sequence encoding fib	1.13e+01
c	15	16	9.2	777	11	Q68695	Human fibroblast inte	1.13e+01
c	16	16	9.2	836	4	N10041	Sequence encoding fib	1.13e+01
c	17	16	9.2	899	4	N10040	Sequence encoding fib	1.13e+01
c	18	16	9.2	926	4	N10038	Sequence encoding hum	1.13e+01
	19	16	9.2	1542	3	Q20695	Mouse OSF.1.	1.13e+01
c	20	16	9.2	1584	24	T35525	Human interferon-beta	1.13e+01
c	21	16	9.2	1836	4	N20057	Interferon-beta gene.	1.13e+01
c	22	16	9.2	2169	33	T79804	Potato starch branchi	1.13e+01
c	23	16	9.2	2169	33	T79783	Potato starch branchi	1.13e+01
c	24	16	9.2	2214	24	T34612	Probe for detecting p	1.13e+01
c	25	16	9.2	11478	33	T79784	Full length potato st	1.13e+01
c	26	16	9.2	11478	33	T79805	Full length potato st	1.13e+01
c	27	16	9.2	580073	27	T58840	Mycoplasma genitalium	1.13e+01
c	28	15	8.6	119	22	T24800	Human gene signature	4.13e+01
c	29	15	8.6	129	19	T21758	Human gene signature	4.13e+01
c	30	15	8.6	189	21	T24427	Human gene signature	4.13e+01
	31	15	8.6	303	8	Q60229	Human brain Expressed	4.13e+01
	32	15	8.6	304	8	Q60068	Human brain Expressed	4.13e+01
	33	15	8.6	312	18	T20989	Human gene signature	4.13e+01
c	34	15	8.6	337	19	T19987	Human gene signature	4.13e+01
	35	15	8.6	359	21	T22621	Human gene signature	4.13e+01
c	36	15	8.6	390	8	Q60080	Human brain Expressed	4.13e+01
	37	15	8.6	405	30	T67805	H. pylori cytoplasmic	4.13e+01
	38	15	8.6	405	30	T77485	H. pylori cytoplasmic	4.13e+01
	39	15	8.6	486	30	T68020	H. pylori cytoplasmic	4.13e+01
c	40	15	8.6	927	30	T66501	Streptococcal inhibit	4.13e+01
c	41	15	8.6	1204	30	T13864	Streptococcal inhibit	4.13e+01
	42	15	8.6	1637	29	T45357	Human ceramide glucos	4.13e+01

c	43	15	8.6	1784	1	Q04487	Sequence of protein H	4.13e+01
c	44	15	8.6	1828	34	T80109	Mouse KF-1 gene fragm	4.13e+01
c	45	15	8.6	1871	34	T73869	Cotton lipid transfer	4.13e+01
	46	15	8.6	2592	33	T62651	DNA encoding type I c	4.13e+01
	47	15	8.6	2604	3	Q14945	Rhizopus PGK2 promote	4.13e+01
c	48	15	8.6	3049	12	Q70890	Tomato spotted wilt v	4.13e+01
c	49	15	8.6	3095	3	Q20292	Sequence of plasmid p	4.13e+01
c	50	15	8.6	3261	34	T80100	Mouse KF-1 gene.	4.13e+01
	51	15	8.6	3280	2	N82053	DNA encoding fibronec	4.13e+01
c	52	15	8.6	3347	1	N80639	Sequence of cDNA enco	4.13e+01
c	53	15	8.6	3856	1	Q03097	Enod2b genomic clone.	4.13e+01
c	54	15	8.6	4440	35	T89487	Human A20 protein cDN	4.13e+01
	55	15	8.6	4447	18	T14078	Pig dihydropyrimidine	4.13e+01
	56	15	8.6	4918	10	Q56696	Sequence of the L. la	4.13e+01
c	57	15	8.6	6511	15	Q95493	Human Cdn-2 DNA.	4.13e+01
c	58	15	8.6	8363	16	Q92408	Human cyclin A gene.	4.13e+01
	59	15	8.6	8920	11	Q62924	Carbamoyl-phosphate-s	4.13e+01
c	60	15	8.6	12151	10	Q54676	Rice starch branching	4.13e+01
c	61	15	8.6	12151	11	Q62137	Rice starch branching	4.13e+01
c	62	15	8.6	24593	3	N50226	Sequence of opine syn	4.13e+01
c	63	15	8.6	24596	3	N50182	Complete nucleotide s	4.13e+01
c	64	15	8.6	133894	17	T13635	AcNPV genomic DNA clo	4.13e+01
	65	14	8.0	40	37	T97401	Synthetic forty mer n	1.44e+02
c	66	14	8.0	40	12	Q72310	Adenomatous polyposis	1.44e+02
c	67	14	8.0	40	37	T96167	3' intron sequence of	1.44e+02
c	68	14	8.0	40	36	T95552	3' intron sequence of	1.44e+02
	69	14	8.0	73	20	T23162	Human gene signature	1.44e+02
	70	14	8.0	188	18	T21028	Human gene signature	1.44e+02
c	71	14	8.0	194	34	T80108	Human KF-1 gene 3' en	1.44e+02
c	72	14	8.0	214	22	T24799	Human gene signature	1.44e+02
	73	14	8.0	240	12	Q78640	Potato PI-IIK gene pr	1.44e+02
	74	14	8.0	253	22	T25894	Human gene signature	1.44e+02
c	75	14	8.0	255	40	V24762	H. pylori ORF 01cp117	1.44e+02
c	76	14	8.0	281	8	Q60250	Human brain Expressed	1.44e+02
	77	14	8.0	299	18	T19557	Human gene signature	1.44e+02
	78	14	8.0	305	8	Q59657	Human brain Expressed	1.44e+02
	79	14	8.0	325	8	Q60687	Human brain Expressed	1.44e+02
c	80	14	8.0	327	30	T68037	H. pylori secreted or	1.44e+02
c	81	14	8.0	351	8	Q60975	Human brain Expressed	1.44e+02
c	82	14	8.0	354	6	Q39766	Expressed Sequence Ta	1.44e+02
c	83	14	8.0	354	8	Q59178	Human brain Expressed	1.44e+02
c	84	14	8.0	423	40	V24615	H. pylori ORF 01ep305	1.44e+02
c	85	14	8.0	765	30	T67682	H. pylori secreted or	1.44e+02
c	86	14	8.0	794	1	Q05683	Islets of Langerhans	1.44e+02
	87	14	8.0	798	30	T77475	H. pylori transporter	1.44e+02
	88	14	8.0	798	30	T67795	H. pylori transporter	1.44e+02
c	89	14	8.0	805	14	Q83839	B. burgdorferi 75 gen	1.44e+02
	90	14	8.0	813	30	T68033	H. pylori transporter	1.44e+02
c	91	14	8.0	830	31	T70299	Pea PsMTA gene promot	1.44e+02
	92	14	8.0	871	4	Q23360	Anther-specific promo	1.44e+02
c	93	14	8.0	1002	5	Q29855	Odorant receptor clon	1.44e+02

c	94	14	8.0	1029	10	Q55186	MS-PO23 Vector.	1.44e+02
	95	14	8.0	1052	1	N90224	Malaria-specific DNA	1.44e+02
	96	14	8.0	1095	38	T91595	Lucilia cuprina proho	1.44e+02
	97	14	8.0	1138	24	T41465	gC1q receptor cDNA.	1.44e+02
c	98	14	8.0	1165	2	N60391	Sequence encoding the	1.44e+02
c	99	14	8.0	1215	28	T58339	Tobacco epi-5-aristol	1.44e+02
	100	14	8.0	1241	3	Q20253	Wound-inducible potat	1.44e+02
	101	14	8.0	1254	40	V24788	H. pylori ORF 07cp217	1.44e+02
c	102	14	8.0	1320	40	V17564	Coding sequence for t	1.44e+02
	103	14	8.0	1334	1	N91223	Em gene 5'flanking re	1.44e+02
	104	14	8.0	1395	11	Q66812	AmEPV spheroidin G1L	1.44e+02
	105	14	8.0	1405	40	V00616	Lycopersicon esculent	1.44e+02
c	106	14	8.0	1526	34	T80107	Human KF-1 gene clone	1.44e+02
	107	14	8.0	1589	24	T39518	Flea calreticulin cDN	1.44e+02
c	108	14	8.0	1589	24	T39516	Flea calreticulin cDN	1.44e+02
	109	14	8.0	1657	31	T68767	ATM mutant 497del7514	1.44e+02
c	110	14	8.0	1666	24	T45050	Beta-globin gene for	1.44e+02
	111	14	8.0	1680	1	N80488	Plasmid clone pPE1.	1.44e+02
	112	14	8.0	1680	6	Q38416	Pectin esterase cDNA	1.44e+02
c	113	14	8.0	1686	4	N10014	Sequence encoding int	1.44e+02
c	114	14	8.0	1704	1	N90725	Groundnut stilbene sy	1.44e+02
	115	14	8.0	1723	4	N40246	DNA segment from pig	1.44e+02
c	116	14	8.0	1859	36	T96468	MSRV-1 clone FBd3.	1.44e+02
	117	14	8.0	1873	39	V15103	Human vascular endoth	1.44e+02
c	118	14	8.0	1977	23	T36794	Prostate-specific mem	1.44e+02
	119	14	8.0	2032	18	T13381	Pea ADPG-PPase SH2 su	1.44e+02
c	120	14	8.0	2038	1	N90367	Fragment of tobacco P	1.44e+02
c	121	14	8.0	2038	1	Q03663	Fragment of tobacco P	1.44e+02
	122	14	8.0	2059	7	Q44295	Pectin esterase clone	1.44e+02
c	123	14	8.0	2080	33	T80098	KF-1 gene.	1.44e+02
	124	14	8.0	2196	19	T04075	M.gallisepticum DNA s	1.44e+02
	125	14	8.0	2316	3	Q14052	KpnI restriction endo	1.44e+02
c	126	14	8.0	2316	3	Q14779	KpnI restriction endo	1.44e+02
	127	14	8.0	2342	3	Q14955	Rat dopamine D1 recep	1.44e+02
	128	14	8.0	2369	12	Q77853	Mycoplasma gallisepti	1.44e+02
c	129	14	8.0	2441	14	Q55144	Enterococcus faecalis	1.44e+02
c	130	14	8.0	2500	21	T30565	Eukaryotic cell growt	1.44e+02
c	131	14	8.0	2550	33	T85402	Human cadherin-8 codi	1.44e+02
c	132	14	8.0	2550	28	T61923	Full length human cad	1.44e+02
	133	14	8.0	2606	32	T72895	Plasmodium E31a gene.	1.44e+02
	134	14	8.0	2606	14	Q83527	P. falciparum E31a ge	1.44e+02
	135	14	8.0	2617	5	Q29753	DNA sequence encoding	1.44e+02
c	136	14	8.0	2636	32	T74133	Zea mays root prefere	1.44e+02
c	137	14	8.0	2956	37	T99140	Full length cytochrom	1.44e+02
	138	14	8.0	3014	3	Q14954	Human dopamine D1 rec	1.44e+02
c	139	14	8.0	3043	28	T61918	Truncated rat cadheri	1.44e+02
c	140	14	8.0	3043	33	T85407	Rat truncated cadheri	1.44e+02
c	141	14	8.0	3056	15	Q90662	Eph-related PTK Cek7'	1.44e+02
c	142	14	8.0	3063	1	N91438	DNA encoding glycine	1.44e+02
c	143	14	8.0	3136	33	T85406	Rat full length cadhe	1.44e+02
c	144	14	8.0	3136	28	T61917	Full length rat cadhe	1.44e+02

	145	14	8.0	3138	2	Q11712	Shuttle vector pMUW16 1.44e+02
c	146	14	8.0	3171	18	T18309	BPGP-1 gene. 1.44e+02
c	147	14	8.0	3182	7	Q42590	Human LIF-R N-termina 1.44e+02
c	148	14	8.0	3182	4	Q25799	hLIF-R. 1.44e+02
c	149	14	8.0	3182	16	Q92270	Human leukaemia inhib 1.44e+02
c	150	14	8.0	3182	10	Q55940	Human leukaemia inhib 1.44e+02
	151	14	8.0	3300	29	T41617	Adzuki bean endo-xylo 1.44e+02
	152	14	8.0	3407	20	T12168	pJG4-5-CDK-BP full le 1.44e+02
c	153	14	8.0	3423	34	T80099	Human KP-1 gene. 1.44e+02
c	154	14	8.0	3526	34	T80106	Human KP-1 gene clone 1.44e+02
	155	14	8.0	3565	14	Q84051	Sequence encoding mel 1.44e+02
c	156	14	8.0	3591	16	Q92272	Human leukaemia inhib 1.44e+02
c	157	14	8.0	3591	10	Q55942	Human leukaemia inhib 1.44e+02
c	158	14	8.0	3591	9	Q58427	Human LIF-R. 1.44e+02
	159	14	8.0	3622	1	N90567	Ovine Y-chromosomal D 1.44e+02
	160	14	8.0	3641	1	Q05752	Microspore-specific c 1.44e+02
	161	14	8.0	3641	28	T48866	Brassica napus micros 1.44e+02
	162	14	8.0	3730	18	T03195	Group B Streptococcal 1.44e+02
	163	14	8.0	3852	30	T62419	Bovine beta-mannosida 1.44e+02
c	164	14	8.0	3865	38	T96833	Intron 3 of human pRb 1.44e+02
	165	14	8.0	3876	22	T31033	Bovine Factor-XIII cD 1.44e+02
	166	14	8.0	4041	20	T13944	VIP2A(a) and VIP1A(a) 1.44e+02
	167	14	8.0	4041	39	V16172	DNA sequence encoding 1.44e+02
	168	14	8.0	4041	38	T73999	B. cereus VIP1A(a)/VI 1.44e+02
	169	14	8.0	4074	38	T73998	B. thuringiensis VIP2 1.44e+02
	170	14	8.0	4074	39	V16171	DNA sequence containi 1.44e+02
	171	14	8.0	4074	20	T13950	B. thuringiensis VIP2 1.44e+02
c	172	14	8.0	4144	6	Q37188	Mouse FosB coding seq 1.44e+02
	173	14	8.0	4200	18	T03190	Group B Streptococcal 1.44e+02
c	174	14	8.0	4253	28	T58361	Tobacco epi-5-aristol 1.44e+02
	175	14	8.0	4300	33	T80042	Pectinase coding sequ 1.44e+02
	176	14	8.0	4319	18	T17215	Adhesion and penetrat 1.44e+02
c	177	14	8.0	4359	9	Q52447	Canine coronavirus 1- 1.44e+02
c	178	14	8.0	4359	9	Q50617	Canine coronavirus 1- 1.44e+02
	179	14	8.0	4364	12	Q71334	Human developmental t 1.44e+02
c	180	14	8.0	4365	9	Q52444	Feline enteric cornon 1.44e+02
c	181	14	8.0	4365	4	Q25539	S gene of FECV. 1.44e+02
c	182	14	8.0	4429	5	Q29685	CCVInsavc-1 spike gen 1.44e+02
c	183	14	8.0	4435	5	Q29686	CCV-C54 spike gene. 1.44e+02
c	184	14	8.0	4486	6	Q35118	Encodes E2 protein of 1.44e+02
c	185	14	8.0	4500	5	Q29684	CCV-6 spike gene. 1.44e+02
c	186	14	8.0	4656	1	N80747	Sequence of 2A11 geno 1.44e+02
c	187	14	8.0	4673	20	T12787	Pea phytochrome gene. 1.44e+02
	188	14	8.0	4788	14	Q85367	MIP-1-alpha. 1.44e+02
c	189	14	8.0	4788	14	Q85367	MIP-1-alpha. 1.44e+02
	190	14	8.0	4937	13	Q72294	Haemophilus high mole 1.44e+02
	191	14	8.0	4937	9	Q49507	Sequence encoding hig 1.44e+02
	192	14	8.0	4937	37	T90995	Non-typeable Haemophi 1.44e+02
	193	14	8.0	5116	13	Q72293	Haemophilus high mole 1.44e+02
	194	14	8.0	5116	9	Q49506	Sequence encoding hig 1.44e+02
	195	14	8.0	5116	37	T90994	Non-typeable Haemophi 1.44e+02

c	196	14	8.0	5244	28	T48869	IamH expression vecto 1.44e+02
	197	14	8.0	5309	2	Q10280	Complete sequence of 1.44e+02
	198	14	8.0	5476	29	T60078	Human SMCY cDNA. 1.44e+02
	199	14	8.0	5648	1	Q03368	Sequence of human lya 1.44e+02
	200	14	8.0	6049	20	T13939	B. cereus VIP2A(a) ge 1.44e+02
	201	14	8.0	6049	39	V16165	DNA encoding vegetati 1.44e+02
	202	14	8.0	6049	38	T73993	B. cereus VIP2A(a) co 1.44e+02
	203	14	8.0	6060	28	T48454	Human beta-globin gen 1.44e+02
	204	14	8.0	6106	13	Q74678	Bacillus cereus vegat 1.44e+02
c	205	14	8.0	6112	28	T48868	IamS expression vecto 1.44e+02
	206	14	8.0	6124	1	Q03568	Sequence encoding the 1.44e+02
	207	14	8.0	6360	2	Q11378	Genes encoding T. hyo 1.44e+02
	208	14	8.0	6766	39	V14507	AmEPV entomopoxvirus 1.44e+02
	209	14	8.0	6768	39	V14517	AmEPV entomopoxvirus 1.44e+02
	210	14	8.0	6768	5	Q28301	AmEPV Spheroidin DNA. 1.44e+02
	211	14	8.0	8341	31	T68772	ATM mutant 4777del830 1.44e+02
	212	14	8.0	8457	11	Q66797	AmEPV spheroidin gene 1.44e+02
	213	14	8.0	8767	31	T68735	ATM mutant 8269del403 1.44e+02
	214	14	8.0	8768	37	T73812	ATM gene variant 8269 1.44e+02
	215	14	8.0	8799	31	T68758	ATM mutant 2467del372 1.44e+02
	216	14	8.0	8816	31	T68750	ATM mutant 5320del355 1.44e+02
	217	14	8.0	8873	31	T68740	ATM mutant 7630del298 1.44e+02
	218	14	8.0	8959	14	Q55133	Staphylococcus aureus 1.44e+02
	219	14	8.0	8970	31	T68771	ATM mutant 1407del201 1.44e+02
	220	14	8.0	8996	31	T68783	ATM mutant 4443del175 1.44e+02
	221	14	8.0	8996	31	T68754	ATM mutant 4437del175 1.44e+02
	222	14	8.0	8997	31	T68774	ATM mutant 3403del174 1.44e+02
	223	14	8.0	8997	31	T68756	ATM mutant 3403del174 1.44e+02
	224	14	8.0	9006	31	T68775	ATM mutant 4612del165 1.44e+02
	225	14	8.0	9006	31	T68753	ATM mutant 4612del165 1.44e+02
	226	14	8.0	9012	31	T68741	ATM mutant 7630del159 1.44e+02
	227	14	8.0	9018	31	T68736	ATM mutant 8269del153 1.44e+02
	228	14	8.0	9022	31	T68745	ATM mutant 6199del149 1.44e+02
	229	14	8.0	9029	31	T68752	ATM mutant 5178del142 1.44e+02
	230	14	8.0	9032	31	T68739	ATM mutant 7789del139 1.44e+02
	231	14	8.0	9044	31	T68755	ATM mutant 4110del127 1.44e+02
	232	14	8.0	9046	31	T68770	ATM mutant 2125del125 1.44e+02
	233	14	8.0	9046	31	T68761	ATM mutant 2125del125 1.44e+02
	234	14	8.0	9066	31	T68744	ATM mutant 6348del105 1.44e+02
	235	14	8.0	9081	31	T68759	ATM mutant 2377del190 1.44e+02
	236	14	8.0	9088	31	T68757	ATM mutant 2839del83 1.44e+02
	237	14	8.0	9160	31	T68749	ATM mutant 5539del11 1.44e+02
	238	14	8.0	9162	31	T68777	ATM mutant 7636del9 c 1.44e+02
	239	14	8.0	9164	31	T68751	ATM mutant 5320del7 c 1.44e+02
	240	14	8.0	9165	31	T68785	ATM mutant 7278del6 c 1.44e+02
	241	14	8.0	9165	31	T68778	ATM mutant 7279del6 c 1.44e+02
	242	14	8.0	9166	31	T68738	ATM mutant 7883del5 c 1.44e+02
	243	14	8.0	9166	31	T68743	ATM mutant 6573del5 c 1.44e+02
	244	14	8.0	9166	31	T68746	ATM mutant 5979del5 c 1.44e+02
	245	14	8.0	9167	31	T68742	ATM mutant 7517del4 c 1.44e+02
	246	14	8.0	9167	31	T68769	ATM mutant 7240del4 c 1.44e+02

	247	14	8.0	9168	31	T68776	ATM mutant 8578del3 c 1.44e+02
	248	14	8.0	9168	31	T68780	ATM mutant 5435del3 c 1.44e+02
	249	14	8.0	9169	31	T68731	ATM mutant 9001delAG 1.44e+02
	250	14	8.0	9169	37	T73808	ATM gene variant 9001 1.44e+02
	251	14	8.0	9169	31	T68760	ATM mutant 2284delCT 1.44e+02
	252	14	8.0	9169	37	T73811	ATM gene variant 8283 1.44e+02
	253	14	8.0	9169	31	T68734	ATM mutant 8283delTC 1.44e+02
	254	14	8.0	9169	31	T68766	ATM mutant 755delGT c 1.44e+02
	255	14	8.0	9169	31	T68763	ATM mutant 1563delAG 1.44e+02
	256	14	8.0	9170	31	T68762	ATM mutant 2113delT c 1.44e+02
	257	14	8.0	9171	31	T68773	ATM mutant C9139T cod 1.44e+02
	258	14	8.0	9171	31	T68768	ATM mutant A538T codi 1.44e+02
	259	14	8.0	9171	31	T68733	ATM mutant G8307A cod 1.44e+02
	260	14	8.0	9171	37	T73810	ATM gene variant G830 1.44e+02
	261	14	8.0	9171	37	T90996	Non-typeable Haemophi 1.44e+02
	262	14	8.0	9171	31	T68765	ATM mutant C1240T cod 1.44e+02
	263	14	8.0	9171	37	T45419	Ataxia-telangiectasia 1.44e+02
	264	14	8.0	9171	37	T43497	ATM open reading fram 1.44e+02
	265	14	8.0	9171	31	T68781	ATM mutant A8711G cod 1.44e+02
	266	14	8.0	9171	31	T68737	ATM mutant C8140T cod 1.44e+02
	267	14	8.0	9171	31	T68782	ATM mutant T128C codi 1.44e+02
	268	14	8.0	9171	31	T68764	ATM mutant C1339T cod 1.44e+02
	269	14	8.0	9172	31	T68748	ATM mutant 5554insC c 1.44e+02
	270	14	8.0	9172	31	T68732	ATM mutant 8946insA c 1.44e+02
	271	14	8.0	9172	31	T68747	ATM mutant 5712insA c 1.44e+02
	272	14	8.0	9180	31	T68779	ATM mutant 5319ins9 c 1.44e+02
	273	14	8.0	9220	9	Q49508	Gene cluster for high 1.44e+02
	274	14	8.0	9323	37	T90997	Non-typeable Haemophi 1.44e+02
	275	14	8.0	9323	9	Q49509	Gene cluster for high 1.44e+02
	276	14	8.0	9620	37	T45424	Murine Ataxia-telangi 1.44e+02
	277	14	8.0	9636	12	Q67190	P. falciparum transmi 1.44e+02
	278	14	8.0	10207	14	Q55134	Staphylococcus aureus 1.44e+02
	279	14	8.0	10266	22	T33007	Mouse SRY-related gen 1.44e+02
c	280	14	8.0	10266	22	T33007	Mouse SRY-related gen 1.44e+02
	281	14	8.0	12778	31	T68784	ATM mutant G9170C cod 1.44e+02
	282	14	8.0	14704	3	Q20685	PKS 741 insert contg. 1.44e+02
c	283	14	8.0	14704	3	Q20685	PKS 741 insert contg. 1.44e+02
	284	14	8.0	37895	24	T34137	Gene for biosynthetic 1.44e+02
	285	14	8.0	37913	10	Q51128	Eicosapentaenoic acid 1.44e+02
	286	14	8.0	40875	33	T80043	Insert from cosmid 10 1.44e+02
	287	14	8.0	133894	17	T13635	AcNPV genomic DNA clo 1.44e+02
	288	14	8.0	580073	27	T58840	Mycoplasma genitalium 1.44e+02
c	289	13	7.5	20	37	T97419	Donor Sequence oligom 4.78e+02
c	290	13	7.5	20	37	T97431	Oligomer s14102 Commo 4.78e+02
c	291	13	7.5	22	36	T97861	PCR primer 10 for DNA 4.78e+02
c	292	13	7.5	24	35	T92483	BRCA2 cancer suscepti 4.78e+02
c	293	13	7.5	29	14	Q81465	PCR primer M3R. 4.78e+02
	294	13	7.5	35	37	T91867	Primer for human insu 4.78e+02
c	295	13	7.5	39	24	T45192	Mycoplasma genitalium 4.78e+02
	296	13	7.5	40	37	T97399	Synthetic forty mer n 4.78e+02
	297	13	7.5	40	37	T97402	Synthetic forty mer n 4.78e+02

	298	13	7.5	40	37	T97400	Synthetic forty mer n 4.78e+02
	299	13	7.5	40	37	T97429	Oligomer s14102 Allel 4.78e+02
	300	13	7.5	40	37	T97430	Oligomer s14102 Allel 4.78e+02
c	301	13	7.5	40	37	T96164	5' intron sequence of 4.78e+02
c	302	13	7.5	40	36	T95549	5' intron sequence of 4.78e+02
c	303	13	7.5	40	12	Q72307	Adenomatous polyposis 4.78e+02
c	304	13	7.5	67	20	T22903	Human gene signature 4.78e+02
c	305	13	7.5	87	20	T23766	Human gene signature 4.78e+02
	306	13	7.5	94	18	T16068	Murine Ikaros gene ex 4.78e+02
	307	13	7.5	94	10	Q61551	Exon boundary sequenc 4.78e+02
	308	13	7.5	99	20	T22374	Human gene signature 4.78e+02
c	309	13	7.5	103	20	T23636	Human gene signature 4.78e+02
	310	13	7.5	105	19	T20377	Human gene signature 4.78e+02
	311	13	7.5	114	8	Q48669	BDA-generated region 4.78e+02
	312	13	7.5	127	20	T22034	Human gene signature 4.78e+02
c	313	13	7.5	133	21	T22550	Human gene signature 4.78e+02
c	314	13	7.5	153	20	T21475	Human gene signature 4.78e+02
c	315	13	7.5	158	18	T19327	Human gene signature 4.78e+02
	316	13	7.5	160	18	T19695	Human gene signature 4.78e+02
	317	13	7.5	180	10	Q76476	Human genome fragment 4.78e+02
c	318	13	7.5	182	18	T19283	Human gene signature 4.78e+02
	319	13	7.5	188	33	T80631	Type II topoisomerase 4.78e+02
	320	13	7.5	192	18	T19663	Human gene signature 4.78e+02
	321	13	7.5	193	20	T21709	Human gene signature 4.78e+02
	322	13	7.5	197	5	Q30027	PCR product prepared 4.78e+02
	323	13	7.5	198	10	Q76543	Human genome fragment 4.78e+02
c	324	13	7.5	210	10	Q76912	Human genome fragment 4.78e+02
c	325	13	7.5	215	19	T20150	Human gene signature 4.78e+02
	326	13	7.5	219	30	T67425	H. pylori membrane pr 4.78e+02
	327	13	7.5	227	19	T20851	Human gene signature 4.78e+02
	328	13	7.5	229	8	Q59375	Human brain Expressed 4.78e+02
	329	13	7.5	231	10	Q57438	Ribosome protein LS 1 4.78e+02
c	330	13	7.5	235	10	Q76644	Human genome fragment 4.78e+02
c	331	13	7.5	259	1	Q06278	Sequence encoding Dow 4.78e+02
	332	13	7.5	262	1	N92948	Latter portion of int 4.78e+02
c	333	13	7.5	275	8	Q61282	Human brain Expressed 4.78e+02
	334	13	7.5	277	18	T19185	Human gene signature 4.78e+02
	335	13	7.5	280	21	T11799	Mycobacterium strain 4.78e+02
c	336	13	7.5	280	40	T99590	Secreted protein AS15 4.78e+02
	337	13	7.5	281	21	T11815	M. scrofulaceum ITG 4 4.78e+02
	338	13	7.5	284	10	Q76600	Human genome fragment 4.78e+02
c	339	13	7.5	286	8	Q59978	Human brain Expressed 4.78e+02
	340	13	7.5	292	18	T13384	ADPG-PPase SH2 subuni 4.78e+02
c	341	13	7.5	294	35	T75603	Canine herpesvirus ge 4.78e+02
c	342	13	7.5	297	20	T23506	Human gene signature 4.78e+02
c	343	13	7.5	297	8	Q60810	Human brain Expressed 4.78e+02
	344	13	7.5	298	19	T09773	Tomato genomic DNA, D 4.78e+02
	345	13	7.5	298	22	T25206	Human gene signature 4.78e+02
c	346	13	7.5	312	22	T26082	Human gene signature 4.78e+02
c	347	13	7.5	317	22	T25118	Human gene signature 4.78e+02
	348	13	7.5	324	20	T22907	Human gene signature 4.78e+02

	349	13	7.5	328	18	T19547	Human gene signature 4.78e+02
	350	13	7.5	332	22	T25321	Human gene signature 4.78e+02
	351	13	7.5	335	8	Q60081	Human brain Expressed 4.78e+02
	352	13	7.5	340	25	T27982	Hepatitis C virus typ 4.78e+02
	353	13	7.5	347	21	T25407	Human gene signature 4.78e+02
c	354	13	7.5	348	22	T24240	Human gene signature 4.78e+02
c	355	13	7.5	350	21	T25419	Human gene signature 4.78e+02
c	356	13	7.5	351	19	T20540	Human gene signature 4.78e+02
c	357	13	7.5	356	20	T21888	Human gene signature 4.78e+02
c	358	13	7.5	371	40	V04475	Secreted protein AS15 4.78e+02
c	359	13	7.5	375	8	Q60808	Human brain Expressed 4.78e+02
	360	13	7.5	387	8	Q59758	Human brain Expressed 4.78e+02
	361	13	7.5	391	19	T20524	Human gene signature 4.78e+02
	362	13	7.5	392	8	Q61361	Human brain Expressed 4.78e+02
	363	13	7.5	415	18	T19071	Human gene signature 4.78e+02
	364	13	7.5	420	19	T19976	Human gene signature 4.78e+02
	365	13	7.5	423	33	T85009	Type II topoisomerase 4.78e+02
c	366	13	7.5	444	1	N91069	Cloned cDNA encoding 4.78e+02
	367	13	7.5	456	8	Q61387	Human brain Expressed 4.78e+02
c	368	13	7.5	458	38	T91259	Human papilloma virus 4.78e+02
	369	13	7.5	516	1	N90685	DNA encoding dihyrofo 4.78e+02
	370	13	7.5	516	1	N90911	Coding sequence of di 4.78e+02
	371	13	7.5	516	30	T67862	H. pylori secreted or 4.78e+02
	372	13	7.5	525	27	T42240	HIV-1 group O strain 4.78e+02
	373	13	7.5	543	5	Q32654	LEMMI 4 cDNA. 4.78e+02
c	374	13	7.5	549	30	T67647	H. pylori cytoplasmic 4.78e+02
	375	13	7.5	552	40	V02986	Human BF245_1 interna 4.78e+02
c	376	13	7.5	552	30	T68258	H. pylori cytoplasmic 4.78e+02
c	377	13	7.5	570	30	T67540	H. pylori inner membr 4.78e+02
c	378	13	7.5	582	40	V15020	JP2 strain leukotoxin 4.78e+02
c	379	13	7.5	584	40	V15022	JP2 strain ltx operon 4.78e+02
c	380	13	7.5	615	30	T67480	H. pylori transporter 4.78e+02
	381	13	7.5	660	13	Q79075	Cytadhesin P1 C-termi 4.78e+02
	382	13	7.5	668	1	N82176	31-1 Repeated Delete. 4.78e+02
c	383	13	7.5	669	40	V10151	Stealth virus nucleic 4.78e+02
	384	13	7.5	674	18	T03506	Papilloma virus major 4.78e+02
c	385	13	7.5	678	30	T67356	H. pylori cytoplasmic 4.78e+02
	386	13	7.5	693	1	Q06408	Sequence hybridising 4.78e+02
	387	13	7.5	693	3	N50336	Sequence coding for m 4.78e+02
c	388	13	7.5	720	30	T68257	H. pylori inner membr 4.78e+02
c	389	13	7.5	737	3	N50355	Storage-specific, lat 4.78e+02
c	390	13	7.5	748	21	T29001	Parietaria allergen P 4.78e+02
c	391	13	7.5	767	13	Q78990	Human immunoglobulin 4.78e+02
c	392	13	7.5	771	40	V24719	H. pylori ORF 06ep103 4.78e+02
c	393	13	7.5	794	15	Q87927	Anther specific gene p 4.78e+02
c	394	13	7.5	810	30	T68188	H. pylori transporter 4.78e+02
	395	13	7.5	819	21	T12557	DNA encoding LIP3. 4.78e+02
	396	13	7.5	825	30	T68152	H. pylori cytoplasmic 4.78e+02
	397	13	7.5	825	2	N60149	Sequence encoding pol 4.78e+02
	398	13	7.5	828	30	T67928	H. pylori surface or 4.78e+02
c	399	13	7.5	830	33	T62455	Lambda 5h-1 clone con 4.78e+02

	400	13	7.5	834	17	T03998	Rice fungal resistanc 4.78e+02
c	401	13	7.5	861	2	Q11098	Encodes gamma heavy c 4.78e+02
c	402	13	7.5	891	30	T67951	H. pylori cytoplasmic 4.78e+02
c	403	13	7.5	919	36	T90048	Chicken urokinase cDN 4.78e+02
c	404	13	7.5	924	3	Q20474	Sequence encoding bac 4.78e+02
	405	13	7.5	940	2	N60906	Sequence encodes mous 4.78e+02
c	406	13	7.5	960	5	Q30850	3' UTR of Type III pr 4.78e+02
	407	13	7.5	963	30	T68017	H. pylori transporter 4.78e+02
	408	13	7.5	964	35	T79676	BRCA2 cancer suscepti 4.78e+02
	409	13	7.5	972	40	V24793	H. pylori ORF 11gel03 4.78e+02
c	410	13	7.5	992	40	V15030	652 strain leukotoxin 4.78e+02
c	411	13	7.5	997	40	V15021	652 strain leukotoxin 4.78e+02
c	412	13	7.5	1008	2	N60420	Sequence encoding ant 4.78e+02
c	413	13	7.5	1013	2	Q12431	Fragment D of urate o 4.78e+02
c	414	13	7.5	1029	6	Q38933	deg-1 gene. 4.78e+02
	415	13	7.5	1041	35	T79675	BRCA2 cancer suscepti 4.78e+02
c	416	13	7.5	1046	30	T68955	Blackcurrant pRIB5 cD 4.78e+02
	417	13	7.5	1050	35	T79674	BRCA2 cancer suscepti 4.78e+02
c	418	13	7.5	1053	30	T67538	H. pylori secreted or 4.78e+02
c	419	13	7.5	1056	35	T79681	BRCA2 cancer suscepti 4.78e+02
c	420	13	7.5	1062	2	N60458	Sequence encoding seg 4.78e+02
	421	13	7.5	1110	30	T68176	H. pylori cytoplasmic 4.78e+02
c	422	13	7.5	1118	3	Q20263	Deg-1 gene. 4.78e+02
c	423	13	7.5	1119	30	T68086	H. pylori secreted or 4.78e+02
c	424	13	7.5	1121	2	Q12430	Urate oxidase coding 4.78e+02
c	425	13	7.5	1121	2	Q10245	Clone 9C encoding a p 4.78e+02
c	426	13	7.5	1128	30	T68185	H. pylori inner membr 4.78e+02
	427	13	7.5	1140	22	T37081	AR781 fragment encodi 4.78e+02
c	428	13	7.5	1144	38	T95782	Stylosanthes humilis 4.78e+02
	429	13	7.5	1158	24	T42983	Adenovirus proteinase 4.78e+02
	430	13	7.5	1159	7	Q40592	JSC1 precursor DNA. 4.78e+02
c	431	13	7.5	1162	4	N30044	Sequence encoding bov 4.78e+02
c	432	13	7.5	1169	38	V12144	Helicobacter pylori r 4.78e+02
c	433	13	7.5	1169	15	Q95252	H. pylori recA gene. 4.78e+02
c	434	13	7.5	1182	4	N30043	Sequence encoding bov 4.78e+02
c	435	13	7.5	1188	1	N82040	La gene. 4.78e+02
	436	13	7.5	1189	7	Q40203	Sequence of JPSIg2 DN 4.78e+02
	437	13	7.5	1234	1	N80494	Sequence encoding hum 4.78e+02
c	438	13	7.5	1254	1	N92937	cDNA from Plasmodium 4.78e+02
	439	13	7.5	1284	3	Q15361	NF-YB. 4.78e+02
c	440	13	7.5	1289	38	T97166	Mouse HPRT gene. 4.78e+02
	441	13	7.5	1290	30	T67838	H. pylori secreted or 4.78e+02
c	442	13	7.5	1296	14	Q79124	Cancer metastasis pro 4.78e+02
c	443	13	7.5	1310	22	T36778	Prostate-specific mem 4.78e+02
	444	13	7.5	1311	9	Q55628	Mouse GM-CSF gene. 4.78e+02
	445	13	7.5	1336	13	Q78233	Alkaline serine prote 4.78e+02
	446	13	7.5	1383	30	T66165	Human interleukin-12 4.78e+02
c	447	13	7.5	1386	24	T38228	Apoptosis-regulating 4.78e+02
	448	13	7.5	1401	3	Q22209	Peroxisome forming fa 4.78e+02
	449	13	7.5	1401	3	Q22208	Peroxisome forming fa 4.78e+02
c	450	13	7.5	1419	11	Q63208	Human c-myc exon 3 en 4.78e+02

c	451	13	7.5	1419	8	Q49480	Human c-myc exon 3. 4.78e+02
c	452	13	7.5	1429	40	V04237	Arabidopsis C-14 ster 4.78e+02
c	453	13	7.5	1432	10	Q44494	Bacillus coagulans pr 4.78e+02
	454	13	7.5	1440	3	N50528	Parasponia rhizobium 4.78e+02
	455	13	7.5	1449	1	N90005	Promoter and coding D 4.78e+02
	456	13	7.5	1449	2	Q11136	NifD gene and promote 4.78e+02
	457	13	7.5	1457	40	V15409	Methicillin resistant 4.78e+02
	458	13	7.5	1468	29	T66425	Human P28 coding sequ 4.78e+02
	459	13	7.5	1470	2	Q10869	Plasmodium vivax 66kD 4.78e+02
	460	13	7.5	1473	1	N92037	Sequence of variant h 4.78e+02
	461	13	7.5	1474	3	N50138	Sequence of the signa 4.78e+02
	462	13	7.5	1475	1	N81558	Pro-UK structural gen 4.78e+02
	463	13	7.5	1475	2	N70356	Sequence of the pre-s 4.78e+02
	464	13	7.5	1475	3	N50141	Sequence encoding pro 4.78e+02
	465	13	7.5	1500	1	N91618	Human pro-urokinase c 4.78e+02
	466	13	7.5	1500	2	N91617	Sequence of entire pr 4.78e+02
	467	13	7.5	1538	31	T69549	Rat pheromone recepto 4.78e+02
	468	13	7.5	1560	4	Q24184	Human GAD65 DNA. 4.78e+02
c	469	13	7.5	1560	14	Q83201	Precursor of novel am 4.78e+02
c	470	13	7.5	1565	24	T39908	Maize acetyl CoA carb 4.78e+02
c	471	13	7.5	1599	1	N90733	Genomic DNA from Plas 4.78e+02
	472	13	7.5	1613	24	T39595	Triticum tauschii CRE 4.78e+02
c	473	13	7.5	1617	10	Q62312	Human somatoliberin r 4.78e+02
c	474	13	7.5	1629	18	T15623	Bacillus stearothermo 4.78e+02
c	475	13	7.5	1636	40	V10435	Human chitinase clone 4.78e+02
c	476	13	7.5	1637	38	T89180	Human chitotriosidase 4.78e+02
c	477	13	7.5	1643	27	T50833	Human chitinase cDNA 4.78e+02
	478	13	7.5	1655	7	Q43720	Sequence which hybrid 4.78e+02
c	479	13	7.5	1656	40	V10436	Human chitinase clone 4.78e+02
	480	13	7.5	1671	4	Q24134	50 kD subunit of SceI 4.78e+02
c	481	13	7.5	1713	27	T50834	Human chitinase cDNA 4.78e+02
	482	13	7.5	1721	6	Q37971	Minimal (bases 123-18 4.78e+02
	483	13	7.5	1728	12	Q73481	Low molecular weight 4.78e+02
c	484	13	7.5	1739	4	Q27539	Tyrosine Kinase recep 4.78e+02
c	485	13	7.5	1752	4	Q25808	Mouse thyrotropin-rel 4.78e+02
	486	13	7.5	1755	8	Q49524	Human interferon rece 4.78e+02
	487	13	7.5	1755	14	Q86458	Human IFN receptor ge 4.78e+02
	488	13	7.5	1755	3	Q14240	Encodes complete inte 4.78e+02
	489	13	7.5	1755	5	Q30533	Sequence encoding a s 4.78e+02
	490	13	7.5	1761	30	T64560	Modified glutamic aci 4.78e+02
	491	13	7.5	1776	33	T51339	Coding sequence for a 4.78e+02
c	492	13	7.5	1777	13	Q79355	Sequence of the exten 4.78e+02
c	493	13	7.5	1794	1	N82201	Beta-amylase from pla 4.78e+02
c	494	13	7.5	1804	40	V20462	Human c-yes oncogene. 4.78e+02
c	495	13	7.5	1804	36	T71262	Human c-yes-2 gene. 4.78e+02
	496	13	7.5	1807	3	Q30243	Mycoplasma-specific D 4.78e+02
	497	13	7.5	1823	35	T75633	Canine herpesvirus ge 4.78e+02
c	498	13	7.5	1825	26	T45951	Rat 25-hydroxyvitamin 4.78e+02
c	499	13	7.5	1831	16	Q97312	Rat A3 adenosine rece 4.78e+02
	500	13	7.5	1833	32	T73363	Konjak mosaic virus o 4.78e+02
	501	13	7.5	1838	37	T79352	Staphylococcus aureus 4.78e+02

c	502	13	7.5	1896	23	T36788	Prostate-specific mem 4.78e+02
c	503	13	7.5	1920	40	V17563	Coding sequence for t 4.78e+02
c	504	13	7.5	1923	9	Q57524	OBP1LT sequence. 4.78e+02
c	505	13	7.5	1924	1	N90684	Gene encoding the 43 4.78e+02
c	506	13	7.5	1927	8	Q47841	Flavonoid-3',5'-hydro 4.78e+02
c	507	13	7.5	1937	38	V12143	Helicobacter pylori r 4.78e+02
c	508	13	7.5	1945	8	Q47816	Sequence encoding 43k 4.78e+02
	509	13	7.5	1954	21	Q94450	Bacterial transferrin 4.78e+02
c	510	13	7.5	1954	7	Q45598	Sequence of the mouse 4.78e+02
	511	13	7.5	1955	38	T49508	Transferrin receptor 4.78e+02
c	512	13	7.5	1969	1	N93087	Human basic fibroblas 4.78e+02
c	513	13	7.5	1981	1	Q04267	Probe containing codi 4.78e+02
c	514	13	7.5	1982	1	N90225	Malaria-specific Plas 4.78e+02
	515	13	7.5	1985	2	N60126	Sequence encoding mur 4.78e+02
	516	13	7.5	2015	2	N70047	Encodes human decay a 4.78e+02
c	517	13	7.5	2022	21	T16016	DNA encoding ubiquiti 4.78e+02
c	518	13	7.5	2031	17	T15544	Borrelia burgdorferi 4.78e+02
	519	13	7.5	2040	24	T45651	Human ST-S1 cDNA, use 4.78e+02
	520	13	7.5	2040	5	Q31246	Chalcone synthase pro 4.78e+02
c	521	13	7.5	2044	35	T75623	Canine herpesvirus ge 4.78e+02
	522	13	7.5	2073	3	N50058	Sequence of nitrogena 4.78e+02
	523	13	7.5	2073	1	N90006	Promoter and coding D 4.78e+02
	524	13	7.5	2073	2	Q11137	NifD gene and promote 4.78e+02
	525	13	7.5	2075	18	T04226	Connective tissue gro 4.78e+02
	526	13	7.5	2075	28	T65360	Human connective tiss 4.78e+02
	527	13	7.5	2075	26	T51234	Connective tissue gro 4.78e+02
c	528	13	7.5	2075	17	T15541	Borrelia burgdorferi 4.78e+02
	529	13	7.5	2096	15	Q99105	Human DAF cDNA. 4.78e+02
	530	13	7.5	2096	33	T84471	Human DAF cDNA. 4.78e+02
c	531	13	7.5	2097	9	Q52316	Coriolus hirsutus ARS 4.78e+02
	532	13	7.5	2099	2	Q10418	Plasmodium knowlesi 6 4.78e+02
	533	13	7.5	2100	15	Q86482	Human glutamic acid d 4.78e+02
	534	13	7.5	2100	6	Q32750	Human GAD 65 gene. 4.78e+02
c	535	13	7.5	2100	1	Q05676	v-PA_gamma. 4.78e+02
c	536	13	7.5	2104	4	Q25273	Sequence encoding the 4.78e+02
	537	13	7.5	2115	13	Q79863	Decay accelerating fa 4.78e+02
	538	13	7.5	2115	1	N91043	Decay accelerating fa 4.78e+02
	539	13	7.5	2125	23	T36787	Prostate-specific mem 4.78e+02
c	540	13	7.5	2127	21	T12414	Ubi7 ubiquitin-lytic 4.78e+02
c	541	13	7.5	2137	1	Q05675	v-PA_beta. 4.78e+02
c	542	13	7.5	2144	15	Q87843	Human calcium channel 4.78e+02
c	543	13	7.5	2153	15	Q87842	Human calcium channel 4.78e+02
	544	13	7.5	2179	13	Q73117	Alpha 2, 3-sialyl tra 4.78e+02
	545	13	7.5	2232	13	Q77739	Human alpha-2,3-sialy 4.78e+02
	546	13	7.5	2233	2	N70048	Encodes human decay a 4.78e+02
	547	13	7.5	2233	1	N91044	Membrane bound decay 4.78e+02
	548	13	7.5	2233	13	Q79864	Decay accelerating fa 4.78e+02
	549	13	7.5	2237	37	T90287	Human glycophosphatid 4.78e+02
	550	13	7.5	2249	3	Q15169	Bacillus thuringiensi 4.78e+02
	551	13	7.5	2249	32	T61097	65 kD human glutamic 4.78e+02
c	552	13	7.5	2252	1	Q05673	v-PA_alphal. 4.78e+02

c	553	13	7.5	2257	1	Q05674	v-PA_alpha2. 4.78e+02
c	554	13	7.5	2264	17	T15542	Borrelia afzelii 66 k 4.78e+02
	555	13	7.5	2266	1	N93079	Sequence encoding pro 4.78e+02
	556	13	7.5	2298	1	N80981	Sequence of the singl 4.78e+02
	557	13	7.5	2303	12	Q73483	Full length human uro 4.78e+02
	558	13	7.5	2303	1	N91740	Sequence of prourokin 4.78e+02
	559	13	7.5	2304	2	N71338	Modified prourokinase 4.78e+02
	560	13	7.5	2304	4	N30030	Sequence encoding ful 4.78e+02
	561	13	7.5	2304	2	N71369	Modified prourokinase 4.78e+02
	562	13	7.5	2304	2	N71368	Modified prourokinase 4.78e+02
	563	13	7.5	2304	2	N71330	Sequence encoding mod 4.78e+02
c	564	13	7.5	2315	6	Q39091	XR79 coding sequence. 4.78e+02
c	565	13	7.5	2322	23	T27627	Malic enzyme coding s 4.78e+02
c	566	13	7.5	2332	9	Q51275	Basic protease gene. 4.78e+02
	567	13	7.5	2333	4	Q27387	Clone pY2 encoding pa 4.78e+02
c	568	13	7.5	2337	1	N91483	NP cDNA to genomic ju 4.78e+02
c	569	13	7.5	2345	17	Q88167	Nr (never ripe) ethyl 4.78e+02
	570	13	7.5	2362	26	T48098	Human interleukin-12 4.78e+02
	571	13	7.5	2362	4	Q23586	Natural killer cell s 4.78e+02
	572	13	7.5	2362	40	V15227	Heterodimeric human i 4.78e+02
	573	13	7.5	2370	14	Q86046	Human GAD gene. 4.78e+02
	574	13	7.5	2370	5	Q31783	Human pancreatic isle 4.78e+02
	575	13	7.5	2377	3	Q20360	Human pro-urokinase c 4.78e+02
c	576	13	7.5	2404	17	Q88163	Tomato ethylene respo 4.78e+02
	577	13	7.5	2412	36	T97869	cDNA encoding a chime 4.78e+02
c	578	13	7.5	2420	8	Q46672	p mysA cDNA clone. 4.78e+02
	579	13	7.5	2425	1	Q05679	btPGS1208 gene. 4.78e+02
	580	13	7.5	2427	1	Q04107	Human pro-urokinase c 4.78e+02
	581	13	7.5	2438	1	N90373	DNA encoding amphirey 4.78e+02
	582	13	7.5	2526	2	N70042	Varicella-zoster viru 4.78e+02
c	583	13	7.5	2542	39	V15094	Human apolipoprotein(4.78e+02
c	584	13	7.5	2547	17	T15543	Borrelia garinii 66 k 4.78e+02
	585	13	7.5	2593	35	T92778	Oryza sativa pathogen 4.78e+02
c	586	13	7.5	2605	2	Q10790	Maize Lc cDNA sequenc 4.78e+02
c	587	13	7.5	2658	29	T60339	Pumpkin ent-kaurene B 4.78e+02
c	588	13	7.5	2669	14	Q82793	TRK2 gene. 4.78e+02
c	589	13	7.5	2675	1	Q06957	Genomic sequence enco 4.78e+02
	590	13	7.5	2692	1	Q05113	Sequence encoding pol 4.78e+02
c	591	13	7.5	2700	38	T97447	Pacific yew taxadiene 4.78e+02
	592	13	7.5	2721	21	T34617	Operon encoding LIP1, 4.78e+02
	593	13	7.5	2722	21	T12675	E. histolytica pyruva 4.78e+02
	594	13	7.5	2722	16	Q94644	Recombinant cold-resi 4.78e+02
	595	13	7.5	2729	30	T62363	Human origin of repli 4.78e+02
	596	13	7.5	2729	31	T73290	Human origin of repli 4.78e+02
	597	13	7.5	2730	3	Q20994	EHV-4 gH gene. 4.78e+02
c	598	13	7.5	2741	38	T96844	Intron 14 of human pR 4.78e+02
	599	13	7.5	2760	14	Q85425	T2 DNA-polymerase wil 4.78e+02
	600	13	7.5	2760	14	Q85426	T4 DNA-polymerase wil 4.78e+02
c	601	13	7.5	2763	1	N95079	Gmhsp26-A heat shock 4.78e+02
	602	13	7.5	2784	2	Q11701	Human alpha-interfero 4.78e+02
	603	13	7.5	2848	30	T47041	Human protein ubiquit 4.78e+02

	604	13	7.5	2861	24	T28639	RNase L inhibitor *H2 4.78e+02
	605	13	7.5	2932	3	Q15325	Beta-amylase pMM2 clo 4.78e+02
c	606	13	7.5	2950	10	Q58744	aMEF2. 4.78e+02
	607	13	7.5	2959	3	Q15326	Beta-amylase pMM3 clo 4.78e+02
c	608	13	7.5	2968	10	Q58740	MEF2. 4.78e+02
	609	13	7.5	3047	26	T51225	Phosphofructokinase c 4.78e+02
	610	13	7.5	3099	2	Q13114	Encodes partial Human 4.78e+02
c	611	13	7.5	3101	1	Q02047	Sequence encoding a c 4.78e+02
c	612	13	7.5	3113	2	Q13870	Cruciferin A gene. 4.78e+02
c	613	13	7.5	3183	37	T79351	Methicillin sensitive 4.78e+02
	614	13	7.5	3188	29	T60035	Blight resistance ind 4.78e+02
	615	13	7.5	3223	11	Q66475	Inulinase coding sequ 4.78e+02
	616	13	7.5	3278	30	T52358	Kluyveromyces lactis 4.78e+02
	617	13	7.5	3278	31	T73285	K. lactis origin of r 4.78e+02
	618	13	7.5	3321	40	V04207	Human metabotropic gl 4.78e+02
	619	13	7.5	3321	40	V04206	Human metabotropic gl 4.78e+02
c	620	13	7.5	3376	13	Q75166	AF-9 cDNA. 4.78e+02
c	621	13	7.5	3410	15	T03887	Human mGluR3 DNA. 4.78e+02
c	622	13	7.5	3442	3	Q20545	Encodes KEX2 endopept 4.78e+02
c	623	13	7.5	3458	8	Q48714	Insecticidal protein 4.78e+02
c	624	13	7.5	3458	3	Q14256	51.4 and 41.9 kD toxi 4.78e+02
c	625	13	7.5	3460	25	T42193	Fragment of pig oestr 4.78e+02
	626	13	7.5	3486	12	Q71820	Chondoitinase ABC cDN 4.78e+02
	627	13	7.5	3504	28	T61018	Bacillus thuringiensi 4.78e+02
	628	13	7.5	3504	37	T89185	Nematode toxin 167P g 4.78e+02
	629	13	7.5	3504	31	T62498	Nematocidal toxin gen 4.78e+02
	630	13	7.5	3562	33	T62170	Candida boidinii S2 A 4.78e+02
	631	13	7.5	3568	24	T28638	RNase L inhibitor *RL 4.78e+02
c	632	13	7.5	3607	37	T45420	3' UTR of the ataxia- 4.78e+02
c	633	13	7.5	3607	31	T43498	ATM gene 3' untransla 4.78e+02
c	634	13	7.5	3637	1	N91475	Human topoisomerase I 4.78e+02
c	635	13	7.5	3645	16	Q91432	Variant human DNA top 4.78e+02
c	636	13	7.5	3706	23	T36392	Lactococcus lactis su 4.78e+02
c	637	13	7.5	3762	35	T73217	ApxICA gene. 4.78e+02
	638	13	7.5	3833	40	V17116	Metabotropic glutamat 4.78e+02
	639	13	7.5	3850	24	T39598	Triticum tauschii nem 4.78e+02
c	640	13	7.5	3852	30	T62419	Bovine beta-mannosida 4.78e+02
	641	13	7.5	3853	3	Q15324	Beta-amylase pMM1 clo 4.78e+02
c	642	13	7.5	3877	37	V02209	cDNA of leaderless pr 4.78e+02
c	643	13	7.5	3878	2	N70494	Sequence encoding hum 4.78e+02
c	644	13	7.5	3878	1	N81502	DNA encoding human ba 4.78e+02
c	645	13	7.5	3919	13	Q80418	Human mGluR3 cDNA. 4.78e+02
	646	13	7.5	3975	15	Q92412	pNF1022 4-kb insert f 4.78e+02
	647	13	7.5	3980	27	T58981	Mutated Proteus vulga 4.78e+02
	648	13	7.5	3980	12	Q74989	P. vulgaris chondroit 4.78e+02
	649	13	7.5	3980	12	Q74988	P. vulgaris chondroit 4.78e+02
	650	13	7.5	3980	27	T58980	Proteus vulgaris chon 4.78e+02
	651	13	7.5	3982	32	T58244	CH8-2a13-1, over abun 4.78e+02
	652	13	7.5	3987	3	Q14629	Human Glioblastoma ce 4.78e+02
	653	13	7.5	3987	22	T34376	Plasmid pTM22 (ATCC 6 4.78e+02
	654	13	7.5	3994	3	N40009	Saccharomyces cerevis 4.78e+02

	655	13	7.5	3994	2	Q11883	Invertase gene. 4.78e+02
	656	13	7.5	4009	40	T85827	Human interleukin-13 4.78e+02
	657	13	7.5	4103	32	T58245	Full length CH8-2a13- 4.78e+02
c	658	13	7.5	4106	15	Q98696	Yeast KEX2 gene. 4.78e+02
	659	13	7.5	4133	15	Q87268	B. sphaericus mtx gen 4.78e+02
	660	13	7.5	4133	8	Q48715	Insecticidal protein 4.78e+02
	661	13	7.5	4133	3	Q14257	Mosquitotoxin gene. 4.78e+02
c	662	13	7.5	4140	36	T77331	Solanum tuberosum col 4.78e+02
	663	13	7.5	4214	28	T58534	Human connective tiss 4.78e+02
	664	13	7.5	4214	29	T59618	Connective tissue gro 4.78e+02
c	665	13	7.5	4249	37	T91326	Arabidopsis thaliana 4.78e+02
c	666	13	7.5	4346	24	T45897	Serine kinase SRPK1 c 4.78e+02
	667	13	7.5	4407	9	Q64637	ced-4 (C1131T) gene. 4.78e+02
	668	13	7.5	4407	9	Q54665	ced-4 gene. 4.78e+02
	669	13	7.5	4407	9	Q64641	ced-4 (C2128T) gene. 4.78e+02
	670	13	7.5	4407	9	Q64639	ced-4 (G1929A) gene. 4.78e+02
	671	13	7.5	4407	9	Q64638	ced-4 (C1428T) gene. 4.78e+02
	672	13	7.5	4407	9	Q64640	ced-4 (T2117A) gene. 4.78e+02
	673	13	7.5	4407	9	Q64642	ced-4 (G3131A) gene. 4.78e+02
c	674	13	7.5	4407	9	Q64639	ced-4 (G1929A) gene. 4.78e+02
c	675	13	7.5	4407	9	Q64640	ced-4 (T2117A) gene. 4.78e+02
c	676	13	7.5	4407	9	Q64637	ced-4 (C1131T) gene. 4.78e+02
c	677	13	7.5	4407	9	Q64641	ced-4 (C2128T) gene. 4.78e+02
c	678	13	7.5	4407	9	Q64642	ced-4 (G3131A) gene. 4.78e+02
c	679	13	7.5	4407	9	Q64638	ced-4 (C1428T) gene. 4.78e+02
c	680	13	7.5	4407	9	Q54665	ced-4 gene. 4.78e+02
	681	13	7.5	4464	2	N71244	Genomic sequence of a 4.78e+02
	682	13	7.5	4465	5	Q28697	Glutamine synthetase. 4.78e+02
	683	13	7.5	4495	24	T32143	Helicobacter-specific 4.78e+02
c	684	13	7.5	4495	24	T32143	Helicobacter-specific 4.78e+02
c	685	13	7.5	4526	8	Q49780	Prosystemin. 4.78e+02
c	686	13	7.5	4527	24	T33899	HIV-1 SF33 genome. 4.78e+02
	687	13	7.5	4590	2	N60472	Sequence encoding the 4.78e+02
c	688	13	7.5	4599	22	T27052	RRP3 telomerase-assoc 4.78e+02
	689	13	7.5	4673	4	Q27189	P. yoelii SSP2 antige 4.78e+02
	690	13	7.5	4699	38	T49500	Transferrin receptor 4.78e+02
	691	13	7.5	4699	21	Q94442	Bacterial transferrin 4.78e+02
c	692	13	7.5	4699	21	Q94442	Bacterial transferrin 4.78e+02
c	693	13	7.5	4699	38	T49500	Transferrin receptor 4.78e+02
	694	13	7.5	4762	1	N81675	Plasmid pBSFOLEK1 con 4.78e+02
	695	13	7.5	4774	1	N81424	Plasmid pBLAK1 stably 4.78e+02
	696	13	7.5	4785	8	Q48797	B.ohbensis Trp(188) c 4.78e+02
	697	13	7.5	4785	3	Q14942	Cyclodextrin glucanot 4.78e+02
	698	13	7.5	4785	3	Q14072	Cyclodextrin glucanot 4.78e+02
	699	13	7.5	4785	8	Q48788	B.ohbensis mutant cyc 4.78e+02
	700	13	7.5	4785	8	Q48794	B.ohbensis Ala(188) c 4.78e+02
	701	13	7.5	4785	8	Q48795	B.ohbensis Ser(188) c 4.78e+02
	702	13	7.5	4785	8	Q48796	B.ohbensis Ser(188) c 4.78e+02
	703	13	7.5	4789	1	N82213	Plasmid pGIF1 encodes 4.78e+02
	704	13	7.5	4852	12	Q70448	Human nestin gene - i 4.78e+02
	705	13	7.5	4854	5	Q28399	Human nestin gene. 4.78e+02

c	706	13	7.5	4875	37	T96058	E. coli colonisation 4.78e+02
	707	13	7.5	4954	14	Q55753	Escherichia coli geno 4.78e+02
c	708	13	7.5	5009	38	T49502	Transferrin receptor 4.78e+02
c	709	13	7.5	5009	21	Q94444	Bacterial transferrin 4.78e+02
c	710	13	7.5	5033	21	Q94443	Bacterial transferrin 4.78e+02
c	711	13	7.5	5033	38	T49501	Transferrin receptor 4.78e+02
c	712	13	7.5	5040	2	N80916	Sequence encoding the 4.78e+02
c	713	13	7.5	5099	21	Q94445	Bacterial transferrin 4.78e+02
c	714	13	7.5	5099	38	T49503	Transferrin receptor 4.78e+02
	715	13	7.5	5108	2	N60388	Sequence encoding hum 4.78e+02
c	716	13	7.5	5144	21	Q94446	Bacterial transferrin 4.78e+02
c	717	13	7.5	5144	38	T49504	Transferrin receptor 4.78e+02
	718	13	7.5	5168	13	Q79074	Cytadhesin P1 gene. 4.78e+02
	719	13	7.5	5169	10	Q55775	Mycoplasma pneumoniae 4.78e+02
	720	13	7.5	5169	2	Q12520	M. pneumoniae P1 cyta 4.78e+02
	721	13	7.5	5181	13	Q80911	Plasmodium falciparum 4.78e+02
c	722	13	7.5	5291	1	Q05768	GPAL2 promoter. 4.78e+02
c	723	13	7.5	5376	1	N90732	Sequence of plasmid p 4.78e+02
c	724	13	7.5	5460	20	T16508	Vector pAC3A1 contg. 4.78e+02
	725	13	7.5	5468	2	Q06540	Human low density lip 4.78e+02
c	726	13	7.5	5760	3	N50530	Sequence encoding the 4.78e+02
	727	13	7.5	5872	35	T90544	Maize male fertility 4.78e+02
c	728	13	7.5	5892	35	T79680	BRCA2 cancer suscepti 4.78e+02
	729	13	7.5	5893	8	Q46099	A. rhizogenes NIAES17 4.78e+02
c	730	13	7.5	5912	37	T45423	Partial ATM gene tran 4.78e+02
c	731	13	7.5	5912	31	T43519	ATM gene clone 7-9. 4.78e+02
	732	13	7.5	6169	24	T39284	Mobilisable cloning v 4.78e+02
c	733	13	7.5	6236	34	T85728	FMR2 gene associated 4.78e+02
	734	13	7.5	6313	5	Q29965	Sugar beet chitinase 4.78e+02
	735	13	7.5	6519	27	T58982	Proteus vulgaris chon 4.78e+02
	736	13	7.5	6519	12	Q74990	P. vulgaris chondroit 4.78e+02
	737	13	7.5	6578	19	T18767	RSV RNA-dependent RNA 4.78e+02
c	738	13	7.5	6767	1	Q05883	Plasmid clone pUC-SK1 4.78e+02
c	739	13	7.5	6854	15	Q99852	Yeast TIH2 gene. 4.78e+02
	740	13	7.5	6901	35	T78853	Human lecithin-choles 4.78e+02
c	741	13	7.5	7240	35	T79668	Partial BRCA2 cancer 4.78e+02
	742	13	7.5	7241	3	Q15140	Genomic clone LE-ACC2 4.78e+02
	743	13	7.5	7244	40	V15705	Tomato ACC synthase L 4.78e+02
c	744	13	7.5	7366	2	Q10342	Modified lac operon. 4.78e+02
c	745	13	7.5	7616	12	Q70194	Vector pOTS208NS1BLmu 4.78e+02
	746	13	7.5	7753	4	Q26727	Encodes exons XV to X 4.78e+02
c	747	13	7.5	7860	24	T44380	Stretch-activated cat 4.78e+02
c	748	13	7.5	8082	40	V20463	Human c-myc oncogene. 4.78e+02
c	749	13	7.5	8082	6	Q37740	Complete human c-myc 4.78e+02
c	750	13	7.5	8082	36	T71259	Human c-myb oncogene. 4.78e+02
c	751	13	7.5	8391	19	T16333	MLL gene 8.3 kb fragm 4.78e+02
c	752	13	7.5	8392	9	Q53478	MLL gene 8.3 kb BamHI 4.78e+02
	753	13	7.5	8797	27	T50856	DNA cassette for amph 4.78e+02
c	754	13	7.5	9060	40	V15702	Zucchini ACC synthase 4.78e+02
	755	13	7.5	9202	2	N70608	VISNA sheep lentiviru 4.78e+02
c	756	13	7.5	9272	13	Q79353	Human genomic clone h 4.78e+02

c	757	13	7.5	9432	17	Q99312	LKP1 operon.	4.78e+02
	758	13	7.5	9919	40	V04631	Arabidopsis thaliana	4.78e+02
	759	13	7.5	10216	24	T39279	Transposon delivery v	4.78e+02
c	760	13	7.5	10614	15	Q89555	Hamster cholesterol 7	4.78e+02
c	761	13	7.5	10798	4	Q28505	Tomato vacuolar inver	4.78e+02
c	762	13	7.5	10965	4	Q28507	Tomato vacuolar inver	4.78e+02
	763	13	7.5	11093	27	T50855	DNA cassette for amph	4.78e+02
c	764	13	7.5	11225	2	N91695	Sequence of fowlpox v	4.78e+02
c	765	13	7.5	11283	35	T84841	Human breast and ovar	4.78e+02
c	766	13	7.5	11385	36	T69707	Human breast cancer s	4.78e+02
c	767	13	7.5	12778	31	T68784	ATM mutant G9170C cod	4.78e+02
	768	13	7.5	13104	8	Q46852	Clone of recombinant	4.78e+02
c	769	13	7.5	13807	32	T62328	Exopolysaccharide ope	4.78e+02
c	770	13	7.5	14176	34	T84564	Swinepox virus HindII	4.78e+02
c	771	13	7.5	14237	39	V05384	Human cathepsin K gen	4.78e+02
c	772	13	7.5	14602	31	T73236	Streptococcus thermop	4.78e+02
	773	13	7.5	15222	36	T78440	Human respiratory syn	4.78e+02
	774	13	7.5	15223	29	T63430	Respiratory syncytial	4.78e+02
c	775	13	7.5	17041	12	Q67057	PACPAP38 DNA.	4.78e+02
c	776	13	7.5	17041	3	Q21065	Genomic DNA of human	4.78e+02
	777	13	7.5	19124	32	T72882	Plasmodium var-7 gene	4.78e+02
	778	13	7.5	19307	20	T27558	Shuttle vector pADBel	4.78e+02
c	779	13	7.5	20303	31	T71699	Human deoxycytidylate	4.78e+02
c	780	13	7.5	26698	20	T08126	Mouse syndecan-1 gene	4.78e+02
c	781	13	7.5	26700	40	V15946	Mouse syndecan gene s	4.78e+02
c	782	13	7.5	26700	11	Q67902	Syndecan gene.	4.78e+02
c	783	13	7.5	26764	31	T71696	Human deoxycytidylate	4.78e+02
	784	13	7.5	28994	39	V15826	Genomic DNA for inter	4.78e+02
	785	13	7.5	29392	40	V15422	Mouse poly Ig recepto	4.78e+02
	786	13	7.5	40352	40	V02032	MAGE-B cluster DNA se	4.78e+02
c	787	13	7.5	40352	40	V02032	MAGE-B cluster DNA se	4.78e+02
	788	13	7.5	134525	1	Q04525	Total base sequence o	4.78e+02
c	789	13	7.5	134525	1	Q04525	Total base sequence o	4.78e+02
	790	12	6.9	40	24	T43052	Helicobacter pylori F	1.49e+03
c	791	12	6.9	42	29	T48509	Synthetic late promot	1.49e+03
	792	12	6.9	62	19	T20449	Human gene signature	1.49e+03
c	793	12	6.9	70	16	Q99648	Swinepox early/late p	1.49e+03
	794	12	6.9	82	39	V04455	Synthetic oligonucleo	1.49e+03
c	795	12	6.9	94	16	Q94699	Mycoplasma gallisepti	1.49e+03
c	796	12	6.9	134	7	Q40250	Sequence of a DNA seg	1.49e+03
	797	12	6.9	152	10	Q76819	Human genome fragment	1.49e+03
c	798	12	6.9	206	34	T65073	Canine genomic micros	1.49e+03
c	799	12	6.9	208	19	T20624	Human gene signature	1.49e+03
c	800	12	6.9	217	40	V05555	Ubiquitin gene intron	1.49e+03
c	801	12	6.9	224	21	T24532	Human gene signature	1.49e+03
c	802	12	6.9	226	24	T43285	Promoter for infectiv	1.49e+03
c	803	12	6.9	230	22	T18576	Human chondrocyte RNA	1.49e+03
	804	12	6.9	267	20	T23253	Human gene signature	1.49e+03
	805	12	6.9	273	8	Q61202	Human brain Expressed	1.49e+03
c	806	12	6.9	277	1	Q06363	Probe for Clostridium	1.49e+03
c	807	12	6.9	294	22	T25203	Human gene signature	1.49e+03

c	808	12	6.9	301	18	T19561	Human gene signature	1.49e+03
	809	12	6.9	309	8	Q60586	Human brain Expressed	1.49e+03
c	810	12	6.9	315	40	V24779	H. pylori ORF 29ge101	1.49e+03
	811	12	6.9	319	21	T25360	Human gene signature	1.49e+03
c	812	12	6.9	335	35	T47089	Partial P. rubrum his	1.49e+03
c	813	12	6.9	338	8	Q60590	Human brain Expressed	1.49e+03
	814	12	6.9	358	8	Q59477	Human brain Expressed	1.49e+03
c	815	12	6.9	360	5	Q30070	Sequence of the nisZ	1.49e+03
	816	12	6.9	378	39	V01094	Humanised light chain	1.49e+03
c	817	12	6.9	397	39	V02300	Human secreted protei	1.49e+03
c	818	12	6.9	397	39	T88079	3' portion of cDNA cl	1.49e+03
c	819	12	6.9	397	38	V02144	Human secreted protei	1.49e+03
	820	12	6.9	413	2	Q13400	Mouse U6 intranuclear	1.49e+03
c	821	12	6.9	413	22	T18132	Maize XA17 acetohydro	1.49e+03
	822	12	6.9	450	1	Q03904	L1 amplified region;	1.49e+03
	823	12	6.9	455	38	T91258	Human papilloma virus	1.49e+03
	824	12	6.9	473	38	V12254	Entodinium caudatum c	1.49e+03
	825	12	6.9	494	28	T59835	Beta-casein gene frag	1.49e+03
c	826	12	6.9	505	39	T88056	Partial cDNA clone en	1.49e+03
c	827	12	6.9	552	40	V02986	Human BF245_1 interna	1.49e+03
	828	12	6.9	557	21	T11338	AFT-1 interacting pro	1.49e+03
	829	12	6.9	582	30	T67762	H. pylori transmembra	1.49e+03
	830	12	6.9	609	3	N50090	Sequence encoding the	1.49e+03
	831	12	6.9	621	1	Q06758	Sequence encoding bov	1.49e+03
c	832	12	6.9	660	32	T58531	Nucleotide sequence o	1.49e+03
c	833	12	6.9	681	40	V24843	H. pylori ORF hp4p113	1.49e+03
c	834	12	6.9	688	40	V10178	Stealth virus nucleic	1.49e+03
c	835	12	6.9	733	40	V04274	Secreted protein C195	1.49e+03
c	836	12	6.9	749	35	T75638	Canine herpesvirus ge	1.49e+03
c	837	12	6.9	808	4	N30106	Sequence on recombina	1.49e+03
	838	12	6.9	830	33	T62455	Lambda 5h-1 clone con	1.49e+03
c	839	12	6.9	841	26	T28529	S. pneumoniae detecti	1.49e+03
c	840	12	6.9	843	1	Q03879	Merozoite surface ant	1.49e+03
	841	12	6.9	852	38	T88140	E. coli AIDA-I autotr	1.49e+03
	842	12	6.9	861	37	T91798	Dihydropyrimidine deh	1.49e+03
c	843	12	6.9	874	4	N10010	Sequence of the Hif-2	1.49e+03
	844	12	6.9	879	31	T71207	Pyemotes tritici tox	1.49e+03
c	845	12	6.9	913	39	V02297	Human secreted protei	1.49e+03
	846	12	6.9	948	11	Q63491	Inhibitor of cyclin d	1.49e+03
c	847	12	6.9	958	39	T96064	Apoptosis associated	1.49e+03
c	848	12	6.9	963	4	N20092	Sequence of leukocyte	1.49e+03
c	849	12	6.9	964	35	T79676	BRCA2 cancer suscepti	1.49e+03
c	850	12	6.9	1005	1	N91784	DNA probe complementa	1.49e+03
c	851	12	6.9	1041	35	T79675	BRCA2 cancer suscepti	1.49e+03
c	852	12	6.9	1048	2	N60313	Equine IFN-alpha-2 fr	1.49e+03
c	853	12	6.9	1050	35	T79674	BRCA2 cancer suscepti	1.49e+03
	854	12	6.9	1063	39	V12466	Human oestrogen sulph	1.49e+03
	855	12	6.9	1067	18	T13033	Cotton fibre-specific	1.49e+03
	856	12	6.9	1086	30	T67941	H. pylori flagella-as	1.49e+03
c	857	12	6.9	1099	4	Q25831	Mutant human BCDF.	1.49e+03
	858	12	6.9	1104	1	N81409	Beta chain gene of DP	1.49e+03

	859	12	6.9	1107	40	V09071	Bovine cGB-PDE clone	1.49e+03
	860	12	6.9	1107	37	T73178	Mycoplasma arthritidi	1.49e+03
c	861	12	6.9	1109	4	N20099	Sequence of leukocyte	1.49e+03
c	862	12	6.9	1110	4	N20027	Human leukocyte inter	1.49e+03
c	863	12	6.9	1112	4	N20028	Human leukocyte inter	1.49e+03
c	864	12	6.9	1145	1	N80340	Sequence encoding Ag5	1.49e+03
c	865	12	6.9	1147	35	T79695	BRCA2 cancer suscepti	1.49e+03
	866	12	6.9	1151	39	T86621	Myxoma T1 gene that e	1.49e+03
c	867	12	6.9	1161	1	N81460	Sequence encoding hum	1.49e+03
	868	12	6.9	1201	17	T04854	Human cDNA encoding T	1.49e+03
	869	12	6.9	1202	28	T48438	Human TGF-beta-like c	1.49e+03
c	870	12	6.9	1236	38	V06459	Esex sequence used fo	1.49e+03
c	871	12	6.9	1250	2	N80316	Transcription control	1.49e+03
c	872	12	6.9	1255	40	V09078	Human C5 gene cDNA cl	1.49e+03
c	873	12	6.9	1264	12	Q67317	Sprague-Dawley rat P1	1.49e+03
	874	12	6.9	1360	1	N90702	Rpt-1 clone cDNA inse	1.49e+03
	875	12	6.9	1390	2	Q13751	Peptidylglycine alpha	1.49e+03
	876	12	6.9	1392	33	T80053	Rice actin 1 gene pro	1.49e+03
c	877	12	6.9	1401	38	V03981	T. litoralis ADPHK co	1.49e+03
c	878	12	6.9	1404	40	V03130	cDNA encoding rat vit	1.49e+03
c	879	12	6.9	1419	6	Q38026	CAD cDNA isolated fro	1.49e+03
c	880	12	6.9	1440	26	T28571	Bacterial antibiotic	1.49e+03
c	881	12	6.9	1440	19	Q79523	Morning Glory dihydro	1.49e+03
c	882	12	6.9	1470	10	Q55185	MS-Le1610 Vector.	1.49e+03
	883	12	6.9	1500	38	T76891	Arabidopsis floral me	1.49e+03
	884	12	6.9	1500	39	V02767	Arabidopsis floral me	1.49e+03
	885	12	6.9	1500	39	V06024	Arabidopsis floral me	1.49e+03
c	886	12	6.9	1593	32	T58563	DNA sequence highly e	1.49e+03
c	887	12	6.9	1612	2	Q10685	Chromosomal HindII fr	1.49e+03
c	888	12	6.9	1614	15	Q93135	Human prepro-BDNF cDN	1.49e+03
	889	12	6.9	1618	8	Q46958	Human cytokine synthe	1.49e+03
	890	12	6.9	1626	35	T88774	Human receptor protei	1.49e+03
	891	12	6.9	1627	8	Q47761	Antisense RNA complem	1.49e+03
	892	12	6.9	1695	39	V15090	Human sperm surface p	1.49e+03
c	893	12	6.9	1710	12	Q70883	Tomato spotted wilt v	1.49e+03
	894	12	6.9	1753	39	V04456	Replication defective	1.49e+03
c	895	12	6.9	1763	2	Q10025	Encodes a portion of	1.49e+03
c	896	12	6.9	1820	7	Q42760	Swine parvovirus B ge	1.49e+03
	897	12	6.9	1828	28	T48238	Neuronal nicotinic ac	1.49e+03
c	898	12	6.9	1923	13	Q77672	Sequence for controll	1.49e+03
c	899	12	6.9	1923	9	Q53305	Sake yeast alcohol ac	1.49e+03
c	900	12	6.9	1926	39	V00255	Rat Ret receptor extr	1.49e+03
	901	12	6.9	1945	8	Q47816	Sequence encoding 43k	1.49e+03
c	902	12	6.9	1953	29	T68316	cDNA encoding human h	1.49e+03
c	903	12	6.9	1969	39	V05125	Haemagglutinin gene s	1.49e+03
c	904	12	6.9	1975	38	T97341	Human protein kinase	1.49e+03
	905	12	6.9	1980	17	Q99227	Heparinase-III gene.	1.49e+03
	906	12	6.9	1985	8	Q51083	H. somnus lppA gene.	1.49e+03
c	907	12	6.9	2004	36	T74638	Glucan elicitor recep	1.49e+03
c	908	12	6.9	2022	36	T96728	Human adducin-like 70	1.49e+03
	909	12	6.9	2028	12	Q70885	Tomato spotted wilt v	1.49e+03

c	910	12	6.9	2078	24	T39930	Tilapia insulin gene.	1.49e+03
	911	12	6.9	2090	1	N90368	Fragment of tobacco P	1.49e+03
	912	12	6.9	2098	10	Q55320	Recombinant hyaluroni	1.49e+03
	913	12	6.9	2125	39	V15093	Mouse sperm surface p	1.49e+03
c	914	12	6.9	2152	12	Q70732	TATA-binding protein-	1.49e+03
c	915	12	6.9	2163	38	T95753	Arabidopsis SCARECROW	1.49e+03
c	916	12	6.9	2254	11	Q66167	Seven transmembrane r	1.49e+03
c	917	12	6.9	2310	20	T09020	Arabidopsis thaliana	1.49e+03
	918	12	6.9	2368	39	T94119	Human haematopoietic	1.49e+03
c	919	12	6.9	2379	38	V05640	Human ribonucleotide	1.49e+03
	920	12	6.9	2413	39	V14997	tvb polypeptide codin	1.49e+03
c	921	12	6.9	2434	10	Q56761	Chitinase derivative	1.49e+03
c	922	12	6.9	2452	11	Q68670	PMGAl.3 Mycoplasma ga	1.49e+03
c	923	12	6.9	2453	28	T51536	Mycobacterium gallise	1.49e+03
c	924	12	6.9	2490	37	T91358	Orf virus strain NZ-7	1.49e+03
c	925	12	6.9	2549	39	V15744	Rice delta-1-pyrrolin	1.49e+03
	926	12	6.9	2663	39	T94120	Alternatively spliced	1.49e+03
	927	12	6.9	2726	36	T97875	Watermellon seed nuce	1.49e+03
c	928	12	6.9	2760	14	Q85426	T4 DNA-polymerase wil	1.49e+03
c	929	12	6.9	2784	36	T65102	Luciferase gene under	1.49e+03
	930	12	6.9	2872	18	T08745	Polycistronic maize U	1.49e+03
	931	12	6.9	2884	14	Q84613	Rat AT2 receptor clon	1.49e+03
	932	12	6.9	2892	35	T92376	Cyclic-isomaltooligosa	1.49e+03
c	933	12	6.9	2903	24	T45666	Partial yeast gene, L	1.49e+03
	934	12	6.9	2982	39	V13266	A. thaliana trehalose	1.49e+03
c	935	12	6.9	3044	28	T64551	Tomato S-ribonuclease	1.49e+03
	936	12	6.9	3182	7	Q42590	Human LIF-R N-termina	1.49e+03
c	937	12	6.9	3291	24	T41405	Human REST protein co	1.49e+03
c	938	12	6.9	3328	24	T43277	DOA4 coding sequence.	1.49e+03
	939	12	6.9	3356	21	T12936	Plasmid pOXcop, compr	1.49e+03
c	940	12	6.9	3383	3	Q20269	Gene encoding AE-III	1.49e+03
	941	12	6.9	3440	40	V15268	Brassica campestris B	1.49e+03
c	942	12	6.9	3486	4	Q27102	Macrophage tissue tra	1.49e+03
c	943	12	6.9	3508	20	T27225	Bone morphogenetic pr	1.49e+03
	944	12	6.9	3527	35	T72171	Altheimer's disease D	1.49e+03
c	945	12	6.9	3600	36	T77330	Solanum tuberosum col	1.49e+03
	946	12	6.9	3607	38	T89990	Cryptosporidium parvu	1.49e+03
	947	12	6.9	3650	1	Q03807	gB gene of RB1B strai	1.49e+03
	948	12	6.9	3687	35	T75487	DNA for Bacillus stea	1.49e+03
c	949	12	6.9	3716	38	V05649	Rice SPK gene signal	1.49e+03
c	950	12	6.9	3786	39	T97229	Human ST receptor pro	1.49e+03
c	951	12	6.9	3786	38	T89300	Human ST receptor pro	1.49e+03
	952	12	6.9	3797	37	T80763	Staphylococcus aureus	1.49e+03
	953	12	6.9	3825	29	T13961	Beta-amylase coding s	1.49e+03
c	954	12	6.9	3889	38	T99948	Human N-acetylgalacto	1.49e+03
	955	12	6.9	3907	11	Q66820	AmEPV spheroidin gene	1.49e+03
c	956	12	6.9	4063	2	Q11540	Human Stem Cell Facto	1.49e+03
	957	12	6.9	4140	36	T77331	Solanum tuberosum col	1.49e+03
c	958	12	6.9	4200	36	T97316	Rabbit (156 kDa) dyst	1.49e+03
c	959	12	6.9	4200	10	Q45166	Sequence of precursor	1.49e+03
c	960	12	6.9	4597	12	Q70536	Human retinoblastoma	1.49e+03

	961	12	6.9	4796	38	T76897	Cauliflower floral me 1.49e+03
	962	12	6.9	5283	35	T91972	DNA encoding rat thro 1.49e+03
c	963	12	6.9	5283	35	T91972	DNA encoding rat thro 1.49e+03
c	964	12	6.9	5361	32	T78868	P. falciparum liver s 1.49e+03
c	965	12	6.9	5510	36	T97317	Human dystrophin-asso 1.49e+03
c	966	12	6.9	5537	34	T79736	Human cholesterol 7-a 1.49e+03
	967	12	6.9	5643	33	T80055	Rice actin 1 gene. 1.49e+03
c	968	12	6.9	5648	24	T41406	Human REST protein cD 1.49e+03
	969	12	6.9	5837	3	Q14461	Polymerase gene. 1.49e+03
	970	12	6.9	6464	8	Q48770	Arylamine N-acetyl-tr 1.49e+03
c	971	12	6.9	6471	18	T06306	Tomato pathogen resis 1.49e+03
	972	12	6.9	6479	39	V05172	Sequence of the spind 1.49e+03
c	973	12	6.9	7011	40	V20464	Human L-myc oncogene. 1.49e+03
c	974	12	6.9	7100	1	Q03360	Complete cDNA of huma 1.49e+03
	975	12	6.9	7351	33	T47529	Canarypox virus DNA c 1.49e+03
	976	12	6.9	7394	11	Q66631	Left-half vector M13E 1.49e+03
	977	12	6.9	7409	4	Q24173	M13IXD03. 1.49e+03
	978	12	6.9	7451	40	V05856	Complete sequence of 1.49e+03
c	979	12	6.9	7587	40	V15703	Zucchini ACC synthase 1.49e+03
c	980	12	6.9	7904	35	T94724	Human papillomavirus 1.49e+03
c	981	12	6.9	7904	35	T94723	Human papillomavirus 1.49e+03
c	982	12	6.9	8710	17	Q89650	pSC11 FIPV E1. 1.49e+03
c	983	12	6.9	8826	35	T74989	Saccharomyces cerevis 1.49e+03
	984	12	6.9	9714	39	T73135	Porcine transmissible 1.49e+03
c	985	12	6.9	9714	39	T73135	Porcine transmissible 1.49e+03
c	986	12	6.9	9919	40	V04631	Arabidopsis thaliana 1.49e+03
	987	12	6.9	9955	5	Q31625	mGP1-2 containing T7 1.49e+03
	988	12	6.9	10079	35	T92317	Human EC-SOD gene. 1.49e+03
	989	12	6.9	10670	38	T99449	Staphylococcus carnos 1.49e+03
	990	12	6.9	13875	4	Q20240	Human centromeric seq 1.49e+03
	991	12	6.9	14237	39	V05384	Human cathepsin K gen 1.49e+03
c	992	12	6.9	14507	32	T73568	Expression augmenting 1.49e+03
c	993	12	6.9	17410	40	V15205	Human osteogenic prot 1.49e+03
c	994	12	6.9	28994	39	V15826	Genomic DNA for inter 1.49e+03
c	995	12	6.9	29392	40	V15422	Mouse poly Ig recepto 1.49e+03
	996	12	6.9	29544	17	T12460	Ovine adenovirus geno 1.49e+03
c	997	12	6.9	31122	7	Q40706	Bacillus subtilis srf 1.49e+03
c	998	12	6.9	35100	40	V20441	Human c-fms oncogene. 1.49e+03
c	999	12	6.9	53526	40	T94101	Human PKD1 gene. 1.49e+03
c	1000	12	6.9	53577	40	T94108	Human PKD1 locus betw 1.49e+03

ALIGNMENTS

```
RESULT    1
ID   T20219 standard; cDNA to mRNA; 174 BP.
AC   T20219;
DT   24-JUL-1996  (first entry)
DE   Human gene signature HUMGS01365.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
```



```
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 589; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   174 BP;    70 A;    20 C;    30 G;    54 T;

  Query Match          100.0%;  Score 174;  DB 19;  Length 174;
  Best Local Similarity 100.0%;  Pred. No. 1.11e-126;
  Matches   174;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db       1 gatctgagcaggcaaagctcaaaagagagtttggaggttaaaaataatttatttttgcag 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy       1 GATCTGAGCAGGCAAAGCTCAAAAGAGAGTTTGGAGGTTAAAAATAATTTATTTTTGCAG 60

Db      61 tagtgtgctttgaaatgtgtaaatcttatttctaatgtatacaaccacatttcacataaa 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      61 TAGTGTGCTTTGAAATGTGTAAATCTTATTTCTAATGTATACAACCACATTTCACATAAA 120

Db     121 aatatgcaatttatatgccagataaaaataaaacaagtgaatttgcaagtgaaa 174
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     121 AATATGCAATTTATATGCCAGATAAAAATAAAACAAGTGAATTTGCAAGTGAAA 174


RESULT    2
ID   T41371 standard; DNA; 20 BP.
```



```
AC   T41371;
DT   04-DEC-1996  (first entry)
DE   Human gene signature HUMGS01365-derived sense primer.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; primer; PCR; amplification;
KW   polymerase chain reaction; ss.
OS   Synthetic.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Example 7; Fig 10; 2245pp; Japanese.
CC   Primers T41001-T41382 are derived from novel human gene signature (GS)
CC   sequences which did not match with sequences deposited in Genbank release
CC   76.  The GS sequences (T19001-T26837) were obtained from 3'-directed cDNA
CC   libraries prepared from various human tissues; synthesis of cDNA was
CC   initiated from the 3'-end of mRNA by using poly(T) as the sole primer.
CC   Each library is constructed so as to reflect accurately the relative
CC   abundance of different mRNAs in the particular tissue from which it was
CC   derived.  The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS sequences)
CC   as a means of diagnosing abnormal cell function or for recognising
CC   different cell types. The primers T41371-2 amplify clone pm2908 which
CC   comprises the GS HUMGS001365 (T20365).  This amplification reaction gave
CC   no product.
SQ   Sequence   20 BP;    5 A;    2 C;    6 G;    7 T;

  Query Match          11.5%;  Score 20;  DB 23;  Length 20;
  Best Local Similarity 100.0%;  Pred. No. 4.30e-02;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db       1 cagtagtgtgctttgaaatg 20
           ||||||||||||||||||||
Qy      58 CAGTAGTGTGCTTTGAAATG 77


RESULT    3
ID   T41372 standard; DNA; 20 BP.
AC   T41372;
DT   04-DEC-1996  (first entry)
DE   Human gene signature HUMGS01365-derived anti-sense primer.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
```



```
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; primer; PCR; amplification;
KW   polymerase chain reaction; ss.
OS   Synthetic.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Example 7; Fig 10; 2245pp; Japanese.
CC   Primers T41001-T41382 are derived from novel human gene signature (GS)
CC   sequences which did not match with sequences deposited in Genbank release
CC   76.  The GS sequences (T19001-T26837) were obtained from 3'-directed cDNA
CC   libraries prepared from various human tissues; synthesis of cDNA was
CC   initiated from the 3'-end of mRNA by using poly(T) as the sole primer.
CC   Each library is constructed so as to reflect accurately the relative
CC   abundance of different mRNAs in the particular tissue from which it was
CC   derived.  The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS sequences)
CC   as a means of diagnosing abnormal cell function or for recognising
CC   different cell types. The primers T41371-2 amplify clone pm2908 which
CC   comprises the GS HUMGS001365 (T20365).  This amplification reaction gave
CC   no product.
SQ   Sequence   20 BP;    4 A;    0 C;    6 G;    10 T;

  Query Match          11.5%;  Score 20;  DB 23;  Length 20;
  Best Local Similarity 100.0%;  Pred. No. 4.30e-02;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db       1 tttatgtgaaatgtggttgt 20
           ||||||||||||||||||||
Cp     120 TTTATGTGAAATGTGGTTGT 101


RESULT    4
ID   T24247 standard; cDNA to mRNA; 168 BP.
AC   T24247;
DT   16-OCT-1996  (first entry)
DE   Human gene signature HUMGS06262.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
```

```
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 1565; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues: synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   168 BP;    57 A;    18 C;    31 G;    53 T;

  Query Match           9.8%;  Score 17;  DB 22;  Length 168;
  Best Local Similarity 100.0%;  Pred. No. 2.95e+00;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     65 gtgaaatgtggttgtat 81
          |||||||||||||||||
Cp    115 GTGAAATGTGGTTGTAT 99


RESULT    5
ID   N90852 standard; DNA; 702 BP.
AC   N90852;
DT   15-JUL-1990 (first entry)
DE   Intron 3 from human follistatin (hFS) gene
KW   Human follistatin genomic DNA; intron 3;
KW   porcine follistatin precursor probe.
OS   Homo sapiens.
PN   WO8901945-A.
PD   09-MAR-1989.
PF   26-AUG-1988; U02971.
PR   26-JUL-1988; US-224432, US-090635.
PA   (SALK) Salk Inst for Biol Stud.
```



```
PI   Ling NCK, Ueno N, Shimasaki S, Ying SY, Guillemin RCL;
DR   WPI; 89-085524/11.
PT   Monomermic mammalian follistatin protein -
PT   isolated from porcine follicular fluid and prepd. by recombinant
PT   techniques, used esp. as male contraceptive
PS   Disclosure;  ; 34pp; English.
CC   The hFS gene was cloned and sequenced from a human genomic library and
CC   3 positives were obtained by hybridisation with a cDNA probe encoding
CC   the first 317 amino acids of the porcine FS (pFS) precursor. Restriction
CC   mapping and nucleotide sequence analyses reveal that one clone, which is
CC   19.1 kb in length, encodes nearly the whole human FS gene
CC   containing 5 exons. Another 15 kb clone contains only the last exon (5).
CC   It is postulated that the genomic organisation of human FS is composed
CC   of a DNA sequence that is more than 5kb in length with 6 exons
CC   interrupted by 5 introns (see n90641). This is the sequence of the third
CC   intron. The first 4 introns possess the consensus 5' - GT ... AG - 3'
CC   donor and acceptor splicing sequences.
SQ   Sequence   702 BP;   235 A;   103 C;   138 G;   225 T;   1 Others;

  Query Match           9.8%;  Score 17;  DB 1;  Length 702;
  Best Local Similarity 100.0%;  Pred. No. 2.95e+00;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    601 aaaataaattattttta 617
          |||||||||||||||||
Cp     55 AAAATAAATTATTTTTA 39


RESULT    6
ID   N80748 standard; DNA; 2207 BP.
AC   N80748;
DT   17-SEP-1990  (first entry)
DE   Sequence of a polygalacturonase genomic  clone.
KW   Polygalacturonase; genomic  clone; tomato plants;
KW   transcription initiation region; 2A11.
PN   WO8809334-A.
PD   01-DEC-1988.
PF   16-MAY-1988; U01811.
PR   26-MAY-1987; US-054369.
PR   5-MAR-1988; US-168190.
PA   (CALG-) Calgene Inc.
PI   Houck CM, Pear JR;
DR   WPI; 88-353919/49.
PT   Fruit specific DNA constructs to modify fruit phenotype -
PT   comprising transcription initiation and termination regions and
PT   desired DNA sequence, eg poly galacturonase gene.
PS   Disclosure; pp; English.
CC   The polygalacturonase genomic clone may be inserted into a fruit specific
CC   DNA construct containing the transcription initiation region 2A11.
CC   The constructs allow modification of plant phenotypes during fruit
CC   maturation and ripening.
```



```
SQ   Sequence   2207 BP;   870 A;   283 C;   270 G;   784 T;

  Query Match           9.8%;  Score 17;  DB 1;  Length 2207;
  Best Local Similarity 100.0%;  Pred. No. 2.95e+00;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1090 aaaaataaattattttt 1106
          |||||||||||||||||
Cp     56 AAAAATAAATTATTTTT 40


RESULT    7
ID   T14077 standard; cDNA; 3957 BP.
AC   T14077;
DT   11-JUN-1996  (first entry)
DE   Human dihydropyrimidine dehydrogenase cDNA.
KW   Dihydropyrimidine dehydrogenase; DPD; DPYD gene; probe; deficiency;
KW   5-fluorouracil; cytostatic; cancer; selectable marker; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             88..3165
FT                   /*tag=  a
FT   primer_bind     complement (36..55)
FT                   /*tag=  b
FT                   /note= "primer RTF1"
FT   primer_bind     1536..1558
FT                   /*tag=  c
FT                   /note= "primer RTR1"
FT   primer_bind     complement (1539..1558)
FT                   /*tag=  d
FT                   /note= "primer H13"
FT   primer_bind     2426..2445
FT                   /*tag=  e
FT                   /note= "primer RTR4"
FT   primer_bind     complement (2424..2447)
FT                   /*tag=  f
FT                   /note= "primer RTR5"
FT   primer_bind     3320..3343
FT                   /*tag=  g
FT                   /note= "primer RTR5"
PN   WO9608568-A2.
PD   21-MAR-1996.
PF   07-SEP-1995; U12016.
PR   12-SEP-1994; US-304309.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI   Fernandez-Salguero P,  Gonzales FJ;
DR   WPI; 96-179943/18.
DR   P-PSDB; R91420.
PT   Pig and human di:hydro:pyrimidine dehydrogenase (DPD) genes and
PT   probes  - useful for detection of DPD deficiencies and
PT   identification of humans at risk of toxic reaction to
```



```
PT   5-fluoro:uracil anti-cancer treatment
PS   Claim 1; Page 43-48; 78pp; English.
CC   Human dihydropyrmidine dehydrogenase (DPD) cDNA (T14077) was
CC   isolated as 3 overlapping fragments obtd. from human liver
CC   cDNA libraries screened using pig DPD cDNA (see also T14077).
CC   The DPD gene, DPYD, is located at chromosome 1p22.  The cDNA
CC   can be inserted into a vector and used for prodn. of DPD (R91420)
CC   in transformed host cells, or used as a selectable marker that
CC   functions in prokaryotic and eukaryotic cells.  It can also be
CC   utilised as a probe to detect DPD deficiency in an individual,
CC   i.e. to identify risk of toxic reaction to 4-fluorouracil.
SQ   Sequence   3957 BP;   1156 A;   787 C;   897 G;   1117 T;

  Query Match           9.8%;  Score 17;  DB 18;  Length 3957;
  Best Local Similarity 100.0%;  Pred. No. 2.95e+00;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    504 atgcaatttatatgcca 520
          |||||||||||||||||
Qy    124 ATGCAATTTATATGCCA 140


RESULT    8
ID   T03133 standard; DNA; 4368 BP.
AC   T03133;
DT   04-JUN-1996  (first entry)
DE   Human dihydropyrimidine dehydrogenase gene.
KW   Bovine; liver; human; dihydropyrimidine dehydrogenase; DPD;
KW   pyrimidine catabolism; 5,6-dihydropyrimidine; pyrimidine analogue;
KW   fluoropyrimidine; anticancer drug; 5-fluorouracil; FUra; cancer;
KW   frameshift mutation; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             49..3126
FT                   /*tag=  a
FT                   /product= DPD
PN   WO9528489-A1.
PD   26-OCT-1995.
PF   13-APR-1995; U04567.
PR   13-APR-1994; US-227357.
PA   (UABR-) UAB RES FOUND.
PI   Cheng X,  Diasio RB,  Johnson M,  Lu Z,  Zhang R;
DR   WPI; 95-373803/48.
DR   P-PSDB; R86001.
PT   Novel di:hydro:pyrimidine dehydrogenase gene - used to optimise
PT   5-fluoro:uracil doses given to cancer patients
PS   Claim 10; Page 140-56; 207pp; English.
CC   This sequence represents the DNA sequence encoding human
CC   dihydropyrimidine dehydrogenase (DPD).  DPD catalyses the initial
CC   and rate limiting step in pyrimidine catabolism, the reduction of
CC   pyrimidines to 5,6-dihydropyrimidines.  DPD is a complex enzyme
```

```
CC   consisting of two identical subunits, containing FMN, FAD and iron-
CC   sulphur centers, and utilising NADPH as a cofactor.  DPD has also been
CC   shown to catalyse the reduction of various pyrimidine analogues
CC   including the fluoropyrimidine anticancer drug 5-fluorouracil (FUra).
CC   Up to 85% of administered FUra may be catabolised by DPD, and it
CC   therefore governs the effectiveness of FUra as an anticancer drug.
CC   DPD genes or fragments of them may be used in the detection of DPD in a
CC   sample, esp. isolated from a cancer patient.  According to the amount
CC   of DPD detected, a therapeutically effective amount of FUra may be
CC   determined and administered.  DPD deficiency, leading to life-
CC   threatening toxicity on exposure to FUra, in a human caused by a
CC   frameshift mutation may be determined by means of a molecular biological
CC   assay to detect the deletion of an A residue at codon 318 within the
CC   DPD-coding region.
SQ   Sequence   4368 BP;    1325 A;    829 C;    932 G;    1282 T;

  Query Match             9.8%;  Score 17;  DB 18;  Length 4368;
  Best Local Similarity 100.0%;  Pred. No. 2.95e+00;
  Matches    17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Db    465 atgcaatttatatgcca 481
          |||||||||||||||||
Qy    124 ATGCAATTTATATGCCA 140


RESULT    9
ID   T03143 standard; cDNA; 4369 BP.
AC   T03143;
DT   05-JUN-1996  (first entry)
DE   Human dihydropyrimidine dehydrogenase cDNA.
KW   Bovine; liver; human; dihydropyrimidine dehydrogenase; DPD;
KW   pyrimidine catabolism; 5,6-dihydropyrimidine; pyrimidine analogue;
KW   fluoropyrimidine; anticancer drug; 5-fluorouracil; FUra; cancer;
KW   frameshift mutation; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             49..3125
FT                   /*tag=  a
FT                   /product= Human lymphocyte DPD
FT   primer_bind     complement (36..60)
FT                   /*tag=  b
FT   primer_bind     complement (716..743)
FT                   /*tag=  c
FT   primer_bind     1260..1288
FT                   /*tag=  d
FT   primer_bind     3117..3139
FT                   /*tag=  e
FT   mutation        1000
FT                   /*tag=  f
FT                   /note= "Position of single base deletion in DPD
FT                   deficient patients"
```



```
PN   WO9528489-A1.
PD   26-OCT-1995.
PF   13-APR-1995; U04567.
PR   13-APR-1994; US-227357.
PA   (UABR-) UAB RES FOUND.
PI   Cheng X,  Diasio RB,  Johnson M,  Lu Z,  Zhang R;
DR   WPI; 95-373803/48.
DR   P-PSDB; R86001.
PT   Novel di:hydro:pyrimidine dehydrogenase gene - used to optimise
PT   5-fluoro:uracil doses given to cancer patients
PS   Disclosure; Fig 7; 207pp; English.
CC   This sequence represents the cDNA sequence encoding human dihydro-
CC   pyrimidine dehydrogenase (DPD).  This sequence was amplified using
CC   the primer sequences given in T03144-47.  DPD catalyses the initial
CC   and rate limiting step in pyrimidine catabolism, the reduction of
CC   pyrimidines to 5,6-dihydropyrimidines.  DPD is a complex enzyme
CC   consisting of two identical subunits, containing FMN, FAD and iron-
CC   sulphur centers, and utilising NADPH as a cofactor.  DPD has also been
CC   shown to catalyse the reduction of various pyrimidine analogues
CC   including the fluoropyrimidine anticancer drug 5-fluorouracil (FUra).
CC   Up to 85% of administered FUra may be catabolised by DPD, and it
CC   therefore governs the effectiveness of FUra as an anticancer drug.
CC   DPD genes or fragments of them may be used in the detection of DPD in a
CC   sample, esp. isolated from a cancer patient.  According to the amount
CC   of DPD detected, a therapeutically effective amount of FUra may be
CC   determined and administered.  DPD deficiency, leading to life-
CC   threatening toxicity on exposure to FUra, in a human caused by a
CC   frameshift mutation may be determined by means of a molecular biological
CC   assay to detect the deletion of an A residue at codon 318 within the
CC   DPD-coding region.
SQ   Sequence   4369 BP;    1325 A;    829 C;    932 G;    1283 T;

  Query Match             9.8%;  Score 17;  DB 18;  Length 4369;
  Best Local Similarity 100.0%;  Pred. No. 2.95e+00;
  Matches    17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Db    465 atgcaatttatatgcca 481
          |||||||||||||||||
Qy    124 ATGCAATTTATATGCCA 140


RESULT   10
ID   N70099 standard; DNA; 609 BP.
AC   N70099;
DT   16-APR-1991  (first entry)
DE   Sequence of DNA fragment containing coding for a modified
DE   interferon and the terminator sequence from yeast alcohol
KW   dehydrogenase
KW   Plasmid SD 922; yeast expression vector; ds.
FH   Key             Location/Qualifiers
FT   misc_feature    1..4
```



```
FT                   /*tag=  b
FT                   /label= Sticky end
PN   EP-214638-A.
PD   18-MAR-1987.
PF   08-SEP-1986; 112385.
PR   09-SEP-1985; US-774219.
PA   (SEAR ) SEARLE G D & CO.
PI   Carter BLA, Doel SM, Goodey AR;
DR   WPI; 87-081158/12.
PT   New plasmid SD922 of molecular wt. 9 KB - useful for expression
PT   of exogenous genes in yeast for prodn. of useful proteins
PS   Disclosure; Fig 3; 16pp; English.
CC   N70100 is used for the construction of a plasmid for expressing a
CC   modified interferon in yeast. Strain SLP 1B transformed with SD 922
CC   containing DNA coding for hybrid interferon (N70099) grew much more
CC   quickly than the equivalent construct containing pyruvate kinase
CC   promoter - 3 days versus 10 days respectively.
SQ   Sequence   609 BP;    189 A;    112 C;    129 G;    179 T;

  Query Match             9.2%;  Score 16;  DB 2;  Length 609;
  Best Local Similarity 100.0%;  Pred. No. 1.13e+01;
  Matches    16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Db    542 aaaataaattattttt 557
          |||||||||||||||||
Cp     55 AAAATAAATTATTTTT 40


RESULT   11
ID   N40108 standard; cDNA; 668 BP.
AC   N40108;
DT   12-FEB-1992  (first entry)
DE   Sequence of the cDNA insert of the recombinant plasmid CG-pBR
DE   322/HLycIFN-betal.
KW   Yeast expression vector; promoter; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             1..468
FT                   /*tag=  a
PN   EP-100561-A.
PD   15-FEB-1984.
PF   08-AUG-1983; 107804.
PR   09-AUG-1982; GB-022883.
PR   31-DEC-1982; GB-037026.
PR   02-JUN-1983; GB-015145.
PR   14-JUL-1983; GB-019099.
PA   (CIBA ) CIBA GEIGY AG.
PI   Hinnen A, Meyhack B, Meyer F;
DR   WPI; 84-044243/08.
DR   P-PSDB; P40124.
PT   Acid phosphatase promoting DNA fragment - for expressing
```



```
PT   peptide(s) in yeasts which are more easily cultured than e.coli
PS   Example; Fig 11; 166pp; English.
CC   The inventors claim: (A) DNA fragment consisting of yeast acid
CC   phosphatase promoter and flanking sequences or its mutants which
CC   retain the promoter function. (B) Hybrid vector consisting of a
CC   yeast acid phosphatase promoter and a yeast or non-yeast polypeptide
CC   coding region controlled by the promoter. (C) A yeast transformed by
CC   the vector. (D) Prodn. of a yeast or non-yeast polypeptide or its
CC   deriv. by culturing the transformed yeast. The polypeptides are
CC   hormones, antiviral and anticancer peptides, enzymes and interferon.
SQ   Sequence   668 BP;    208 A;    125 C;    147 G;    188 T;

  Query Match             9.2%;  Score 16;  DB 3;  Length 668;
  Best Local Similarity 100.0%;  Pred. No. 1.13e+01;
  Matches    16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Db    641 aaaataaattattttt 656
          |||||||||||||||||
Cp     55 AAAATAAATTATTTTT 40


RESULT   12
ID   N30101 standard; DNA; 668 BP.
AC   N30101;
DT   14-JUN-1992  (first entry)
DE   Sequence on recombinant plasmid CG-pBR 322/HL gamma cIFN-betal
DE   encoding polypeptide with human lymphoblastoid interferon activity.
KW   Immunomodulator; antiviral agent; antitumor; anticancer; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             1..465
FT                   /*tag=  a
PN   EP--76489-A.
PD   13-APR-1983.
PF   01-SEP-1982; 024871.
PR   03-OCT-1981; GB-029937.
PR   26-MAR-1982; GB-008988.
PR   01-SEP-1982; GB-024871.
PA   (CIBA ) CIBA GEIGY AG.
PI   Meyer F;
DR   WPI; 83-37194K/16.
DR   P-PSDB; P30180.
PT   Human interferon polypeptide - produced by host cells, esp.
PT   E.coli transformed with recombinant DNA contg. sequence
PT   obtainable from human lymphoblastoid cells
PS   Claim 45; Page 101 and Fig 5; 142pp; German.
CC   The inventors claim DNA encoding a polypeptide with human
CC   lymphoblastoid interferon activity, hosts which have been
CC   transformed with the DNA and the polypeptide.  Variants of the DNA
CC   sequence are also claimed.  The polypeptides are useful as immuno-
CC   modulators, esp. as antiviral, antitumour and anticancer agents.
```

```
SQ   Sequence   668 BP;    208 A;    124 C;    147 G;    189 T;

  Query Match           9.2%;  Score 16;  DB 4;  Length 668;
  Best Local Similarity 100.0%;  Pred. No. 1.13e+01;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    641 aaaataaattattttt 656
          ||||||||||||||||
Cp     55 AAAATAAATTATTTTT 40


RESULT   13
ID   N20031 standard; DNA; 761 BP.
AC   N20031;
DT   11-AUG-1992  (first entry)
DE   DNA sequence encoding human fibroblast interferon.
KW   Interferon; virucide; antitumor; ss.
OS   Escherichia coli, Bacillus subtilis, Saccharomyces cerevisiae.
FH   Key             Location/Qualifiers
FT   cds             1..562
FT                   /*tag= a
FT                   /label= human interferon
PN   EP--48970-A.
PD   07-APR-1982.
PF   24-SEP-1981; 107621.
PR   25-SEP-1980; US-190799.
PR   11-AUG-1981; US-291892.
PA   (GENE-) GENENTECH INC.
PI   Crea R, Goeddel DVN;
DR   WPI; 82-28974E/15  (28974E).
DR   P-PSDB; P20026.
PT   Microbially produced mature human fibroblast interferon - obtd.
PT   by using recombinant DNA coding for amino acid interferon
PT   sequences.
PS   Disclosure; Fig 3; 40pp; English.
CC   Human interferon is expressed in large amounts in E. coli, B.
CC   subtilis or S. cerevisiae and used as a virucide or antitumour
CC   agent.
SQ   Sequence   761 BP;    232 A;    152 C;    159 G;    218 T;

  Query Match           9.2%;  Score 16;  DB 4;  Length 761;
  Best Local Similarity 100.0%;  Pred. No. 1.13e+01;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    737 aaaataaattattttt 752
          ||||||||||||||||
Cp     55 AAAATAAATTATTTTT 40


RESULT   14
ID   N10009 standard; cDNA; 777 BP.
```



```
AC   N10009;
DT   19-OCT-1992  (first entry)
DE   Sequence encoding fibroblast interferon and its putative signal
DE   peptide in plasmid TpIF 319-13.
KW   Interferon; antiviral agent; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   signal_peptide  7..69
FT                   /*tag= a
FT   mat_peptide     70..567
FT                   /*tag= b
PN   EP--28033-A.
PD   06-MAY-1981.
PF   30-OCT-1980; 106685.
PR   30-OCT-1979; JP-139289.
PR   19-MAR-1980; JP-033931.
PA   (NICA-) JAPANESE CANCER RES.
PI   Sugano H, Muramatsu M, Taniguchi T;
DR   WPI; 81-34772D/20.
DR   P-PSDB; P10016.
PT   DNA coding for poly:peptide with interferon activity - useful in
PT   prepn. of human interferon in large amounts
PS   Example; Table 5, Page 18; 22pp; English.
CC   The inventors claim recombinant plasmid TpIF 319-13 which contains
CC   cDNA (N10009) prepd. from mRNA extracted from human fibroblasts.
CC   TpIF 319-13 transformed in E. coli has been deposited under
CC   accession number ATCC 31712, which is claimed.
SQ   Sequence   777 BP;    240 A;    155 C;    162 G;    220 T;

  Query Match           9.2%;  Score 16;  DB 4;  Length 777;
  Best Local Similarity 100.0%;  Pred. No. 1.13e+01;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    743 aaaataaattattttt 758
          ||||||||||||||||
Cp     55 AAAATAAATTATTTTT 40


RESULT   15
ID   Q68695 standard; DNA; 777 BP.
AC   Q68695;
DT   17-FEB-1995  (first entry)
DE   Human fibroblast interferon beta-1 gene.
KW   Fibroblast; interferon; antiviral; anti-tumour; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             7..570
FT                   /*tag= a
FT                   /product= Interferon beta 1.
PN   US5326859-A.
PD   05-JUL-1994.
```



```
PF   27-OCT-1980; 201359.
PR   27-OCT-1980; US-201359.
PA   (NICA-) JAPANESE FOUND CANCER RES.
PI   Muramatsu M,  Sugano H,  Taniguchi T;
DR   WPI; 94-217099/26.
DR   P-PSDB; R56987.
PT   New DNA encoding human fibroblast interferon Beta 1 - as
PT   precursor or mature protein, useful as antiviral and antitumour
PT   agent, allowing large scale prepn of recombinant protein
PS   Claim 3; Columns 11-13; 11pp; English.
CC   This DNA sequence can be introduced to host cells e.g. E. coli and
CC   then expressed in those cells, allowing large scale production of
CC   recombinant human fibroblast interferon beta 1 which exhibits
CC   antiviral and anti-tumour activity.
SQ   Sequence   777 BP;    240 A;    155 C;    162 G;    220 T;

  Query Match           9.2%;  Score 16;  DB 11;  Length 777;
  Best Local Similarity 100.0%;  Pred. No. 1.13e+01;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    743 aaaataaattattttt 758
          ||||||||||||||||
Cp     55 AAAATAAATTATTTTT 40


RESULT   16
ID   N10041 standard; cDNA; 836 BP.
AC   N10041;
DT   12-OCT-1992  (first entry)
DE   Sequence encoding fibroblast interferon-like protein.
KW   Interferon; antiviral agent; ss.
OS   Homo sapiens.
PN   BE-887397-A.
PD   01-JUN-1981.
PF   28-AUG-1980; 105113.
PR   06-FEB-1980; GB-003947.
PR   28-FEB-1980; GB-006712.
PR   17-APR-1980; GB-012666.
PR   24-APR-1980; GB-013592.
PR   12-MAY-1980; GB-015646.
PR   18-NOV-1980; GB-036951.
PA   (SEAR ) SEARLE G D & CO.
DR   WPI; 81-44110D/25.
PT   Gene for expressing protein similar to human interferon - derived
PT   from plasmid recombined materials, and modified bacterial cells
PS   Claim 4; Page 58-59; 71pp; French.
CC   N10040 is cDNA prepd. from mRNA isolated from induced fibroblast
CC   cells using oligo-dT initiator and reverse transcriptase.
CC   Double-stranded DNA is then made using reverse transcriptase of E.
CC   coli DNA polymerase, pref. together with an initator.  The genes
CC   produced are identified after cloning by using an initiator to
```



```
CC   select colonies.
SQ   Sequence   836 BP;    247 A;    173 C;    177 G;    239 T;

  Query Match           9.2%;  Score 16;  DB 4;  Length 836;
  Best Local Similarity 100.0%;  Pred. No. 1.13e+01;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    809 aaaataaattattttt 824
          ||||||||||||||||
Cp     55 AAAATAAATTATTTTT 40


RESULT   17
ID   N10040 standard; cDNA; 899 BP.
AC   N10040;
DT   12-OCT-1992  (first entry)
DE   Sequence encoding fibroblast interferon-like protein.
KW   Interferon; antiviral agent; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   misc_feature    1..869
FT                   /*tag= a
FT                   /note= "claim 1"
PN   BE-887397-A.
PD   01-JUN-1981.
PF   28-AUG-1980; 105113.
PR   06-FEB-1980; GB-003947.
PR   28-FEB-1980; GB-006712.
PR   17-APR-1980; GB-012666.
PR   24-APR-1980; GB-013592.
PR   12-MAY-1980; GB-015646.
PR   18-NOV-1980; GB-036951.
PA   (SEAR ) SEARLE G D & CO.
DR   WPI; 81-44110D/25.
PT   Gene for expressing protein similar to human interferon - derived
PT   from plasmid recombined materials, and modified bacterial cells
PS   Claim 2; Page 57; 71pp; French.
CC   N10040 is cDNA prepd. from mRNA isolated from induced fibroblast
CC   cells using oligo-dT initiator and reverse transcriptase.
CC   Double-stranded DNA is then made using reverse transcriptase of E.
CC   coli DNA polymerase, pref. together with an initator.  The genes
CC   produced are identified after cloning by using an initiator to
CC   select colonies.
SQ   Sequence   899 BP;    269 A;    184 C;    189 G;    257 T;

  Query Match           9.2%;  Score 16;  DB 4;  Length 899;
  Best Local Similarity 100.0%;  Pred. No. 1.13e+01;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    833 aaaataaattattttt 848
          ||||||||||||||||
```

```
Cp     55 AAAATAAATTATTTTT 40


RESULT   18
ID   N10038 standard; DNA; 926 BP.
AC   N10038;
DT   14-AUG-1992  (first entry)
DE   Sequence encoding human fibroblast interferon which is derived from
DE   a combination of the data from at least two of the plasmids
DE   pHFIF1-13.
KW   Viral infection; therapy; cancer; tumour; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   signal_peptide  64..126
FT                   /*tag=  a
FT   mat_peptide     127..624
FT                   /*tag=  b
FT                   /note= "claimed; CDS for sig-peptide and
FT                   mat-peptide combined are also claimed"
PN   EP--41313-A.
PD   09-DEC-1981.
PF   01-APR-1981; 301414.
PR   03-APR-1980; GB-011306.
PR   06-JUN-1980; GB-018701.
PA   (BIOG-) BIOGEN NV.
PI   Fiers WC;
DR   WPI; 81-93390D/51.
DR   P-PSDB; P10043.
PT   DNA sequences, recombinant DNA molecules transformed hosts etc. -
PT   for prodn. of antiviral and anticancer polypeptide(s) (NO
PT   26.10.81)
PS   Example; Fig 4; 118pp; English.
CC   The inventors claim a DNA sequence consisting of the DNA inserts of
CC   G-pBR322(Pst)/HFIF1, /HFIF3, /HFIF6 or /HFIF7 and DNA sequences
CC   which hybridise any of these 4 inserts.  A polypeptide or its
CC   fragments and derivs. showing an immunological or biological
CC   activity of human fibroblast interferon produced by the transformed
CC   host is also claimed.  The polypeptide is useful for treating viral
CC   infections, cancers or tumours in humans, or for treating bovine
CC   viral infections.
SQ   Sequence   926 BP;    286 A;    187 C;    196 G;    257 T;

  Query Match           9.2%;  Score 16;  DB 4;  Length 926;
  Best Local Similarity 100.0%;  Pred. No. 1.13e+01;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    890 aaaataaattattttt 905
          ||||||||||||||||
Cp     55 AAAATAAATTATTTTT 40
```



```
RESULT   19
ID   Q20695 standard; cDNA; 1542 BP.
AC   Q20695;
DT   24-APR-1992  (first entry)
DE   Mouse OSP.1.
KW   MK; calvarial cell; probe; diagnosis; osteoblast;
KW   cranial nerve cell; ss.
OS   Mus musculus.
FH   Key             Location/Qualifiers
FT   cds             229..735
FT                   /*tag=  a
FT                   /label= I-I
PN   WO9200324-A.
PD   09-JAN-1992.
PF   27-JUN-1991; J00871.
PR   29-JUN-1990; JP-169824.
PR   28-SEP-1990; JP-256810.
PA   (FARH ) HOECHST JAPAN LTD.
PI   Hashimoto T, Tezuka K, Kumegawa S, Takagi C;
DR   WPI; 92-041516/05.
DR   P-PSDB; R20238.
PT   Protein from mouse calvarial cells - differentiates growth of
PT   osteoblast(s) and cranial nerve cells, for treating and
PT   diagnosing osteoporosis and dementia
PS   Disclosure; Fig 3(1-5); 43pp; Japanese.
CC   A cDNA library was constructed from mouse calvarial cell line
CC   MC3T3E1.  Differential screening in NIH3T3 cells (ATCC CRL-1658) and
CC   cloning, then insertion into vector pUC118 gave plasmid pMCO31 which
CC   contains the DNA coding for mouse OSP-1 (Q20695).  This plasmid was
CC   used as probe to clone by plaque hybridisation from a human cDNA
CC   library to give a phage clone, HBR1.  Insertion into vector pUC118
CC   gave pHBR1 which may be used diagnostically or the OSP-1 cDNA
CC   (Q20696) incorporated (with poly-A signal sequence and SV40 T-cell
CC   antigen promoter) into a vector (such as pHSG-757) for expression
CC   (e.g. in CHO cells) of the protein.
CC   See also Q20695-97.
SQ   Sequence   1542 BP;    488 A;    277 C;    360 G;    417 T;

  Query Match           9.2%;  Score 16;  DB 3;  Length 1542;
  Best Local Similarity 100.0%;  Pred. No. 1.13e+01;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1488 aaaataaaacaagtga 1503
          ||||||||||||||||
Qy    145 AAAATAAAACAAGTGA 160


RESULT   20
ID   T35525 standard; DNA; 1584 BP.
AC   T35525;
DT   14-JAN-1997  (first entry)
```



```
DE   Human interferon-beta 5'UTR and coding region.
KW   Gene targeting; gene activation; homologous recombination;
KW   interferon-beta; IFN-beta; multiple sclerosis; gene therapy; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   misc_feature    276
FT                   /*tag=  b
FT                   /function= CAP site
FT   cds             357..920
FT                   /*tag=  c
PN   WO9629411-A1.
PD   26-SEP-1996.
PF   12-MAR-1996; U03377.
PR   17-MAR-1995; US-406030.
PA   (TRAN-) TRANSKARYOTIC THERAPIES INC.
PI   Hauge BM,  Heartlein MW,  Selden RF,  Treco DA;
DR   WPI; 96-443186/44.
PT   Altering expression of genes encoding thrombopoietin, DNase I or
PT   beta-interferon  - using DNA constructs useful in gene therapy to
PT   treat, e.g. cystic fibrosis and multiple sclerosis.
PS   Claim 47; Fig 14A-B; 115pp; English.
CC   A genomic DNA fragment (T35525) corresponds to nucleotides -356 to
CC   +669 relative to the initiation codon of the human interferon-beta
CC   (IFN-beta) gene and encodes a 21-amino acid signal peptide and 120-
CC   amino acid mature protein.  It forms part of a DNA clone isolated
CC   from a human leukocyte genomic DNA library; the 5' untranslated
CC   region is given in T35524.  Non-coding genomic sequences within and
CC   upstream of the transcribed regions of the gene may be used as
CC   targeting sequences in DNA constructs useful for gene activation by
CC   homologous recombination.  Novel genes are generated in which
CC   exogenous and endogenous exons are operatively linked.  Expression
CC   of IFN-beta encoded by the endogenous gene can be activated in
CC   order to alleviate the symptoms of multiple sclerosis.
SQ   Sequence   1584 BP;    492 A;    309 C;    317 G;    466 T;

  Query Match           9.2%;  Score 16;  DB 24;  Length 1584;
  Best Local Similarity 100.0%;  Pred. No. 1.13e+01;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1093 aaaataaattattttt 1108
          ||||||||||||||||
Cp     55 AAAATAAATTATTTTT 40


RESULT   21
ID   N20057 standard; DNA; 1836 BP.
AC   N20057;
DT   27-NOV-1992  (first entry)
DE   Interferon-beta gene.
KW   E.coli; mouse; pBR322; pCR1; pMB9; pSC1; ss.
OS   Homo sapiens.
```



```
FH   Key             Location/Qualifiers
FT   cds             358..921
FT                   /*tag=  a
FT   tata_signal     254..259
FT                   /*tag=  b
FT   polya_signal    1096..1101
FT                   /*tag=  c
PN   WO8202715-A.
PD   19-AUG-1982.
PF   04-FEB-1982; 900385.
PR   04-FEB-1981; JP-014373.
PR   11-JUL-1981; JP-108539.
PA   (NICA-) JAPAN FOUND CANCER.
PA   (SUGA-) SUGANO H.
PA   (GANK-) GAN KENKYU-KAI.
PA   (JURI-) JURIDICAL FOUND.
PI   Sugano H, Taniguchi T, Ono S;
DR   WPI; 82-72320E/34  (72320E).
DR   P-PSDB; P20049.
PT   Human interferon beta gene in recombinant DNA - can be used to
PT   produce human interferon beta in E.coli or material with
PT   identical chemical structure in eukaryotic cells
PS   Disclosure; Fig 2; 19pp; Japanese.
CC   The sequence given is the interferon-beta gene and the
CC   transcriptional control sequences associated with this from human
CC   DNA.  This sequence can be used in recombinant DNA technology to
CC   produce interferon-beta from E.coli or a chemically identical
CC   substance from mouse cells.  The interferon-beta gene is introduced
CC   into E. coli or mouse cells using a vector, pref. lambda phage,
CC   Charon phage or the plasmids pBR322, pCR1, pMB9 or pSC1.
SQ   Sequence   1836 BP;    552 A;    372 C;    370 G;    542 T;

  Query Match           9.2%;  Score 16;  DB 4;  Length 1836;
  Best Local Similarity 100.0%;  Pred. No. 1.13e+01;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1094 aaaataaattattttt 1109
          ||||||||||||||||
Cp     55 AAAATAAATTATTTTT 40


RESULT   22
ID   T79804 standard; DNA; 2169 BP.
AC   T79804;
DT   11-NOV-1997  (first entry)
DE   Potato starch branching enzyme gene promoter region.
KW   Starch branching enzyme; SBE; potato; genetic engineering;
KW   transgenic plant; starch production; modification; ss.
OS   Solanum tuberosum.
PN   WO9704112-A2.
PD   06-FEB-1997.
```

```
PF   12-JUL-1996; E03052.
PR   14-JUL-1995; GB-014435.
PA   (DANI-) DANISCO AS.
PI   Poulsen P;
DR   WPI; 97-132650/12.
PT   Affecting enzymatic activity in plant by expressing anti-sense
PT   intron sequence - esp. for inhibiting starch branching enzyme thus
PT   creating modified starch without post-harvest derivatisation
PS   Claim 9; Page 37; 70pp; English.
CC   T79804 is the promoter region of the potato starch-branching enzyme
CC   (SBE) gene. This sequence is used in genetic constructs to control the
CC   expression of sequences antisense to introns of the SBE gene. Such
CC   constructs are used to produce transgenic plants producing modified
CC   or new forms of starch for industrial use. The sequences allow starches
CC   to be made without post-harvest modification (which uses hazardous
CC   chemicals and requires much energy). Using the antisense intron
CC   sequences is more reliable, efficient and selective than known methods
CC   for regulating enzymatic activity because there is almost no homology
CC   between introns in different SBE genes. The SBE gene promoter is also
CC   useful in other genetic engineering methods.
SQ   Sequence   2169 BP;    777 A;    292 C;    311 G;    789 T;

  Query Match           9.2%;  Score 16;  DB 33;  Length 2169;
  Best Local Similarity 100.0%;  Pred. No. 1.13e+01;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1381 aaaataaattattttt 1396
            ||||||||||||||||
Cp     55 AAAATAAATTATTTTT 40


RESULT   23
ID   T79783 standard; DNA; 2169 BP.
AC   T79783;
DT   10-NOV-1997  (first entry)
DE   Potato starch branching enzyme gene promoter region.
KW   Starch branching enzyme; SBE; potato; genetic engineering;
KW   transgenic plant; starch production; modification; ss.
OS   Solanum tuberosum.
PN   WO9704113-A2.
PD   06-FEB-1997.
PF   12-JUL-1996; E03053.
PR   14-JUL-1995; GB-014437.
PA   (DANI-) DANISCO AS.
PI   Poulsen P;
DR   WPI; 97-132651/12.
PT   Affecting enzymatic activity in plant by expressing sense intron
PT   sequence - esp. for inhibiting starch branching enzyme to create
PT   modified starch without post-harvest derivatisation
PS   Claim 9; Page 33; 70pp; English.
CC   T79783 is the promoter region of the potato starch-branching enzyme
```



```
CC   (SBE) gene. This sequence is used in genetic constructs to control the
CC   expression of sequences antisense to introns of the SBE gene. Such
CC   constructs are used to produce transgenic plants producing modified
CC   or new forms of starch for industrial use. The sequences allow starches
CC   to be made without post-harvest modification (which uses hazardous
CC   chemicals and requires much energy). Using the antisense intron
CC   sequences is more reliable, efficient and selective than known methods
CC   for regulating enzymatic activity because there is almost no homology
CC   between introns in different SBE genes. The SBE gene promoter is also
CC   useful in other genetic engineering methods.
SQ   Sequence   2169 BP;    777 A;    292 C;    311 G;    789 T;

  Query Match           9.2%;  Score 16;  DB 33;  Length 2169;
  Best Local Similarity 100.0%;  Pred. No. 1.13e+01;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1381 aaaataaattattttt 1396
            ||||||||||||||||
Cp     55 AAAATAAATTATTTTT 40


RESULT   24
ID   T34612 standard; cDNA; 2214 BP.
AC   T34612;
DT   02-JAN-1997  (first entry)
DE   Probe for detecting pathology in peripheral nervous system.
KW   Probe; central nervous system; peripheral nervous system; CNS; PNS;
KW   axon; neurone; neuronal cells; glial cells; Schwann cells; trauma;
KW   pathology; regeneration; MS2; metalloprotease; ds.
OS   Homo sapiens.
PN   WO9617865-A2.
PD   13-JUN-1996.
PF   05-DEC-1995; E04777.
PR   05-DEC-1994; DE-443159.
PR   27-JAN-1995; DE-002525.
PA   (BOEF ) BOEHRINGER MANNHEIM GMBH.
PI   Gillen C,  Gleichmann M,  Mueller H;
DR   WPI; 96-287114/29.
PT   Nucleic acids associated with damaged or regenerating nerve cells -
PT   for diagnostic or therapeutic use.
PS   Claim 1; Page 37-39; 54pp; German.
CC   Nucleic acids described in T34587-T34616 are useful as probes for
CC   detecting traumatic and pathological changes in the peripheral
CC   nervous system.  They may also be used to identify hybridising
CC   sequences from an ischial nerve cDNA library, these sequences encode
CC   proteins expressed in damaged and/or regenerating nerve cells. They
CC   can be used to transfect cells which then express the protein which
CC   can then be harvested and studied.  The nucleic acids are preferably
CC   selected from NTII-11 which corresponds to plasmolipin cDNA with a
CC   234 bp N-terminal extension and is expressed in glial cells of the
CC   central nervous system and Schwann cells of the peripheral nervous
```



```
CC   system; NTII-1, which codes for a protein having 79% homology with
CC   human CDC4L in three domains and may be involved in regulating
CC   Schwann cell growth as well as being therapeutically useful for
CC   regenerating nerve cells; and CRII-7 which codes for a protein which
CC   is similar to the macrophage-specific cysteine-rich mouse protein
CC   MS2 and a metalloprotease  and which inhibits interaction between
CC   Schwann cells and the extracellular matrix and is useful for the
CC   therapy of nerve damage by facilitating regeneration of damaged
CC   axons.  This sequence is designated NTI-11.
SQ   Sequence   2214 BP;    614 A;    471 C;    578 G;    551 T;

  Query Match           9.2%;  Score 16;  DB 24;  Length 2214;
  Best Local Similarity 100.0%;  Pred. No. 1.13e+01;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1136 atttcaaagcacacta 1151
            ||||||||||||||||
Cp     76 ATTTCAAAGCACACTA 61


RESULT   25
ID   T79784 standard; DNA; 11478 BP.
AC   T79784;
DT   10-NOV-1997  (first entry)
DE   Full length potato starch branching enzyme gene.
KW   Starch branching enzyme; SBE; potato; genetic engineering;
KW   transgenic plant; starch production; modification; ss.
OS   Solanum tuberosum.
FH   Key             Location/Qualifiers
FT   promoter        1..2131
FT                   /*tag= a
FT   tata_signal     2048..2051
FT                   /*tag= b
FT   cds             2132..9639
FT                   /*tag= c
FT   exon            2132..2209
FT                   /*tag= d
FT                   /number= 1
FT   intron          2210..3376
FT                   /*tag= e
FT                   /number= 1
FT   exon            3377..3496
FT                   /*tag= f
FT                   /number= 2
FT   intron          3497..3817
FT                   /*tag= g
FT                   /number= 2
FT   exon            3818..4037
FT                   /*tag= h
FT                   /number= 3
FT   intron          4038..4541
```



```
FT                   /*tag= i
FT                   /number= 3
FT   exon            4542..4611
FT                   /*tag= j
FT                   /number= 4
FT   intron          4612..4757
FT                   /*tag= k
FT                   /number= 4
FT   exon            4758..5027
FT                   /*tag= l
FT                   /number= 5
FT   intron          5028..5245
FT                   /*tag= m
FT                   /number= 5
FT   exon            5246..6152
FT                   /*tag= n
FT                   /number= 6
FT   intron          6153..6350
FT                   /*tag= o
FT                   /number= 6
FT   exon            6351..6467
FT                   /*tag= p
FT                   /number= 7
FT   intron          6468..6675
FT                   /*tag= q
FT                   /number= 7
FT   exon            6676..6738
FT                   /*tag= r
FT                   /number= 8
FT   intron          6739..7031
FT                   /*tag= s
FT                   /number= 8
FT   exon            7032..7139
FT                   /*tag= t
FT                   /number= 9
FT   intron          7140..7515
FT                   /*tag= u
FT                   /number= 9
FT   exon            7516..7617
FT                   /*tag= v
FT                   /number= 10
FT   intron          7618..7789
FT                   /*tag= w
FT                   /number= 10
FT   exon            7790..7857
FT                   /*tag= x
FT                   /number= 11
FT   intron          7858..8002
FT                   /*tag= y
FT                   /number= 11
FT   exon            8003..8084
```

```
FT                   /*tag=  z
FT                   /number= 12
FT   intron          8085..8326
FT                   /*tag=  aa
FT                   /number= 12
FT   exon            8327..8443
FT                   /*tag=  ab
FT                   /number= 13
FT   intron          8444..9240
FT                   /*tag=  ac
FT                   /number= 13
FT   exon            9241..9636
FT                   /*tag=  ad
FT                   /number= 14
FT   3'utr           9637..11478
FT                   /*tag=  ae
PN   WO9704113-A2.
PD   06-FEB-1997.
PF   12-JUL-1996; E03053.
PR   14-JUL-1995; GB-014437.
PA   (DANI-) DANISCO AS.
PI   Poulsen P;
DR   WPI; 97-132651/12.
PT   Affecting enzymatic activity in plant by expressing sense intron
PT   sequence - esp. for inhibiting starch branching enzyme to create
PT   modified starch without post-harvest derivatisation
PS   Disclosure; Page 38-42; 70pp; English.
CC   T79784 represents the full length potato starch-branching enzyme
CC   (SBE) gene. Sequences antisense to introns of the SBE gene are used
CC   to produce transgenic plants producing modified or new forms of starch
CC   for industrial use. The sequences allow starches to be made without
CC   post-harvest modification (which uses hazardous chemicals and requires
CC   much energy). Using the antisense intron sequences is more reliable,
CC   efficient and selective than known methods for regulating enzymatic
CC   activity because there is almost no homology between introns in
CC   different SBE genes. The SBE gene promoter (see T79783) is also
CC   useful in genetic engineering.
SQ   Sequence   11478 BP;    3615 A;    1863 C;    2127 G;    3864 T;

  Query Match           9.2%;  Score 16;  DB 33;  Length 11478;
  Best Local Similarity 100.0%;  Pred. No. 1.13e+01;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1381 aaaataaattattttt 1396
            ||||||||||||||||
Cp     55 AAAATAAATTATTTTT 40


RESULT   26
ID   T79805 standard; DNA; 11478 BP.
AC   T79805;
```



```
DT   11-NOV-1997  (first entry)
DE   Full length potato starch branching enzyme gene.
KW   Starch branching enzyme; SBE; potato; genetic engineering;
KW   transgenic plant; starch production; modification; ss.
OS   Solanum tuberosum.
FH   Key             Location/Qualifiers
FT   promoter        1..2131
FT                   /*tag=  a
FT   tata_signal     2048..2051
FT                   /*tag=  b
FT   cds             2132..9639
FT                   /*tag=  c
FT   exon            2132..2209
FT                   /*tag=  d
FT                   /number= 1
FT   intron          2210..3376
FT                   /*tag=  e
FT                   /number= 1
FT   exon            3377..3496
FT                   /*tag=  f
FT                   /number= 2
FT   intron          3497..3817
FT                   /*tag=  g
FT                   /number= 2
FT   exon            3818..4037
FT                   /*tag=  h
FT                   /number= 3
FT   intron          4038..4541
FT                   /*tag=  i
FT                   /number= 3
FT   exon            4542..4611
FT                   /*tag=  j
FT                   /number= 4
FT   intron          4612..4757
FT                   /*tag=  k
FT                   /number= 4
FT   exon            4758..5027
FT                   /*tag=  l
FT                   /number= 5
FT   intron          5028..5245
FT                   /*tag=  m
FT                   /number= 5
FT   exon            5246..6152
FT                   /*tag=  n
FT                   /number= 6
FT   intron          6153..6350
FT                   /*tag=  o
FT                   /number= 6
FT   exon            6351..6467
FT                   /*tag=  p
FT                   /number= 7
```



```
FT   intron          6468..6675
FT                   /*tag=  q
FT                   /number= 7
FT   exon            6676..6738
FT                   /*tag=  r
FT                   /number= 8
FT   intron          6739..7031
FT                   /*tag=  s
FT                   /number= 8
FT   exon            7032..7139
FT                   /*tag=  t
FT                   /number= 9
FT   intron          7140..7515
FT                   /*tag=  u
FT                   /number= 9
FT   exon            7516..7617
FT                   /*tag=  v
FT                   /number= 10
FT   intron          7618..7789
FT                   /*tag=  w
FT                   /number= 10
FT   exon            7790..7857
FT                   /*tag=  x
FT                   /number= 11
FT   intron          7858..8002
FT                   /*tag=  y
FT                   /number= 11
FT   exon            8003..8084
FT                   /*tag=  z
FT                   /number= 12
FT   intron          8085..8326
FT                   /*tag=  aa
FT                   /number= 12
FT   exon            8327..8443
FT                   /*tag=  ab
FT                   /number= 13
FT   intron          8444..9240
FT                   /*tag=  ac
FT                   /number= 13
FT   exon            9241..9636
FT                   /*tag=  ad
FT                   /number= 14
FT   3'utr           9637..11478
FT                   /*tag=  ae
PN   WO9704112-A2.
PD   06-FEB-1997.
PF   12-JUL-1996; E03052.
PR   14-JUL-1995; GB-014435.
PA   (DANI-) DANISCO AS.
PI   Poulsen P;
DR   WPI; 97-132650/12.
```



```
PT   Affecting enzymatic activity in plant by expressing anti-sense
PT   intron sequence - esp. for inhibiting starch branching enzyme thus
PT   creating modified starch without post-harvest derivatisation
PS   Disclosure; Page 42-46; 70pp; English.
CC   T79805 represents the full length potato starch-branching enzyme
CC   (SBE) gene. Sequences antisense to introns of the SBE gene are used
CC   to produce transgenic plants producing modified or new forms of starch
CC   for industrial use. The sequences allow starches to be made without
CC   post-harvest modification (which uses hazardous chemicals and requires
CC   much energy). Using the antisense intron sequences is more reliable,
CC   efficient and selective than known methods for regulating enzymatic
CC   activity because there is almost no homology between introns in
CC   different SBE genes. The SBE gene promoter (see T79804) is also
CC   useful in genetic engineering.
SQ   Sequence   11478 BP;    3615 A;    1863 C;    2127 G;    3864 T;

  Query Match           9.2%;  Score 16;  DB 33;  Length 11478;
  Best Local Similarity 100.0%;  Pred. No. 1.13e+01;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1381 aaaataaattattttt 1396
            ||||||||||||||||
Cp     55 AAAATAAATTATTTTT 40


RESULT   27
ID   T58840 standard; DNA; 580073 BP.
AC   T58840;
DT   27-MAR-1997  (first entry)
DE   Mycoplasma genitalium genome.
KW   M. genitalium; DNAA; DNA gyrase; origin of replication;
KW   megabase shotgun sequencing method; open reading frame; ORF; ss.
OS   Mycoplasma genitalium.
FH   Key             Location/Qualifiers
FT   cds             8552..9184
FT                   /*tag=  a
FT                   /label= MG006
FT                   /note= "Previously identified as MORF-20076, the
FT                   encoded protein shows 27.59 percentage
FT                   identity to thymidylate kinase (CDC8)
FT                   from Saccharomyces cerevisiae"
FT   cds             11252..12040
FT                   /*tag=  b
FT                   /label= MG009
FT                   /note= "Previously identified as MORF-20078, the
FT                   encoded protein shows 35.43 percentage
FT                   identity to the Bacillus subtilis hypothetical
FT                   protein covered in accession number
FT                   GB:D26185_102"
FT   cds             12069..12725
FT                   /*tag=  c
```

```
FT                   /label= MG010
FT                   /note= "Previously identified as MORF-20079, the
FT                   encoded protein shows 25.73 percentage
FT                   identity to DNA primase (dnaE) from
FT                   Clostridium acetobutylicum"
FT   cds             complement (13570..14247)
FT                   /*tag= d
FT                   /label= MG012
FT                   /note= "Previously identified as MORF-20080, the
FT                   encoded protein shows 31.50 percentage
FT                   identity to the ribosomal protein S6
FT                   modification protein (rimK) from Escherichia
FT                   coli"
FT   cds             complement (14396..15217)
FT                   /*tag= e
FT                   /label= MG013
FT                   /note= "Previously identified as MORF-19823, MORF-20080
FT                   and MORF-20081, the encoded protein shows 33.04
FT                   percentage identity to 5,10-methylene-tetra-
FT                   hydrofolate dehydrogenase (folD) from E. coli"
FT   cds             17474..19243
FT                   /*tag= f
FT                   /label= MG015
FT                   /note= "Previously identified as MORF-20084, the
FT                   encoded protein shows 32.23 percentage
FT                   identity to transport ATP-binding protein
FT                   (msbA) from E. coli"
FT   cds             26478..27344
FT                   /*tag= g
FT                   /label= MG023
FT                   /note= "Previously identified as MORF-20092, the
FT                   encoded protein shows 45.96 percentage
FT                   identity to fructose-bisphosphate aldolase
FT                   (tsr) from B. subtilis"
FT   cds             27345..28448
FT                   /*tag= h
FT                   /label= MG024
FT                   /note= "Previously identified as MORF-19826 and
FT                   MORF-20093, the encoded protein shows 46.84
FT                   percentage identity to GTP-binding protein
FT                   from E. coli"
FT   cds             36987..38978
FT                   /*tag= i
FT                   /label= MG032
FT                   /note= "Previously identified as MORF-20099, the
FT                   encoded protein shows 26.82 percentage
FT                   identity to ATP-dependent nuclease (addA)
FT                   from B. subtilis"
FT   cds             39242..39904
FT                   /*tag= j
FT                   /label= MG033
```



```
FT                   /note= "Previously identified as MORF-20100, the
FT                   encoded protein shows 35.90 percentage
FT                   identity to glycerol uptake facilitator
FT                   (glpF) from B. subtilis"
FT   cds             complement (39873..40514)
FT                   /*tag= k
FT                   /label= MG034
FT                   /note= "Previously identified as MORF-20101, the
FT                   encoded protein shows 48.13 percentage
FT                   identity to thymidylate kinase (tdk)
FT                   from B. subtilis"
FT   cds             40543..41787
FT                   /*tag= l
FT                   /label= MG035
FT                   /note= "Previously identified as MORF-20102, the
FT                   encoded protein shows 30.71 percentage
FT                   identity to histidyl-tRNA synthetase (hisS)
FT                   from Mycobacterium leprae"
FT   cds             complement (44751..46277)
FT                   /*tag= m
FT                   /label= MG038
FT                   /note= "Previously identified as MORF-20105, the
FT                   encoded protein shows 46.83 percentage
FT                   identity to glycerol kinase (glpK)
FT                   from E. coli"
FT   cds             complement (46268..47422)
FT                   /*tag= n
FT                   /label= MG039
FT                   /note= "Previously identified as MORF-19831 and
FT                   MORF-20106, the encoded protein shows 43.20
FT                   percentage identity to glycerol-3-phosphate
FT                   dehydrogenase (GUT2) from S. cerevisiae"
FT   cds             49377..49643
FT                   /*tag= o
FT                   /label= MG041
FT                   /note= "The encoded protein shows 48.86 percentage
FT                   identity to phosphohistidinoprotein-hexose
FT                   phosphotransferase (ptsH) from Mycoplasma
FT                   capricolum"
FT   cds             50060..51520
FT                   /*tag= p
FT                   /label= MG042
FT                   /note= "Previously identified as MORF-19832 and
FT                   MORF-20108, the encoded protein shows 41.92
FT                   percentage identity to spermidine/
FT                   putrescine transport ATP-binding protein
FT                   (potA) from E. coli"
FT   cds             51525..52382
FT                   /*tag= q
FT                   /label= MG043
FT                   /note= "Previously identified as MORF-20110, the
```



```
FT                   encoded protein shows 26.51 percentage
FT                   identity to spermidine/putrescine transport
FT                   system permease protein (potB) from E. coli"
FT   cds             52366..53220
FT                   /*tag= r
FT                   /label= MG044
FT                   /note= "Previously identified as MORF-20111, the
FT                   encoded protein shows 29.45 percentage
FT                   identity to spermidine/putrescine transport
FT                   system permease protein C (potC) from E. coli"
FT   cds             54658..55605
FT                   /*tag= s
FT                   /label= MG046
FT                   /note= "Previously identified as MORF-20112, the
FT                   encoded protein shows 36.60 percentage
FT                   identity to sialoglycoprotease (gcp)
FT                   from Pasteurella haemolytica"
FT   cds             complement (56970..58310)
FT                   /*tag= t
FT                   /label= MG048
FT                   /note= "Previously identified as MORF-19834,
FT                   MORF-20114 and MORF-20115, the encoded protein
FT                   shows 43.02 percentage identity to signal
FT                   recognition particle protein (ffh) from B.
FT                   subtilis"
FT   cds             58117..59079
FT                   /*tag= u
FT                   /label= MG049
FT                   /note= "Previously identified as MORF-20114 and
FT                   MORF-20115, the encoded protein shows 44.78
FT                   percentage identity to purine-nucleoside
FT                   phosphorylase (deoD) from E. coli"
FT   cds             59083..59754
FT                   /*tag= v
FT                   /label= MG050
FT                   /note= "Previously identified as MORF-20117, the
FT                   encoded protein shows 83.03 percentage
FT                   identity to deoxyribose-phosphate aldolase
FT                   (deoC) from Mycoplasma pneumoniae"
FT   cds             complement (64898..65731)
FT                   /*tag= w
FT                   /label= MG056
FT                   /note= "Previously identified as MORF-20122, the
FT                   encoded protein shows 30.25 percent
FT                   identity to the protein disclosed in
FT                   GB:D26185_99 from B. subtilis"
FT   cds             complement (65713..66249)
FT                   /*tag= x
FT                   /label= MG057
FT                   /note= "Previously identified as MORF-20123, the
FT                   encoded protein shows 38.90 percentage
```



```
FT                   identity to the protein disclosed in
FT                   GB:D26185_104 from B. subtilis"
FT   cds             81047..82597
FT                   /*tag= y
FT                   /label= MG067
FT                   /note= "Previously identified as MORF-19845, the
FT                   encoded protein shows 28.84 percentage
FT                   identity to glutamic acid specific protease
FT                   (SPase) from Staphylococcus aureus"
FT   cds             91065..91919
FT                   /*tag= z
FT                   /label= MG070
FT                   /note= "Previously identified as MORF-20136, the
FT                   encoded protein shows 34.8 percentage
FT                   identity to ribosomal protein S2 (rpS2)
FT                   from Spirulina plantensis"
FT   cds             103104..104324
FT                   /*tag= aa
FT                   /label= MG077
FT                   /note= "Previously identified as MORF-20140, the
FT                   encoded protein shows 28.05 percentage
FT                   identity to oligopeptide transport system
...
Note: remainder of annotations omitted.

  Query Match           9.2%;  Score 16;  DB 27;  Length 580073;
  Best Local Similarity 100.0%;  Pred. No. 1.13e+01;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db 448919 cttgttttatttttat 448934
          ||||||||||||||||
Cp    157 CTTGTTTTATTTTTAT 142


RESULT   28
ID   T24800 standard; cDNA to mRNA; 119 BP.
AC   T24800;
DT   30-OCT-1996  (first entry)
DE   Human gene signature HUMGS06885.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
```

```
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 1698; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   119 BP;    44 A;    17 C;    21 G;    37 T;

  Query Match           8.6%;  Score 15;  DB 22;  Length 119;
  Best Local Similarity 100.0%;  Pred. No. 4.13e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     33 tgcaaaaataaatta 47
          ||||||||||||||||
Cp     59 TGCAAAAATAAATTA 45


RESULT   29
ID   T21758 standard; cDNA to mRNA; 129 BP.
AC   T21758;
DT   01-AUG-1996  (first entry)
DE   Human gene signature HUMGS03214.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
```



```
PT   tissues
PS   Claim 1; Page 961; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   129 BP;    45 A;    28 C;    19 G;    33 T;

  Query Match           8.6%;  Score 15;  DB 19;  Length 129;
  Best Local Similarity 93.8%;  Pred. No. 4.13e+01;
  Matches   15;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db    106 aaataaaantattttta 121
          ||||||| ||||||||
Cp     54 AAATAAATTATTTTTA 39


RESULT   30
ID   T24427 standard; cDNA to mRNA; 189 BP.
AC   T24427;
DT   10-OCT-1996  (first entry)
DE   Human gene signature HUMGS06460.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 1608; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
```



```
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   189 BP;    74 A;    23 C;    18 G;    74 T;

  Query Match           8.6%;  Score 15;  DB 21;  Length 189;
  Best Local Similarity 100.0%;  Pred. No. 4.13e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    169 aaaaataaattattt 183
          ||||||||||||||||
Cp     56 AAAAATAAATTATTT 42


RESULT   31
ID   Q60229 standard; DNA; 303 BP.
AC   Q60229;
DT   16-MAR-1994  (first entry)
DE   Human brain Expressed Sequence Tag EST02222.
KW   Gene transcription product; genetic markers; tagging; in vivo;
KW   transcription; mapping; locations; chromosomes; chromosomal; ss.
OS   Homo sapiens.
PN   WO9316178-A.
PD   19-AUG-1993.
PF   12-FEB-1993; U01294.
PR   12-FEB-1992; US-837195.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
PI   Adams MD, Moreno RF, Venter CJ;
DR   WPI; 93-272882/34.
PT   Enriched oligonucleotides and corresp. sequences - used as
PT   markers for human genes transcribed in-vivo, facilitate tagging
PT   of most human genes
PS   Example 4; Page 301-302; 500pp; English.
CC   The Expressed Sequence Tag was isolated from a human brain cDNA
CC   library as part of a large set of ESTs which can be used as markers
CC   for human genes transcribed in vivo. They can be used to facilitate
CC   tagging of most human genes, for mapping locations of expressed genes
CC   on chromosomes, for individual or forensic identification, for mapping
CC   locations of disease-associated genes, for identification of tissue
CC   type, and for prepn. of antisense sequences, probes and constructs.
```



```
CC   EST02222 has a "poor" coding probability as evaluated using the
CC   coding-region prediction program CRM. See also Q59041-Q61440.
SQ   Sequence   303 BP;    120 A;    44 C;    44 G;    89 T;

  Query Match           8.6%;  Score 15;  DB 8;  Length 303;
  Best Local Similarity 100.0%;  Pred. No. 4.13e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     25 aaaatatgcaaattta 39
          ||||||||||||||||
Qy    119 AAAATATGCAATTTA 133


RESULT   32
ID   Q60068 standard; DNA; 304 BP.
AC   Q60068;
DT   16-MAR-1994  (first entry)
DE   Human brain Expressed Sequence Tag EST02052.
KW   Gene transcription product; genetic markers; tagging; in vivo;
KW   transcription; mapping; locations; chromosomes; chromosomal; ss.
OS   Homo sapiens.
PN   WO9316178-A.
PD   19-AUG-1993.
PF   12-FEB-1993; U01294.
PR   12-FEB-1992; US-837195.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
PI   Adams MD, Moreno RF, Venter CJ;
DR   WPI; 93-272882/34.
PT   Enriched oligonucleotides and corresp. sequences - used as
PT   markers for human genes transcribed in-vivo, facilitate tagging
PT   of most human genes
PS   Example 4; Page 276; 500pp; English.
CC   The Expressed Sequence Tag was isolated from a human brain cDNA
CC   library as part of a large set of ESTs which can be used as markers
CC   for human genes transcribed in vivo. They can be used to facilitate
CC   tagging of most human genes, for mapping locations of expressed genes
CC   on chromosomes, for individual or forensic identification, for mapping
CC   locations of disease-associated genes, for identification of tissue
CC   type, and for prepn. of antisense sequences, probes and constructs.
CC   EST02052 has a "poor" coding probability as evaluated using the
CC   coding-region prediction program CRM. See also Q59041-Q61440.
SQ   Sequence   304 BP;    57 A;    67 C;    49 G;    124 T;

  Query Match           8.6%;  Score 15;  DB 8;  Length 304;
  Best Local Similarity 100.0%;  Pred. No. 4.13e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    215 aatcttatttctaat 229
          ||||||||||||||||
Qy     82 AATCTTATTTCTAAT 96
```

```
RESULT   33
ID   T20989 standard; cDNA to mRNA; 312 BP.
AC   T20989;
DT   03-JUL-1996  (first entry)
DE   Human gene signature HUMGS02268.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 774; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   312 BP;    73 A;    58 C;    74 G;    106 T;

  Query Match           8.6%;  Score 15;  DB 18;  Length 312;
  Best Local Similarity 100.0%;  Pred. No. 4.13e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    273 ctttgaaatgtgtaa 287
          ||||||||||||||||
Qy     68 CTTTGAAATGTGTAA 82


RESULT   34
ID   T19987 standard; cDNA to mRNA; 337 BP.
```



```
AC   T19987;
DT   17-JUL-1996  (first entry)
DE   Human gene signature HUMGS01122.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 529; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   337 BP;    107 A;    43 C;    70 G;    116 T;

  Query Match           8.6%;  Score 15;  DB 19;  Length 337;
  Best Local Similarity 100.0%;  Pred. No. 4.13e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     87 cttgttttatttttta 101
          ||||||||||||||||
Cp    157 CTTGTTTTATTTTTA 143


RESULT   35
ID   T22621 standard; cDNA to mRNA; 359 BP.
AC   T22621;
DT   01-OCT-1996  (first entry)
DE   Human gene signature HUMGS04273.
```



```
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 1171; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   359 BP;    123 A;    48 C;    54 G;    128 T;

  Query Match           8.6%;  Score 15;  DB 21;  Length 359;
  Best Local Similarity 100.0%;  Pred. No. 4.13e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    181 tcttatttctaatgt 195
          ||||||||||||||||
Qy     84 TCTTATTTCTAATGT 98


RESULT   36
ID   Q60080 standard; DNA; 390 BP.
AC   Q60080;
DT   16-MAR-1994  (first entry)
DE   Human brain Expressed Sequence Tag EST02064.
KW   Gene transcription product; genetic markers; tagging; in vivo;
KW   transcription; mapping; locations; chromosomes; chromosomal; ss.
OS   Homo sapiens.
```



```
PN   WO9316178-A.
PD   19-AUG-1993.
PF   12-FEB-1993; U01294.
PR   12-FEB-1992; US-837195.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
PI   Adams MD, Moreno RF, Venter CJ;
DR   WPI; 93-272882/34.
PT   Enriched oligonucleotides and corresp. sequences - used as
PT   markers for human genes transcribed in-vivo, facilitate tagging
PT   of most human genes
PS   Example 4; Page 278; 500pp; English.
CC   The Expressed Sequence Tag was isolated from a human brain cDNA
CC   library as part of a large set of ESTs which can be used as markers
CC   for human genes transcribed in vivo. They can be used to facilitate
CC   tagging of most human genes, for mapping locations of expressed genes
CC   on chromosomes, for individual or forensic identification, for mapping
CC   locations of disease-associated genes, for identification of tissue
CC   type, and for prepn. of antisense sequences, probes and constructs.
CC   EST02064 has a "poor" coding probability as evaluated using the
CC   coding-region prediction program CRM. See also Q59041-Q61440.
SQ   Sequence   390 BP;    141 A;    72 C;    64 G;    109 T;

  Query Match           8.6%;  Score 15;  DB 8;  Length 390;
  Best Local Similarity 100.0%;  Pred. No. 4.13e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    133 ataaattatttttaa 147
          ||||||||||||||||
Cp     52 ATAAATTATTTTTAA 38


RESULT   37
ID   T67805 standard; DNA; 405 BP.
AC   T67805;
DT   29-JUL-1997  (first entry)
DE   H. pylori cytoplasmic protein ORF 36520792.aa.
KW   Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW   identification; binding compound; bacterium; life cycle; activator;
KW   bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;
KW   diagnosis; ds.
OS   Helicobacter pylori.
FH   Key             Location/Qualifiers
FT   cds             1..405
FT                   /*tag=  a
FT                   /transl_except= (pos: 376..378, aa: Xaa)
FT                   /note= "Xaa = Unknown"
PN   WO9640893-A1.
PD   19-DEC-1996.
PF   06-JUN-1996; U09122.
PR   07-JUN-1995; US-487032.
PR   01-APR-1996; US-630405.
```

```
PA   (ASTR ) ASTRA AB.
PI   Berglindh OT,  Smith D,  Mellgaerd BL;
DR   WPI; 97-052306/05.
DR   P-PSDB; W20461.
PT   Helicobacter pylori nucleic acid sequences and related
PT   polypeptide(s)  - useful for vaccines to treat or prevent H. pylori
PT   infection, and to detect Helicobacter
PS   Claim 9; Page -; 148lpp; English.
CC   This sequence encodes a H. pylori cytoplasmic protein.
CC   The protein may be used in a vaccine to prevent or treat H. pylori
CC   infection or to identify H. pylori polypeptide binding compounds,
CC   useful as potential H. pylori life cycle activators or inhibitors.
CC   The genomic sequence of H. pylori (ATCC 55679) was determined from
CC   overlapping contigs generated by mechanically shearing the bacterial
CC   DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC   and the predicted coding regions defined by computer evaluation. To
CC   identify likely H. pylori antigens for vaccine development, the amino
CC   acid sequences predicted from various ORF were analysed for significant
CC   homology to other known or exported membrane proteins. Having identified
CC   and determined the sequences of interest, particular regions can be
CC   isolated from H. pylori by PCR amplification for recombinant polypeptide
CC   production, e.g. in E. coli hosts.
CC   Note: This DNA sequence is not reproduced in the specification and
CC   has been derived from the related specification, WO9719098.
SQ   Sequence   405 BP;    140 A;    66 C;    89 G;    109 T;

  Query Match           8.6%;  Score 15;  DB 30;  Length 405;
  Best Local Similarity 100.0%;  Pred. No. 4.13e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    379 aaacaagtgaatttg 393
          |||||||||||||||
Qy    151 AAACAAGTGAATTTG 165


RESULT   38
ID   T77485 standard; DNA; 405 BP.
AC   T77485;
DT   12-AUG-1997  (first entry)
DE   H. pylori cytoplasmic protein ORF 36520792.aa.
KW   Transmembrane; cytoplasmic; cell envelope; flagella; transport;
KW   secreted; periplasmic; chronic gastritis; duodenal ulcer disease;
KW   activator; inhibitor; bacterial life cycle; vaccine; immunise;
KW   detection; antisense; inhibition; ds.
OS   Helicobacter pylori.
FH   Key             Location/Qualifiers
FT   cds             1..405
FT                   /*tag= a
FT                   /transl_except= (pos: 376..378, aa: Xaa)
PN   WO9719098-A1.
PD   29-MAY-1997.
```



```
PF   15-NOV-1996; U18542.
PR   17-NOV-1995; US-561469.
PA   (ASTR ) ASTRA AB.
PI   Smith DH;
DR   WPI; 97-298052/27.
DR   P-PSDB; W24667.
PT   Helicobacter pylori nucleic acid sequences and related proteins -
PT   used for diagnostics and therapeutics
PS   Claim 1; Page 115; 235pp; English.
CC   This sequence encodes an H. pylori cytoplasmic protein. This sequence
CC   showed homology to methyl-accepting chemotaxis protein.
CC   Helicobacter pylori has been strongly linked to chronic gastritis and
CC   duodenal ulcer disease.  The nucleic acid sequences of the invention
CC   are used to evaluate compounds, especially activators or inhibitors of
CC   bacterial life cycle, for the ability to bind an H. pylori nucleic acid
CC   sequence.  The nucleic acid sequences, and corresponding proteins, are
CC   also useful for generating vaccines for immunising subjects against H.
CC   pylori or for use in detecting the presence of Helicobacter species in
CC   a sample.  Antisense nucleic acid sequences of these sequences are
CC   used to inhibit expression of a gene from Helicobacter species.  H.
CC   pylori whole genomic DNA was isolated and nebulised to a median size of
CC   2000 bp.  Purified DNA fragments were blunt-ended and ligated to unique
CC   BstXI-linker adapters in 100-1000 fold molar excess.  These linkers are
CC   complementary to the BstXI-cut pMPX vectors, while the overhang is not
CC   self-complementary.  Therefore the linkers will not concatemerise nor
CC   will the cut vector re-ligate itself easily.  The linker-adapter inserts
CC   were ligated to each of the 20 pMPX vectors to construct a series of
CC   shotgun subclone libraries.  The purified DNA samples were then
CC   sequenced.
CC   Note: The ORF/protein reference number for this sequence was obtained
CC   from the related specification, WO9640893.
SQ   Sequence   405 BP;    140 A;    66 C;    89 G;    109 T;

  Query Match           8.6%;  Score 15;  DB 30;  Length 405;
  Best Local Similarity 100.0%;  Pred. No. 4.13e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    379 aaacaagtgaatttg 393
          |||||||||||||||
Qy    151 AAACAAGTGAATTTG 165


RESULT   39
ID   T68020 standard; DNA; 486 BP.
AC   T68020;
DT   15-JUL-1997  (first entry)
DE   H. pylori cytoplasmic protein ORF 07ee20513orf1.
KW   Cytoplasmic; vaccine; prevention; treatment; infection; identification;
KW   binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW   duodenal ulcer disease; chronic gastritis; diagnosis; envelope; ds.
OS   Helicobacter pylori.
```



```
FH   Key             Location/Qualifiers
FT   cds             1..486
FT                   /*tag= a
FT                   /note= "no stop codon given in sequence"
PN   WO9640893-A1.
PD   19-DEC-1996.
PF   06-JUN-1996; U09122.
PR   07-JUN-1995; US-487032.
PR   01-APR-1996; US-630405.
PA   (ASTR ) ASTRA AB.
PI   Berglindh OT,  Smith D,  Mellgaerd BL;
DR   WPI; 97-052306/05.
DR   P-PSDB; W20767.
PT   Helicobacter pylori nucleic acid sequences and related
PT   polypeptide(s)  - useful for vaccines to treat or prevent H. pylori
PT   infection, and to detect Helicobacter
PS   Claim 9; Pages 849; 1481pp; English.
CC   The present sequence encodes a Helicobacter pylori cytoplasmic protein.
CC   The protein may be used in a vaccine to prevent or treat H. pylori
CC   infection or to identify H. pylori polypeptide binding compounds, useful
CC   as potential H. pylori life cycle activators or inhibitors.  The genomic
CC   sequence of H. pylori (ATCC 55679) was determined from overlapping
CC   contigs generated by mechanically shearing the bacterial DNA. The
CC   sequences were analysed for ORF of at least 180 nucleotides, and the
CC   predicted coding regions defined by computer evaluation. To identify
CC   likely H. pylori antigens for vaccine development, the amino acid
CC   sequences predicted from various ORF were analysed for significant
CC   homology to other known or exported membrane proteins. Having identified
CC   and determined the sequences of interest, particular regions can be
CC   isolated from H. pylori by PCR amplification for recombinant polypeptide
CC   production, e.g. in E. coli hosts.
SQ   Sequence   486 BP;    159 A;    87 C;    114 G;    126 T;

  Query Match           8.6%;  Score 15;  DB 30;  Length 486;
  Best Local Similarity 100.0%;  Pred. No. 4.13e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    463 aaacaagtgaatttg 477
          |||||||||||||||
Qy    151 AAACAAGTGAATTTG 165


RESULT   40
ID   T66501 standard; DNA; 927 BP.
AC   T66501;
DT   16-JUL-1997  (first entry)
DE   Streptococcal inhibitor of complement mediated lysis (SIC) gene.
KW   Streptococcal inhibitor of complemented mediated lysis; SIC;
KW   clusterin; histidine-rich glycoprotein; vaccine; diagnosis; ds.
OS   Streptococcus pyogenes strain API(40/58), M1 serotype.
FH   Key             Location/Qualifiers
```



```
FT   cds             10..927
FT                   /*tag= a
FT   signal_peptide  10..105
FT                   /*tag= b
FT   mat_peptide     106..924
FT                   /*tag= c
FT   repeat_region   106..563
FT                   /*tag= d
FT                   /note= "N-terminal short repeat region"
FT   repeat_unit     130..142
FT                   /*tag= e
FT                   /note= "encodes peptide repeat I"
FT   repeat_unit     142..166
FT                   /*tag= f
FT                   /note= "encodes peptide repeat II"
FT   repeat_unit     160..172
FT                   /*tag= g
FT                   /note= "encodes peptide repeat I"
FT   repeat_unit     175..187
FT                   /*tag= h
FT                   /note= "encodes peptide repeat III"
FT   repeat_unit     205..217
FT                   /*tag= i
FT                   /note= "encodes peptide repeat III"
FT   repeat_unit     211..223
FT                   /*tag= j
FT                   /rpt_type= TANDEM
FT                   /note= "encodes peptide repeat IV"
FT   repeat_unit     226..238
FT                   /*tag= k
FT                   /rpt_type= TANDEM
FT                   /note= "encodes peptide repeat IV"
FT   repeat_unit     238..251
FT                   /*tag= l
FT                   /note= "encodes peptide repeat V"
FT   repeat_unit     265..289
FT                   /*tag= m
FT                   /note= "encodes peptide repeat II"
FT   repeat_unit     286..299
FT                   /*tag= n
FT                   /note= "encodes peptide repeat V"
FT   repeat_unit     328..414
FT                   /*tag= o
FT                   /rpt_type= TANDEM
FT                   /note= "encodes peptide repeat R1"
FT   repeat_unit     415..500
FT                   /*tag= p
FT                   /rpt_type= TANDEM
FT                   /note= "encodes peptide repeat R2"
FT   repeat_unit     501..563
FT                   /*tag= q
```

```
FT                   /rpt_type= TANDEM
FT                   /note= "encodes peptide repeat R3"
PN   WO9713786-A1.
PD   17-APR-1997.
PF   02-OCT-1996; SE1238.
PR   09-OCT-1995; SE-003495.
PA   (HIGH-) HIGHTECH RECEPTOR AB.
PI   Akesson P,  Bjorck L;
DR   WPI; 97-235840/21.
DR   P-PSDB; W16476.
PT   Protein designated streptococcal inhibitor of complement mediated
PT   lysis - useful for detecting, and vaccinating against Streptococcus
PT   infection
PS   Claim 8; Page 39-40; 63pp; English.
CC   A DNA sequence (T66501) codes for a novel protein (W16476),
CC   designated SIC (streptococcal inhibitor of complement mediated
CC   lysis), that plays a role in Streptococcus pyogenes pathogenicity
CC   and virulence, and which inhibits haemolysis by interacting with
CC   clusterin and members of the cystatin protein super family such as
CC   histidine rich glycoprotein (HRG).  The sic gene is located approx.
CC   188 bp downstream of the protein H gene (sph) and approx. 2.1 kb
CC   upstream of the C5a peptidase gene in S. pyogenes AP1.  Variant
CC   sic partial genes (see also T66502-03) have also been obtd. from
CC   other S. pyogenes strains.  The isolated sic genes can be used to
CC   produce recombinant SIC proteins in host cells for use in vaccines,
CC   or to design probes used to determine the presence of virulent S.
CC   pyogenes in a sample.
SQ   Sequence   927 BP;    293 A;    184 C;    225 G;    225 T;

  Query Match           8.6%;  Score 15;  DB 30;  Length 927;
  Best Local Similarity 100.0%;  Pred. No. 4.13e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     16 attagaaataagatt 30
          ||||||||||||||| 
Cp     96 ATTAGAAATAAGATT 82


RESULT   41
ID   T13864 standard; DNA; 1204 BP.
AC   T13864;
DT   18-JUL-1997  (first entry)
DE   Streptococcal inhibitor of complement mediated lysis (SIC) gene.
KW   Streptococcal inhibitor of complemented mediated lysis; SIC;
KW   clusterin; histidine-rich glycoprotein; vaccine; diagnosis; ds.
OS   Streptococcus pyogenes strain API(40/58), M1 serotype.
FH   Key             Location/Qualifiers
FT   -35_signal      120..124
FT                   /*tag=  a
FT   -10_signal      143..147
FT                   /*tag=  b
```



```
FT   rbs             173..179
FT                   /*tag=  c
FT   cds             189..1106
FT                   /*tag=  d
FT   signal_peptide  189..284
FT                   /*tag=  e
FT   mat_peptide     285..1103
FT                   /*tag=  f
FT   repeat_region   285..742
FT                   /*tag=  g
FT                   /note= "N-terminal short repeat region"
FT   repeat_unit     309..321
FT                   /*tag=  h
FT                   /note= "encodes peptide repeat I"
FT   repeat_unit     321..345
FT                   /*tag=  i
FT                   /note= "encodes peptide repeat II"
FT   repeat_unit     339..351
FT                   /*tag=  j
FT                   /note= "encodes peptide repeat I"
FT   repeat_unit     354..366
FT                   /*tag=  k
FT                   /note= "encodes peptide repeat III"
FT   repeat_unit     384..396
FT                   /*tag=  l
FT                   /note= "encodes peptide repeat III"
FT   repeat_unit     390..402
FT                   /*tag=  m
FT                   /rpt_type= TANDEM
FT                   /note= "encodes peptide repeat IV"
FT   repeat_unit     405..417
FT                   /*tag=  n
FT                   /rpt_type= TANDEM
FT                   /note= "encodes peptide repeat IV"
FT   repeat_unit     417..430
FT                   /*tag=  o
FT                   /note= "encodes peptide repeat V"
FT   repeat_unit     444..468
FT                   /*tag=  p
FT                   /note= "encodes peptide repeat II"
FT   repeat_unit     465..478
FT                   /*tag=  q
FT                   /note= "encodes peptide repeat V"
FT   repeat_unit     507..593
FT                   /*tag=  r
FT                   /rpt_type= TANDEM
FT                   /note= "encodes peptide repeat R1"
FT   repeat_unit     594..679
FT                   /*tag=  s
FT                   /rpt_type= TANDEM
FT                   /note= "encodes peptide repeat R2"
```



```
FT   repeat_unit     680..742
FT                   /*tag=  t
FT                   /rpt_type= TANDEM
FT                   /note= "encodes peptide repeat R3"
PN   WO9713786-A1.
PD   17-APR-1997.
PF   02-OCT-1996; SE1238.
PR   09-OCT-1995; SE-003495.
PA   (HIGH-) HIGHTECH RECEPTOR AB.
PI   Akesson P,  Bjorck L;
DR   WPI; 97-235840/21.
DR   P-PSDB; W16476.
PT   Protein designated streptococcal inhibitor of complement mediated
PT   lysis - useful for detecting, and vaccinating against Streptococcus
PT   infection
PS   Example 2; Fig 1; 63pp; English.
CC   A DNA sequence (T13864) codes for a novel protein (W16476),
CC   designated SIC (streptococcal inhibitor of complement mediated
CC   lysis), that plays a role in Streptococcus pyogenes pathogenicity
CC   and virulence, and which inhibits haemolysis by interacting with
CC   clusterin and members of the cystatin protein super family such as
CC   histidine rich glycoprotein (HRG).  The sic gene is located approx.
CC   188 bp downstream of the protein H gene (sph) and approx. 2.1 kb
CC   upstream of the C5a peptidase gene in S. pyogenes AP1.  Variant
CC   sic partial genes (see also T66502-03) have also been obtd. from
CC   other S. pyogenes strains.  The isolated sic genes can be used to
CC   produce recombinant SIC proteins in host cells for use in vaccines,
CC   or to design probes used to determine the presence of virulent S.
CC   pyogenes in a sample.
SQ   Sequence   1204 BP;    389 A;    222 C;    281 G;    312 T;

  Query Match           8.6%;  Score 15;  DB 30;  Length 1204;
  Best Local Similarity 100.0%;  Pred. No. 4.13e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    195 attagaaataagatt 209
          ||||||||||||||| 
Cp     96 ATTAGAAATAAGATT 82


RESULT   42
ID   T45357 standard; cDNA; 1637 BP.
AC   T45357;
DT   03-JUL-1997  (first entry)
DE   Human ceramide glucosyltransferase encoding cDNA.
KW   Ceramide glucosyltransferase; GlcT-1; glycosphingolipid biosynthesis;
KW   glycosylation; glycosphingoceramide; glucosylceramide; human; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             291..1475
FT                   /*tag=  a
```



```
FT                   /product= ceramide_glucosyltransferase
FT                   /EC_number= 2.4.1.80
PN   EP-748868-A2.
PD   18-DEC-1996.
PF   13-JUN-1996; 109447.
PR   15-JUN-1995; JP-148472.
PA   (RIKA ) INST PHYSICAL & CHEM RES.
PI   Hirabayashi Y,  Ichikawa S;
DR   WPI; 97-036152/04.
DR   P-PSDB; W09088.
PT   New isolated human ceramide glucosyl:transferase - useful for prodn.
PT   of glucosyl:ceramide, a precursor of glycosphingolipid biosynthesis
PS   Claim 7; Page 15; 18pp; English.
CC   The present sequence encodes a novel ceramide glucosyltransferase
CC   (GlcT-1), which has catalytic activity for glucose transfer from
CC   UDP-glucose to ceramide. The present sequence can be used to produce
CC   large amounts of ceramide glucosyltransferase which catalyses the first
CC   glycosylation step of glycosphingoceramide to produce glucosylceramide,
CC   a precursor to more than 300 glycosphingolipid biosyntheses. It can
CC   also be used as a probe in clinical tests and diagnosis of diseases
CC   based on ceramide glucosyltransferase.
SQ   Sequence   1637 BP;    416 A;    375 C;    394 G;    452 T;

  Query Match           8.6%;  Score 15;  DB 29;  Length 1637;
  Best Local Similarity 100.0%;  Pred. No. 4.13e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1598 taatttatttttgca 1612
          ||||||||||||||| 
Qy     45 TAATTTATTTTTGCA 59


RESULT   43
ID   Q04487 standard; DNA; 1784 BP.
AC   Q04487;
DT   09-OCT-1990  (first entry)
DE   Sequence of protein H encoding gene.
KW   Protein H; IgG; autoimmune; HIV; AIDS; ds.
OS   Streptococcus G148.
FH   Key             Location/Qualifiers
FT   cds             328..1453
FT                   /*tag=  a
PN   EP-371199-A.
PD   6-JUN-1990.
PF   21-JUL-1989; 113430.
PR   21-NOV-1988; JP-295527.
PR   9-MAR-1989; JP-058434.
PA   (SUMO) Sumitomo Chem Ind KK.
PI   Gomi H, Hozumi T, Hattori S, Tagawa C, Kishimoto F, Bjorck L;
DR   WPI; 90-172687/23.
DR   P-PSDB; R05137.
```

```
PT   New protein H and gene encoding it -
PT   capable of binding specifically to human immunoglobulin of all
PT   sub-classes and non-human IgA, IgD, IgE or IgM.
PS   Claim 8; Fig 4; 28pp; English.
CC   Protein specifically binds to human IgG and is useful in assay,
CC   purification and adsorption of excess IgG from the blood in
CC   treatment of autoimmune diseases.
SQ   Sequence   1784 BP;   734 A;   343 C;   338 G;   369 T;

  Query Match           8.6%;  Score 15;  DB 1;  Length 1784;
  Best Local Similarity 100.0%;  Pred. No. 4.13e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1653 attagaaaataaagatt 1667
            ||||||||||||||||
Cp     96 ATTAGAAATAAGATT 82


RESULT   44
ID   T80109 standard; cDNA; 1828 BP.
AC   T80109;
DT   28-NOV-1997  (first entry)
DE   Mouse KF-1 gene fragment.
KW   Alzheimer's disease; KF-1; human; mouse; KF361 gene; antibody; ss.
OS   Mus musculus.
PN   J09215495-A.
PD   19-AUG-1997.
PF   20-JUN-1996; 181514.
PR   07-DEC-1995; JP-345659.
PA   (SUMU ) SUMITOMO SEIYAKU KK.
DR   WPI; 97-465073/43.
PT   A brain-specific expression gene - used in the diagnosis of
PT   Alzheimer's disease
PS   Example 4; Page 16-17; 23pp; Japanese.
CC   This sequence represents a fragment of a clone of the mouse KF-1 gene
CC   (see T80100 for wild type sequence). The full length sequence, T80098,
CC   and T80099 all represent cDNA sequences of the invention. These genes,
CC   their products and antibodies specific for the proteins are useful for
CC   the diagnosis of Alzheimer's disease.
SQ   Sequence   1828 BP;   547 A;   318 C;   373 G;   590 T;

  Query Match           8.6%;  Score 15;  DB 34;  Length 1828;
  Best Local Similarity 100.0%;  Pred. No. 4.13e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1791 aaaataaatttattt 1805
            ||||||||||||||||
Cp     55 AAAATAAATTATTTT 41


RESULT   45
```



```
ID   T73869 standard; DNA; 1871 BP.
AC   T73869;
DT   26-JAN-1998  (first entry)
DE   Cotton lipid transfer protein coding sequence.
KW   promoter; fibre-specific; transcriptional factor; promoter;
KW   altered phenotype; colour; melanin; indigo; ss.
OS   Gossypium hirsutum cv. coker 130.
FH   Key             Location/Qualifiers
FT   promoter        1..1104
FT                   /*tag=  a
FT                   /note= "contains promoter region"
FT   CDS             1104..1334
FT                   /*tag=  b
FT                   /product= lipid_transfer_protein
PN   WO9640924-A2.
PD   19-DEC-1996.
PF   07-JUN-1996; U09897.
PR   07-JUN-1995; US-480178.
PR   01-JUL-1996; ZA-005572.
PA   (CALJ ) CALGENE INC.
PI   Mcbride K,  Pear JR,  Perez-Grau L,  Stalker DM;
DR   WPI; 97-052325/05.
DR   P-PSDB; W21902.
PT   DNA construct contg. gene of interest controlled by cotton fibre
PT   transcriptional factor  - used to produce altered phenotype cotton
PT   fibre cells expressing genes affecting pigmentation
PS   Claim 27; Fig 7A-B; 95pp; English.
CC   The present DNA from cotton encodes a lipid transfer protein (Ltp). The
CC   Ltp promoter is active throughout fibre development. DNA constructs
CC   containing cotton fibre-specific transcriptional factor promoters are
CC   useful to produce cotton fibre cells with altered phenotype, especially
CC   altered colour. Genes involved in the production of melanin (e.g.
CC   tyrosinase gene and ORF438 encoded protein from Streptomyces
CC   antibioticus) and indigo (mono-oxygenase genes possibly in conjunction
CC   with a tryptophanase gene) are of interest. The promoters of the
CC   invention are reliable and permit expression of a protein selectively
CC   in cotton fibre to affect qualities such as fibre strength, length,
CC   colour and dyability as required. The construct and methods can also be
CC   used for the introduction of other advantageous genes into a cotton
CC   plant, e.g. a plant hormone. In particular, fibres from a plant producing
CC   coloured fibres may be used to produce yarns and/or fabrics that do not
CC   require dyeing.
SQ   Sequence   1871 BP;   542 A;   344 C;   356 G;   628 T;

  Query Match           8.6%;  Score 15;  DB 34;  Length 1871;
  Best Local Similarity 100.0%;  Pred. No. 4.13e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    269 cttgtttttatttttа 283
            ||||||||||||||||
Cp    157 CTTGTTTTTATTTTTA 143
```



```
RESULT   46
ID   T62651 standard; DNA; 2592 BP.
AC   T62651;
DT   24-NOV-1997  (first entry)
DE   DNA encoding type I carnitine palmitoyl transferase-like protein.
KW   Human; type I; carnitine palmitoyl transferase-like protein; CPTI;
KW   recombinant production; diagnosis; detection; disease; screen; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             40..2358
FT                   /*tag=  a
PN   J09009969-A.
PD   14-JAN-1997.
PF   29-JUN-1995; 163854.
PR   29-JUN-1995; JP-163854.
PA   (SAKA ) OTSUKA SEIYAKU KOGYO KK.
DR   WPI; 97-126426/12.
DR   P-PSDB; W14438.
PT   Human type I carnitine palmitoyl transferase (CPTI)-like protein
PT   gene  - used for expression and detection of CPTI-like protein for
PT   diagnosis of diseases related to CPTI expression
PS   Claim 2; Page 7-13; 13pp; Japanese.
CC   This DNA encodes human type I carnitine palmitoyl transferase (CPTI)-like
CC   protein (W14438). The gene can be used for the expression and the
CC   detection of human CPTI-like protein and thus can be used for the
CC   diagnosis of various diseases related to the protein. The CPTI-like
CC   protein and DNA sequences may also be used to screen agents that may be
CC   useful for the treatment of such diseases.
SQ   Sequence   2592 BP;   563 A;   769 C;   702 G;   556 T;

  Query Match           8.6%;  Score 15;  DB 33;  Length 2592;
  Best Local Similarity 100.0%;  Pred. No. 4.13e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2482 tgctttgaaatgtgt 2496
            ||||||||||||||||
Qy     66 TGCTTTGAAATGTGT 80


RESULT   47
ID   Q14945 standard; DNA; 2604 BP.
AC   Q14945;
DT   16-MAR-1992  (first entry)
DE   Rhizopus PGK2 promoter region.
KW   Phosphoglycerate kinase; yeast; promoter; ss.
OS   Rhizopus niveus IFO 4810.
FH   Key             Location/Qualifiers
FT   caat_signal     173..176
FT                   /*tag=  a
```



```
FT   tata_signal     234..238
FT                   /*tag=  b
PN   J03247286-A.
PD   05-NOV-1991.
PF   26-FEB-1990; 044715.
PR   26-FEB-1990; JP-044715.
PA   (NISO ) NIPPON SHOKUHIN KAK KK.
DR   WPI; 91-366333/50.
PT   Promoter gene for protein or yeast prodn. - comprises Rhizopus
PT   sp. derived phospho-glycerate kinase gene with concensus sequence
PT   upstream of initiation codon
PS   Example; Fig 2; 9pp; Japanese.
CC   This promoter sequence is an example of the claimed promoter
CC   sequences of the invention. It was derived from the PGK2 gene of R.
CC   niveus. The 6 "others" in the sequence are all unknown bases.
CC   See also Q14944.
SQ   Sequence   2604 BP;   769 A;   499 C;   503 G;   827 T;

  Query Match           8.6%;  Score 15;  DB 3;  Length 2604;
  Best Local Similarity 100.0%;  Pred. No. 4.13e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     60 aaaaaataaaacaagt 74
            ||||||||||||||||
Qy    144 AAAAATAAAACAAGT 158


RESULT   48
ID   Q70890 standard; cDNA; 3049 BP.
AC   Q70890;
DT   13-JUN-1995  (first entry)
DE   Tomato spotted wilt virus S RNA genome (cDNA copy).
KW   Tomato spotted wilt virus; TSWV; infection; resistance;
KW   nucleocapsid; Tospovirus; nonstructural gene; lettuce; ss.
OS   Tomato spotted wilt virus.
FH   Key             Location/Qualifiers
FT   cds             88..1491
FT                   /*tag=  a
FT                   /product= Non-structural (NS) protein.
FT   misc_difference 694..696
FT                   /*tag=  b
FT                   /transl_except= TTC encodes Glutamic acid.
PN   WO9416550-A.
PD   04-AUG-1994.
PF   27-JAN-1994; U01046.
PR   29-JAN-1993; US-010410.
PA   (CORR ) CORNELL RES FOUND INC.
PI   Gonsalves D,  Pang S;
DR   WPI; 94-263639/32.
DR   P-PSDB; R70890.
PT   Nucleotide sequences for tomato spotted wilt virus nucleo-capsid
```

```
PT   protein - used to confer resistance to Tospoviruses by
PT   transformation of susceptible plant genomes
PS   Claim 1; Page 27-29; 89pp; English.
CC   Viruses in the Tospovirus genus (of which the tomato spotted wilt
CC   virus is one), infect a wide variety of plant species. Tomato
CC   spotted wilt virus (TSWV) has a wide host range, infecting more than
CC   360 plant species of 50 families and causing large economic losses
CC   to vegetables and ornamental plants worldwide. By inserting the
CC   gene (or fragments) encoding the nucleocapsid of TSWV into
CC   susceptible plants, it is possible to render plants resistant to
CC   infection. This sequence is a cDNA copy of the complete S RNA
CC   genome of a brazilian isolate of TSWV designated TSWV-B. The
CC   nucleocapsid protein is encoded on the complementary strand. The
CC   gene sequence of the nucleocapsid protein and the protein itself are
CC   described in Q70891 and R56548 respectively.
SQ   Sequence  3049 BP;  932 A;  593 C;  513 G;  1011 T;

  Query Match          8.6%;  Score 15;  DB 12;  Length 3049;
  Best Local Similarity 100.0%;  Pred. No. 4.13e+01;
  Matches   15;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1841 ttgtttttatttttat 1855
          ||||||||||||||||
Cp    156 TTGTTTTATTTTTAT 142


RESULT   49
ID   Q20292 standard; DNA; 3095 BP.
AC   Q20292;
DT   15-APR-1992  (first entry)
DE   Sequence of plasmid p26 contg. the entire sequence for protein H
DE   plus a truncated M1 gene.
KW   Protein H; immunoglobulin G; IgG; antibody; autoimmune disease; ss.
OS   Streptococcus M type 1, strain 40/58.
FH   Key             Location/Qualifiers
FT   rbs             569..572
FT                   /*tag= a
FT   rbs             1857..1861
FT                   /*tag= b
FT   cds             582..1427
FT                   /*tag= c
FT                   /label= M1 gene
FT                   /note= "R20128"
FT   cds             1869..2999
FT                   /*tag= d
FT                   /label= protein H
FT                   /note= "R21662"
FT   cds             582..656
FT                   /*tag= e
FT                   /product= p16H1
FT   cds             657..791
```



```
FT                   /*tag= f
FT                   /product= C1
FT   cds             792..917
FT                   /*tag= g
FT                   /product= C2
FT   cds             918..1046
FT                   /*tag= h
FT                   /product= C3
FT   cds             1047..1109
FT                   /*tag= i
FT                   /product= C4
FT   cds             1110..1427
FT                   /*tag= j
FT                   /product= D
FT   cds             1869..1991
FT                   /*tag= k
FT                   /product= protein H S
FT   cds             1992..2231
FT                   /*tag= l
FT                   /product= A
FT   cds             2232..2342
FT                   /*tag= m
FT                   /product= B
FT   cds             2343..2465
FT                   /*tag= n
FT                   /product= C1
FT   cds             2466..2591
FT                   /*tag= o
FT                   /product= C2
FT   cds             2592..2717
FT                   /*tag= p
FT                   /product= C3
FT   cds             2718..2999
FT                   /*tag= q
FT                   /product= D
PN   WO9119740-A.
PD   26-DEC-1991.
PF   20-JUN-1991; SE0447.
PR   21-JUN-1990; SE-002212.
PA   (HIGH-) HIGHTECH RECEPTOR A.
PI   Schmidt KH, Akesson P, Cooney J, Bjorck L;
DR   WPI; 92-024366/03.
DR   P-PSDB; R20128, R21662.
PT   New IgG binding proteins H' lacking an albumin binding sequence -
PT   useful in purifica. of excess IgG from blood and to diagnose
PT   autoimmune diseases
PS   Disclosure; Fig 8; 37pp; English.
CC   The inventors claim a protein prod. by a strain of Gp.A Streptococci.
CC   The protein has the AA sequence of protein H but lacks at least some
CC   part of the C and D regions (responsible for binding albumin), esp.
CC   it lacks the whole of these regions and extends for AA1 to AA158.
```



```
CC   Compared with matural protein H, it is more specific and may be used
CC   as part of a kit for the binding, separation and identification of
CC   human IgG.  The same sequences appear in WO9119741.
SQ   Sequence  3095 BP;  1223 A;  595 C;  599 G;  678 T;

  Query Match          8.6%;  Score 15;  DB 3;  Length 3095;
  Best Local Similarity 100.0%;  Pred. No. 4.13e+01;
  Matches   15;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2964 attagaaataagatt 2978
          ||||||||||||||||
Cp     96 ATTAGAAATAAGATT 82


RESULT   50
ID   T80100 standard; cDNA; 3261 BP.
AC   T80100;
DT   25-NOV-1997  (first entry)
DE   Mouse KF-1 gene.
KW   Alzheimer's disease; KF-1; human; mouse; KF361 gene; antibody; ss.
OS   Mus musculus.
FH   Key             Location/Qualifiers
FT   cds             769..2820
FT                   /*tag= a
PN   J09215495-A.
PD   19-AUG-1997.
PF   20-JUN-1996; 181514.
PR   07-DEC-1995; JP-345659.
PA   (SUMU ) SUMITOMO SEIYAKU KK.
DR   WPI; 97-465073/43.
DR   P-PSDB; W26163.
PT   A brain-specific expression gene - used in the diagnosis of
PT   Alzheimer's disease
PS   Claim 3; Page 8-9; 23pp; Japanese.
CC   This sequence represents the mouse KF-1 gene. This sequence contains the
CC   open reading frame represented by T80101. This sequence, T80098, and
CC   T80099 all represent cDNA sequences of the invention. These sequences are
CC   all KF-1 genes, and contain the open reading frames shown in T80101,
CC   T80102 and T80103 respectively. These genes, their products and
CC   antibodies specific for the proteins are useful for the diagnosis of
CC   Alzheimer's disease.
SQ   Sequence  3261 BP;  790 A;  752 C;  825 G;  894 T;

  Query Match          8.6%;  Score 15;  DB 34;  Length 3261;
  Best Local Similarity 100.0%;  Pred. No. 4.13e+01;
  Matches   15;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   3241 aaaataaattatttt 3255
          ||||||||||||||||
Cp     55 AAAATAAATTATTTT 41
```



```
Search completed: Tue Dec 22 20:50:20 1998
Job time : 152 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Tue Dec 22 20:31:33 1998; MasPar time 286.43 Seconds
1338.737 Million cell updates/sec

Tabular output not generated.

Title: >US-08-530-112A-1219
Description: (1-174) from US08530112A.seq
Perfect Score: 174
N.A. Sequence: 1 GATCTGAGCAGGCAAAGCTC.....AAGTGAATTTGCAAGTGAAA 174
Comp: CTAGACTCGTCCGTTTCGAG.....TTCACTTAAACGTTCACTTT

Scoring table: TABLE jmetric
Gap 60

Kmatch STD : Dbase 0; Query 0

Searched: 567134 seqs, 1101898692 bases x 2

Post-processing: Minimum Match 0%
Listing first 1000 summaries

Database: embl55
1:em_ba 2:em_fun 3:em_htg 4:em_huml 5:em_hum2 6:em_in
7:em_om 8:em_or 9:em_ov 10:em_pat 11:em_ph 12:em_pl
13:em_ro 14:em_vi
Database: genbank107
15:gb_ba1 16:gb_ba2 17:gb_htg 18:gb_in 19:gb_om 20:gb_ov
21:gb_pat 22:gb_ph 23:gb_pl1 24:gb_pl2 25:gb_pr1
26:gb_pr2 27:gb_pr3 28:gb_ro 29:gb_st 30:gb_sts 31:gb_sy
32:gb_un 33:gb_vi

Statistics: Mean 7.935; Variance 3.352; scale 2.367

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
	1	174	100.0	6513	27	AF016833	Homo sapiens maltase-g	1.51e-142
	2	167	96.0	174	30	G25624	human STS EST51018.	1.14e-135
	3	167	96.0	174	30	G27635	human STS SHGC-32830.	1.14e-135
c	4	21	12.1	1267	23	KMARS1G	K.marxianus (ATCC 1242	1.94e-01
	5	20	11.5	76850	26	AC003109	Human DNA from overlap	8.92e-01
	6	20	11.5	124752	17	AC004986	*** SEQUENCING IN PROG	8.92e-01
c	7	20	11.5	213863	27	AC003037	Homo sapiens Xp22 BAC	8.92e-01
	8	19	10.9	247	25	HUMPROS05	Human protein S alpha	3.93e+00
	9	19	10.9	511	25	HUMPS05	Human S protein-alpha	3.93e+00
	10	19	10.9	540	25	HUMPSPS02	Human protein S-beta p	3.93e+00
	11	19	10.9	19108	23	SPAC1B9	S.pombe chromosome I c	3.93e+00
c	12	19	10.9	24725	18	CEK01C8	Caenorhabditis elegans	3.93e+00
c	13	19	10.9	27885	18	CEL2K488	Caenorhabditis elegans	3.93e+00
	14	19	10.9	29718	26	HS310H5	Human DNA sequence fro	3.93e+00
c	15	19	10.9	32793	18	CELC18F3	Caenorhabditis elegans	3.93e+00
	16	19	10.9	37145	18	CER09A8	Caenorhabditis elegans	3.93e+00
	17	19	10.9	39188	26	AC004754	Homo sapiens chromosom	3.93e+00
	18	19	10.9	111448	18	AC005334	Drosophila melanogaste	3.93e+00
c	19	19	10.9	126263	26	HS127B14	Human DNA sequence fro	3.93e+00
c	20	19	10.9	127270	26	AC004740	Homo sapiens PAC clone	3.93e+00
c	21	19	10.9	147709	17	AC004962	*** SEQUENCING IN PROG	3.93e+00
	22	19	10.9	149566	26	AC003081	Human BAC clone RG009H	3.93e+00
c	23	19	10.9	152115	17	HS404K8	Human DNA sequence ***	3.93e+00
	24	19	10.9	167034	17	HS77O6	Human DNA sequence ***	3.93e+00
c	25	19	10.9	211468	17	CEY1E3	Caenorhabditis elegans	3.93e+00
	26	18	10.3	400	30	G14255	human STS SHGC-7426 cl	1.66e+01
	27	18	10.3	436	18	BOU17773	Blatta orientalis mito	1.66e+01
c	28	18	10.3	662	27	HUMZD55D10	Homo sapiens full leng	1.66e+01
	29	18	10.3	4037	18	DDU41289	Dictyostelium discoide	1.66e+01
c	30	18	10.3	4678	28	MMYPT14	Mouse yptl gene for ra	1.66e+01
	31	18	10.3	4880	15	STYSINRPHO	Salmonella typhimurium	1.66e+01
c	32	18	10.3	5837	23	TOBPG13A	Nicotiana tabacum PG13	1.66e+01
	33	18	10.3	10702	33	RBFPS	Malignant rabbit fibro	1.66e+01
	34	18	10.3	11763	28	MUSBA	Mouse gene for basigin	1.66e+01
c	35	18	10.3	16550	16	MI11CG	Methanococcus jannasch	1.66e+01
c	36	18	10.3	23210	18	CET0C3	Caenorhabditis elegans	1.66e+01
c	37	18	10.3	26969	18	CET07F10	Caenorhabditis elegans	1.66e+01
	38	18	10.3	31975	24	SPCC553	S.pombe chromosome III	1.66e+01
	39	18	10.3	32346	18	CEB0240	Caenorhabditis elegans	1.66e+01
c	40	18	10.3	36087	18	CEL2C21	C. elegans cosmid ZC21	1.66e+01
	41	18	10.3	41754	18	CELB0238	Caenorhabditis elegans	1.66e+01
c	42	18	10.3	45131	18	CELF55A8	Caenorhabditis elegans	1.66e+01

	43	18	10.3	61290	24	AB017065	Arabidopsis thaliana g	1.66e+01
	44	18	10.3	74918	26	AC004140	Homo sapiens PAC clone	1.66e+01
	45	18	10.3	77137	18	AC004422	Drosophila melanogaste	1.66e+01
c	46	18	10.3	95167	23	ATAC003105	Arabidopsis thaliana c	1.66e+01
c	47	18	10.3	105127	26	AC004548	Homo sapiens PAC clone	1.66e+01
c	48	18	10.3	122638	26	AB001523	Homo sapiens gene for	1.66e+01
c	49	18	10.3	130019	17	HS194O8	Human DNA sequence ***	1.66e+01
	50	18	10.3	131047	17	AC002421	*** SEQUENCING IN PROG	1.66e+01
c	51	18	10.3	140060	17	AC005101	*** SEQUENCING IN PROG	1.66e+01
c	52	18	10.3	140679	17	HS246H3	Human DNA sequence ***	1.66e+01
c	53	18	10.3	159424	26	AF064865	Homo sapiens chromosom	1.66e+01
c	54	18	10.3	177328	17	HS462D8	Human DNA sequence ***	1.66e+01
c	55	18	10.3	185994	17	AC002042	*** SEQUENCING IN PROG	1.66e+01
c	56	18	10.3	187253	17	AC005486	*** SEQUENCING IN PROG	1.66e+01
	57	18	10.3	205559	23	ATFCA4	Arabidopsis thaliana D	1.66e+01
	58	18	10.3	224645	17	AC004153	*** SEQUENCING IN PROG	1.66e+01
c	59	18	10.3	244649	17	AC005308	*** SEQUENCING IN PROG	1.66e+01
	60	17	9.8	243	25	HS2D9T3	Human genomic DNA sequ	6.65e+01
	61	17	9.8	318	25	HS186F12R	H.sapiens CpG island D	6.65e+01
c	62	17	9.8	348	30	G24412	human STS WI-13825.	6.65e+01
c	63	17	9.8	390	30	G08476	human STS CHLC.GATA81C	6.65e+01
	64	17	9.8	400	30	G14387	human STS SHGC-9208 cl	6.65e+01
c	65	17	9.8	514	20	BB051654	Barbus barbus x Barbus	6.65e+01
	66	17	9.8	719	33	HIVU69471	HIV-1 patient z samp1	6.65e+01
c	67	17	9.8	721	30	G39227	Z20157 Zebrafish AB Da	6.65e+01
	68	17	9.8	731	18	TETDELAA	T.thermophila macronuc	6.65e+01
c	69	17	9.8	887	18	AF033378	Plasmodium falciparum	6.65e+01
	70	17	9.8	1021	20	OMLYRNA	O.mykiss mRNA for lyso	6.65e+01
c	71	17	9.8	1247	18	MDGTPS	M.domestica glutathion	6.65e+01
	72	17	9.8	1262	23	AB010900	Schizosaccharomyces po	6.65e+01
	73	17	9.8	1324	18	TETDELIA	T.thermophila micronuc	6.65e+01
c	74	17	9.8	1333	18	MDNF10	M.domestica nf10 gene.	6.65e+01
c	75	17	9.8	1381	23	SCYDL012C	S.cerevisiae chromosom	6.65e+01
	76	17	9.8	1427	15	HEAIIIM	Haemophilus aegyptius	6.65e+01
c	77	17	9.8	1979	23	SCYDL011C	S.cerevisiae chromosom	6.65e+01
c	78	17	9.8	2115	18	OTU19288	Oxytricha trifallax mi	6.65e+01
c	79	17	9.8	2207	23	LEPG5	Lycopersicon esculentu	6.65e+01
c	80	17	9.8	2207	21	I09042	Sequence 5 from Patent	6.65e+01
	81	17	9.8	2559	23	BV161PROM	B.vulgaris BV-16/1 gen	6.65e+01
	82	17	9.8	2938	18	TTU12971	Tetrahymena thermophil	6.65e+01
	83	17	9.8	3050	15	AF051375	Haemophilus influenzae	6.65e+01
c	84	17	9.8	3631	25	HUMCALP09	Homo sapiens calpastat	6.65e+01
c	85	17	9.8	3631	25	HUMCPA4	Human calpastatin gene	6.65e+01
c	86	17	9.8	3927	23	MIWMCOX	W.makrii mitochondrial	6.65e+01
	87	17	9.8	3951	25	HSU09178	Human dihydropyrimidin	6.65e+01
	88	17	9.8	4241	15	AF051374	Fusobacterium nucleatu	6.65e+01
	89	17	9.8	4409	26	HSU20938	Human lymphocyte dihyd	6.65e+01
c	90	17	9.8	6499	15	AF027167	Streptococcus thermoph	6.65e+01
c	91	17	9.8	7456	23	TOMPGAAA	Tomato polygalacturona	6.65e+01
c	92	17	9.8	7456	23	LEPOLYGA	Tomato gene for cell w	6.65e+01
c	93	17	9.8	7471	32	PCCCHRAF	Capsicum annuum chromo	6.65e+01

	94	17	9.8	8335	20	AF002166	Xenopus laevis Ig mu h	6.65e+01
c	95	17	9.8	10995	15	STRMSM	Streptococcus mutans r	6.65e+01
c	96	17	9.8	11772	18	PFSC03022	Plasmodium falciparum	6.65e+01
	97	17	9.8	13132	18	PFSC03021	Plasmodium falciparum	6.65e+01
c	98	17	9.8	14320	18	CELPAR2	Caenorhabditis elegans	6.65e+01
c	99	17	9.8	16343	18	AMPGENOM	Apis mellifera ligusti	6.65e+01
c	100	17	9.8	21552	18	CEM04D8	Caenorhabditis elegans	6.65e+01
c	101	17	9.8	22197	15	D83967	Bacillus subtilis geno	6.65e+01
	102	17	9.8	22880	18	CELT04A6	Caenorhabditis elegans	6.65e+01
c	103	17	9.8	23449	24	AF072725	Zea mays starch branch	6.65e+01
	104	17	9.8	25943	18	CEC38C6	Caenorhabditis elegans	6.65e+01
	105	17	9.8	26590	18	CELF47B7	Caenorhabditis elegans	6.65e+01
	106	17	9.8	26711	18	CELF28A12	Caenorhabditis elegans	6.65e+01
c	107	17	9.8	26790	18	CEC14H10	Caenorhabditis elegans	6.65e+01
	108	17	9.8	30698	18	CER11D1	Caenorhabditis elegans	6.65e+01
	109	17	9.8	30961	18	CEF47B10	Caenorhabditis elegans	6.65e+01
	110	17	9.8	32056	18	CELT28C12	Caenorhabditis elegans	6.65e+01
	111	17	9.8	32527	18	CELT09A12	Caenorhabditis elegans	6.65e+01
	112	17	9.8	32677	18	CEC08F8	Caenorhabditis elegans	6.65e+01
	113	17	9.8	32750	18	CELZK484	Caenorhabditis elegans	6.65e+01
	114	17	9.8	34860	18	CELR173	Caenorhabditis elegans	6.65e+01
	115	17	9.8	35549	18	CELH19M22	Caenorhabditis elegans	6.65e+01
	116	17	9.8	36383	26	HS86C3	Human DNA sequence fro	6.65e+01
c	117	17	9.8	36767	26	AC004646	Homo sapiens chrosom	6.65e+01
c	118	17	9.8	36874	26	AC005336	Homo sapien chromosome	6.65e+01
c	119	17	9.8	37738	18	CELF32B5	Caenorhabditis elegans	6.65e+01
c	120	17	9.8	38103	18	CELR09A1	Caenorhabditis elegans	6.65e+01
c	121	17	9.8	38546	18	CICOS2	Ciona intestinalis DNA	6.65e+01
	122	17	9.8	39121	18	CEF09C6	Caenorhabditis elegans	6.65e+01
	123	17	9.8	40912	18	CELC41D11	Caenorhabditis elegans	6.65e+01
c	124	17	9.8	41372	23	SCCHRIV42	S. cerevisiae chromoso	6.65e+01
	125	17	9.8	41533	26	AC004523	Homo sapiens chromosom	6.65e+01
c	126	17	9.8	41564	18	CELF22F7	Caenorhabditis elegans	6.65e+01
c	127	17	9.8	42393	17	AC005780	Homo sapiens chromosom	6.65e+01
c	128	17	9.8	44371	18	CEC47E8	Caenorhabditis elegans	6.65e+01
c	129	17	9.8	44888	25	HS130N4B	Human DNA sequence fro	6.65e+01
	130	17	9.8	45193	17	AC005672	*** SEQUENCING IN PROG	6.65e+01
c	131	17	9.8	45678	25	HSV351F8	Human DNA sequence fro	6.65e+01
c	132	17	9.8	53042	17	HS361H15	Human DNA sequence ***	6.65e+01
	133	17	9.8	55328	23	PWU02970	Prototheca wickerhamii	6.65e+01
c	134	17	9.8	61849	18	AC004306	Drosophila melanogaste	6.65e+01
c	135	17	9.8	63769	17	CER08A2	Caenorhabditis elegans	6.65e+01
	136	17	9.8	64827	24	AB016873	Arabidopsis thaliana g	6.65e+01
	137	17	9.8	71168	17	AC005633	*** SEQUENCING IN PROG	6.65e+01
	138	17	9.8	71803	17	AC005504	*** SEQUENCING IN PROG	6.65e+01
c	139	17	9.8	79073	23	AB008271	Arabidopsis thaliana g	6.65e+01
	140	17	9.8	81837	26	AC004775	Homo sapiens chromosom	6.65e+01
	141	17	9.8	83822	24	AC004260	Arabidopsis thaliana c	6.65e+01
c	142	17	9.8	84170	26	HS130N4	Human DNA sequence fro	6.65e+01
c	143	17	9.8	84883	18	AC004118	Drosophila melanogaste	6.65e+01
c	144	17	9.8	89840	24	ATAC004705	Arabidopsis thaliana c	6.65e+01

c	145	17	9.8	92651	26	AC002527	Human BAC clone GS180J 6.65e+01
c	146	17	9.8	95506	26	AC002407	Human Chromosome X, co 6.65e+01
c	147	17	9.8	108390	17	HS312L17	Human DNA sequence *** 6.65e+01
	148	17	9.8	110000	17	CEY102A5	Caenorhabditis elegans 6.65e+01
c	149	17	9.8	112392	26	AC004028	Human PAC clone DJ0800 6.65e+01
c	150	17	9.8	114784	27	HS1189B24	Human DNA sequence fro 6.65e+01
	151	17	9.8	115641	24	F8K4	Arabidopsis thaliana c 6.65e+01
c	152	17	9.8	121019	26	AF042091	Homo sapiens chromosom 6.65e+01
c	153	17	9.8	121024	23	CHMPXX	Liverwort Marchantia p 6.65e+01
c	154	17	9.8	122325	27	HS1044O17	Human DNA sequence *** 6.65e+01
	155	17	9.8	127247	25	HS326L13	Human DNA sequence fro 6.65e+01
	156	17	9.8	127816	17	AC004822	*** SEQUENCING IN PROG 6.65e+01
c	157	17	9.8	134199	15	SYCSLRF	Synechocystis sp. PCC6 6.65e+01
	158	17	9.8	134534	17	HS1163J1	Human DNA sequence *** 6.65e+01
c	159	17	9.8	141120	26	AC002528	Human BAC clone GS542D 6.65e+01
c	160	17	9.8	146831	26	HS523C21	Homo sapiens DNA seque 6.65e+01
	161	17	9.8	148198	26	HS438G17	Human DNA sequence fro 6.65e+01
c	162	17	9.8	152383	26	HSF18108	Homo sapiens chromosom 6.65e+01
	163	17	9.8	152386	17	AC005711	*** SEQUENCING IN PROG 6.65e+01
	164	17	9.8	153053	17	HSAJ9612	Homo sapiens *** SEQUE 6.65e+01
	165	17	9.8	154252	26	AC004062	Homo sapiens chromosom 6.65e+01
c	166	17	9.8	154788	27	HS325F22	Human DNA sequence fro 6.65e+01
c	167	17	9.8	155164	26	AC005165	Homo sapiens BAC clone 6.65e+01
c	168	17	9.8	155407	26	AF042090	Homo sapiens chromosom 6.65e+01
	169	17	9.8	167217	26	AC002350	Homo sapiens 12q24 PAC 6.65e+01
c	170	17	9.8	171112	26	AC002452	Human BAC clone GS310A 6.65e+01
c	171	17	9.8	177821	17	AC005030	*** SEQUENCING IN PROG 6.65e+01
c	172	17	9.8	180527	17	HS269M15	Human DNA sequence *** 6.65e+01
c	173	17	9.8	180565	17	HS149A16	Human DNA sequence *** 6.65e+01
	174	17	9.8	185364	17	AC005477	*** SEQUENCING IN PROG 6.65e+01
	175	17	9.8	191439	17	AC005040	*** SEQUENCING IN PROG 6.65e+01
c	176	17	9.8	192335	17	AC005042	*** SEQUENCING IN PROG 6.65e+01
	177	17	9.8	197273	26	AC002461	Human BAC clone RG204I 6.65e+01
	178	17	9.8	198570	26	HS343C1	Human DNA sequence fro 6.65e+01
	179	17	9.8	200000	17	AC004618	*** SEQUENCING IN PROG 6.65e+01
	180	17	9.8	208430	15	BSUB0005	Bacillus subtilis comp 6.65e+01
	181	17	9.8	209071	18	AC005285	Drosophila melanogaste 6.65e+01
	182	17	9.8	210364	17	AC005537	*** SEQUENCING IN PROG 6.65e+01
c	183	17	9.8	221273	17	CEY40H4	Caenorhabditis elegans 6.65e+01
	184	17	9.8	225676	17	HS199L16	Human DNA sequence *** 6.65e+01
	185	17	9.8	231329	17	AC002353	*** SEQUENCING IN PROG 6.65e+01
	186	17	9.8	307360	17	HS796E4	Human DNA sequence *** 6.65e+01
	187	17	9.8	323900	17	HS402G11	Human DNA sequence *** 6.65e+01
c	188	17	9.8	329751	17	HS407F11	Human DNA sequence *** 6.65e+01
c	189	16	9.2	123	25	HSPA2SG12	H.sapiens flow-sorted 2.52e+02
c	190	16	9.2	134	30	HUMSWX810	Human chromosome X STS 2.52e+02
c	191	16	9.2	250	30	G15412	human STS SHGC-16753. 2.52e+02
	192	16	9.2	251	20	CHKCOLA04	Chicken alpha-2 collag 2.52e+02
	193	16	9.2	257	30	G05635	human STS WI-6074. 2.52e+02
	194	16	9.2	303	33	AF058112	HIV-1 patient 344 from 2.52e+02
	195	16	9.2	303	33	AF058113	HIV-1 patient 375 from 2.52e+02

	196	16	9.2	346	25	HSPA20G6	H.sapiens flow-sorted 2.52e+02
c	197	16	9.2	358	30	G25740	human STS EST386001. 2.52e+02
c	198	16	9.2	387	30	G29752	human STS SHGC-35172, 2.52e+02
	199	16	9.2	446	30	G23099	human STS WI-14076. 2.52e+02
	200	16	9.2	514	15	BUHARGSD	Buchnera aphidicola Ar 2.52e+02
c	201	16	9.2	594	28	MMU010694	mus musculus Cyrnl gen 2.52e+02
	202	16	9.2	626	23	LUU94847	Linum usitatissimum ch 2.52e+02
c	203	16	9.2	668	21	E00171	cDNA encoding human in 2.52e+02
c	204	16	9.2	668	21	E00098	DNA coding of interfer 2.52e+02
c	205	16	9.2	668	21	A08452	DNA sequence for CG-pB 2.52e+02
c	206	16	9.2	668	21	A07160	Artificial sequence fo 2.52e+02
	207	16	9.2	679	18	PFU31953	Plasmodium falciparum 2.52e+02
c	208	16	9.2	680	18	DD58RNA	D.discoideum 5S riboso 2.52e+02
	209	16	9.2	719	20	AF005923	Oncorhynchus mykiss gr 2.52e+02
	210	16	9.2	723	23	NPDTNP1	N.plumbaginifolia mRNA 2.52e+02
c	211	16	9.2	757	25	HSIFR5	Human messenger RNA fo 2.52e+02
c	212	16	9.2	761	21	I03922	Sequence 51 from Paten 2.52e+02
c	213	16	9.2	767	30	G39595	Z6634 Zebrafish AB Dan 2.52e+02
c	214	16	9.2	792	18	MDGST4A	M.domesticus mRNA for 2.52e+02
c	215	16	9.2	797	27	HSTHSMT4	Human thiopurine methy 2.52e+02
c	216	16	9.2	828	25	HSIFR4	Messenger RNA for huma 2.52e+02
c	217	16	9.2	836	21	E00012	DNA coding for human i 2.52e+02
c	218	16	9.2	836	21	E00017	DNA coding for human i 2.52e+02
c	219	16	9.2	836	21	A01302	Nucleotide sequence 21 2.52e+02
c	220	16	9.2	838	27	HUMIFNB	Human interferon-beta 2.52e+02
	221	16	9.2	839	18	DROGPAS1	D.melanogaster G0 prot 2.52e+02
c	222	16	9.2	840	25	HUMIFNB1	Human interferon beta- 2.52e+02
c	223	16	9.2	869	21	A01299	Nucleotide sequence 19 2.52e+02
	224	16	9.2	883	18	AF019228	Spraguea lophii chitin 2.52e+02
c	225	16	9.2	899	21	A00131	Synthetic DNA for huma 2.52e+02
c	226	16	9.2	899	21	A01300	Nucleotide sequence 24 2.52e+02
c	227	16	9.2	899	21	A01301	Nucleotide sequence 20 2.52e+02
c	228	16	9.2	934	18	DROMSP316	D.melanogaster sex-spe 2.52e+02
c	229	16	9.2	997	25	HSIFD6	Gene for human fibrobl 2.52e+02
c	230	16	9.2	1012	21	E01154	Promoter of saccharomy 2.52e+02
	231	16	9.2	1309	27	S72871	GATA-2=transcription f 2.52e+02
	232	16	9.2	1414	18	PFTRXR	P.falciparum thioredox 2.52e+02
c	233	16	9.2	1456	18	CEU21282	Caenorhabditis elegans 2.52e+02
	234	16	9.2	1514	28	MUSOSF1	Mouse mRNA for OSF-1. 2.52e+02
c	235	16	9.2	1522	28	D50586	Mus musculus mRNA for 2.52e+02
	236	16	9.2	1542	21	E03564	cDNA encoding OSF-1 ca 2.52e+02
c	237	16	9.2	1580	25	HSB98D12E	Human partial cDNA seq 2.52e+02
	238	16	9.2	1629	18	MIDTTR01	Drosophila yakuba mito 2.52e+02
c	239	16	9.2	1630	23	MTWSORFA	W.suaveolens mitochond 2.52e+02
c	240	16	9.2	1695	20	S98891	hat-2=bcr oncogene hom 2.52e+02
c	241	16	9.2	1752	28	RNU12006	Rattus norvegicus puta 2.52e+02
c	242	16	9.2	1758	19	BMDNA3	Blue Whale common ceta 2.52e+02
c	243	16	9.2	1790	23	SCYBR273C	S.cerevisiae chromosom 2.52e+02
c	244	16	9.2	1834	21	I00848	Sequence 1 from Patent 2.52e+02
c	245	16	9.2	1835	21	E00038	DNA coding for human i 2.52e+02
c	246	16	9.2	1835	25	HSIFD4	Human gene for fibrobl 2.52e+02

c	247	16	9.2	1835	26	HUMIFNB1F	Human fibroblast inter 2.52e+02
c	248	16	9.2	1955	23	SCYBR272C	S.cerevisiae chromosom 2.52e+02
	249	16	9.2	2017	33	ADLGPROB	Adelaide River virus G 2.52e+02
	250	16	9.2	2085	18	PFTHIORED	P.falciparum mRNA for 2.52e+02
	251	16	9.2	2169	21	A59666	Sequence 28 from Paten 2.52e+02
	252	16	9.2	2169	21	A59615	Sequence 28 from Paten 2.52e+02
c	253	16	9.2	2169	21	A59601	Sequence 14 from Paten 2.52e+02
c	254	16	9.2	2169	21	A59652	Sequence 14 from Paten 2.52e+02
c	255	16	9.2	2214	21	A51791	Sequence 26 from Paten 2.52e+02
c	256	16	9.2	2236	28	AF064706	Rattus norvegicus puta 2.52e+02
c	257	16	9.2	2243	19	BOVM6PR2	Bovine 46 kd mannose-6 2.52e+02
	258	16	9.2	2323	28	MMMFH1	M.musculus mRNA for MF 2.52e+02
	259	16	9.2	2433	23	SCURP2G	S.cerevisiae genes for 2.52e+02
	260	16	9.2	2484	23	ATPATY3	A. thaliana genes for 2.52e+02
	261	16	9.2	2610	25	HUMGATA2A	Human transcription fa 2.52e+02
	262	16	9.2	2712	28	MMICRMFH1	M.musculus mRNA for me 2.52e+02
	263	16	9.2	2729	25	HUMGATA	Human GATA-binding pro 2.52e+02
	264	16	9.2	2772	26	AF016050	Homo sapiens vascular 2.52e+02
	265	16	9.2	2961	28	MMGAS1MR	M.musculus gas1 mRNA. 2.52e+02
	266	16	9.2	2961	18	TETRPL37	T.thermophila ribosoma 2.52e+02
	267	16	9.2	2978	26	HUMXQJ	Human mRNA, Xq termina 2.52e+02
	268	16	9.2	3133	28	AB000096	Mouse mRNA for GATA-2 2.52e+02
c	269	16	9.2	3282	27	HSU95367	Human GABA-A receptor 2.52e+02
c	270	16	9.2	3346	23	SOYGLYMAX	Glycine max vspA gene, 2.52e+02
c	271	16	9.2	3577	23	LEPOOXE	L.esculentum gene for 2.52e+02
c	272	16	9.2	3714	28	MMCRISP1P	M.musculus CRISP-1 gen 2.52e+02
c	273	16	9.2	3763	16	ECOASLAB	Escherichia coli putat 2.52e+02
	274	16	9.2	3914	33	ADLGPROA	Adelaide River virus G 2.52e+02
c	275	16	9.2	4085	23	S81751	RNA polymerase homolog 2.52e+02
c	276	16	9.2	4697	28	MUSPD	Mouse mRNA for peptidy 2.52e+02
c	277	16	9.2	4718	19	OAKERB2A	Sheep B2A and B2D gene 2.52e+02
	278	16	9.2	4860	18	CEVF39H2L	Caenorhabditis elegans 2.52e+02
	279	16	9.2	4904	23	CHEGPSB	Euglena gracilis chlor 2.52e+02
	280	16	9.2	5036	15	LLMOBAMAT	L.lactis DNA for mobA 2.52e+02
	281	16	9.2	5253	18	PFSC04092	Plasmodium falciparum 2.52e+02
	282	16	9.2	5359	18	AF092920	Chlorohydra viridissim 2.52e+02
	283	16	9.2	5380	18	SMPRDIISA	S.mansoni protein disu 2.52e+02
	284	16	9.2	5616	25	HSSISG1	Human sis proto-oncoge 2.52e+02
c	285	16	9.2	5641	31	AF016404	Chloroplast transforma 2.52e+02
	286	16	9.2	5770	27	AB007972	Homo sapiens mRNA, chr 2.52e+02
c	287	16	9.2	6019	16	AF022812	Dehalospirillum multiv 2.52e+02
	288	16	9.2	6021	28	MMMFHEAD1	M.musculus MFH-1 gene. 2.52e+02
c	289	16	9.2	6053	18	CELC30H7	Caenorhabditis elegans 2.52e+02
c	290	16	9.2	6828	33	PPNRNA	Pepper ringspot viral 2.52e+02
	291	16	9.2	6898	15	LLU50902	Lactococcus lactis lac 2.52e+02
c	292	16	9.2	7218	15	CVPME131	C.viguieri phS gene, g 2.52e+02
c	293	16	9.2	7390	33	EV70CG	Enterovirus 70 genomic 2.52e+02
	294	16	9.2	7421	18	PFSC04063	Plasmodium falciparum 2.52e+02
c	295	16	9.2	7514	15	U39719	Mycoplasma genitalium 2.52e+02
	296	16	9.2	7547	19	RABHBBB1B1	Rabbit beta1-globin ge 2.52e+02
c	297	16	9.2	7820	18	AF040725	Cryptosporidium parvum 2.52e+02

c	298	16	9.2	9271	18	CELY40D12A	Caenorhabditis elegans 2.52e+02
	299	16	9.2	9757	15	U67599	Methanococcus jannasch 2.52e+02
	300	16	9.2	10028	23	PSBLOX13	P.sativum (Birte) Lox1 2.52e+02
c	301	16	9.2	10596	18	PFSC04056	Plasmodium falciparum 2.52e+02
	302	16	9.2	10626	18	PFSC03027	Plasmodium falciparum 2.52e+02
	303	16	9.2	10647	18	CEH05G16	Caenorhabditis elegans 2.52e+02
	304	16	9.2	10784	16	AE001290	Chlamydia trachomatis 2.52e+02
c	305	16	9.2	10901	18	PFSC03070	Plasmodium falciparum 2.52e+02
c	306	16	9.2	11478	21	A59616	Sequence 29 from Paten 2.52e+02
c	307	16	9.2	11478	21	A59667	Sequence 29 from Paten 2.52e+02
c	308	16	9.2	11568	15	ECAE000456	Escherichia coli K-12 2.52e+02
c	309	16	9.2	11577	15	ECAE000443	Escherichia coli K-12 2.52e+02
	310	16	9.2	11843	26	HS1109B5	Human DNA sequence fro 2.52e+02
c	311	16	9.2	12742	15	U67556	Methanococcus jannasch 2.52e+02
c	312	16	9.2	13132	18	PFSC03021	Plasmodium falciparum 2.52e+02
c	313	16	9.2	14001	18	PFCOMPIRB	P.falciparum complete 2.52e+02
	314	16	9.2	14890	18	CEE01F3	Caenorhabditis elegans 2.52e+02
	315	16	9.2	15152	15	AE001171	Borrelia burgdorferi (2.52e+02
	316	16	9.2	15376	25	HSU82210	Human chromosome 10 cl 2.52e+02
	317	16	9.2	15875	18	CELF49E8	Caenorhabditis elegans 2.52e+02
	318	16	9.2	16019	18	MIDYRRN	Drosophila yakuba mito 2.52e+02
	319	16	9.2	16402	19	MIBMCG	B.musculus mitochondri 2.52e+02
c	320	16	9.2	16986	18	CEM153	Caenorhabditis elegans 2.52e+02
	321	16	9.2	19480	18	CET08A11	Caenorhabditis elegans 2.52e+02
c	322	16	9.2	21217	18	CELF54G2	Caenorhabditis elegans 2.52e+02
c	323	16	9.2	21737	19	SSTMP835	S.scrofa tmp83.5 gene. 2.52e+02
	324	16	9.2	22333	18	CELC02F5	C. elegans cosmid C02F 2.52e+02
	325	16	9.2	22920	18	CET14G10	Caenorhabditis elegans 2.52e+02
	326	16	9.2	23911	15	BS233DEG	B.subtilis 23.9kb frag 2.52e+02
	327	16	9.2	25195	18	CELC24A3	Caenorhabditis elegans 2.52e+02
c	328	16	9.2	25195	18	CELC24A3	Caenorhabditis elegans 2.52e+02
c	329	16	9.2	25523	18	CEF39D8	Caenorhabditis elegans 2.52e+02
c	330	16	9.2	25560	18	CEZK809	Caenorhabditis elegans 2.52e+02
c	331	16	9.2	26667	18	CEF14H3	Caenorhabditis elegans 2.52e+02
c	332	16	9.2	26802	17	AC005708	*** SEQUENCING IN PROG 2.52e+02
	333	16	9.2	28000	18	CEC15H7	Caenorhabditis elegans 2.52e+02
c	334	16	9.2	28409	18	CELC06A6	Caenorhabditis elegans 2.52e+02
c	335	16	9.2	28690	18	CEW08G11	Caenorhabditis elegans 2.52e+02
	336	16	9.2	28740	18	CER11	Caenorhabditis elegans 2.52e+02
	337	16	9.2	28954	18	CELC03G5	Caenorhabditis elegans 2.52e+02
	338	16	9.2	29039	18	CEF56F3	Caenorhabditis elegans 2.52e+02
	339	16	9.2	29276	23	YSCH9433	Saccharomyces cerevisi 2.52e+02
	340	16	9.2	29716	18	CELC39H7	Caenorhabditis elegans 2.52e+02
	341	16	9.2	29731	18	CEB0391	Caenorhabditis elegans 2.52e+02
c	342	16	9.2	30281	18	CELF41G3	Caenorhabditis elegans 2.52e+02
c	343	16	9.2	30323	18	CEF58E6	Caenorhabditis elegans 2.52e+02
c	344	16	9.2	31001	18	CEZK1307	Caenorhabditis elegans 2.52e+02
	345	16	9.2	31180	18	DMC155E2	Drosophila melanogaste 2.52e+02
c	346	16	9.2	31230	18	CEF07H5	Caenorhabditis elegans 2.52e+02
	347	16	9.2	31455	18	CEF48C11	Caenorhabditis elegans 2.52e+02
	348	16	9.2	31821	18	CEF58G11	Caenorhabditis elegans 2.52e+02

	349	16	9.2	31929	25	HSL139H8	Human DNA sequence fro 2.52e+02
c	350	16	9.2	32978	18	CEC09D8	Caenorhabditis elegans 2.52e+02
	351	16	9.2	33013	18	DMU87286	Drosophila melanogaste 2.52e+02
c	352	16	9.2	33238	18	CET06E6	Caenorhabditis elegans 2.52e+02
	353	16	9.2	33573	18	CELF09E10	Caenorhabditis elegans 2.52e+02
	354	16	9.2	33651	18	CELF44E2	C. elegans cosmid F44E 2.52e+02
	355	16	9.2	33783	18	CELW09G10	Caenorhabditis elegans 2.52e+02
	356	16	9.2	34100	18	CEM110	Caenorhabditis elegans 2.52e+02
c	357	16	9.2	34347	18	CELK08A8	Caenorhabditis elegans 2.52e+02
	358	16	9.2	34349	18	CELK10C2	Caenorhabditis elegans 2.52e+02
c	359	16	9.2	34372	18	CELZC250	Caenorhabditis elegans 2.52e+02
	360	16	9.2	34511	18	CELZC404	Caenorhabditis elegans 2.52e+02
	361	16	9.2	35033	18	CELT05C7	Caenorhabditis elegans 2.52e+02
	362	16	9.2	35103	18	CELDYNEINH	Caenorhabditis elegans 2.52e+02
	363	16	9.2	35220	18	CET04D3	Caenorhabditis elegans 2.52e+02
c	364	16	9.2	35253	18	CELC24A8	Caenorhabditis elegans 2.52e+02
c	365	16	9.2	35350	18	CEC30H6	Caenorhabditis elegans 2.52e+02
	366	16	9.2	35365	18	CELF27D9	Caenorhabditis elegans 2.52e+02
	367	16	9.2	35820	18	CELT21E12	Caenorhabditis elegans 2.52e+02
c	368	16	9.2	36405	17	AC005673	*** SEQUENCING IN PROG 2.52e+02
c	369	16	9.2	36428	26	AC005362	Homo sapiens chromosom 2.52e+02
c	370	16	9.2	36587	18	AC003700	Drosophila melanogaste 2.52e+02
c	371	16	9.2	36697	27	AF038667	Homo sapiens clone c10 2.52e+02
	372	16	9.2	36864	18	CELT03D3	Caenorhabditis elegans 2.52e+02
	373	16	9.2	36873	18	CELC34E10	Caenorhabditis elegans 2.52e+02
c	374	16	9.2	37019	18	CELF27C1	Caenorhabditis elegans 2.52e+02
	375	16	9.2	37218	26	HSN11D4	Human DNA sequence fro 2.52e+02
	376	16	9.2	37445	18	CELC12D12	Caenorhabditis elegans 2.52e+02
c	377	16	9.2	37881	18	CELK09H9	Caenorhabditis elegans 2.52e+02
	378	16	9.2	37888	25	HSU74C11	Human DNA sequence fro 2.52e+02
	379	16	9.2	37940	25	HUM19ANON1	Human DNA from cosmid 2.52e+02
	380	16	9.2	38117	27	HSN10C3	Human DNA sequence fro 2.52e+02
	381	16	9.2	38162	26	AC004095	Human Cosmid g1248a143 2.52e+02
c	382	16	9.2	38335	18	CELF42G8	Caenorhabditis elegans 2.52e+02
	383	16	9.2	38514	18	CEC27A7	Caenorhabditis elegans 2.52e+02
c	384	16	9.2	38926	25	HSF45C1B	Human DNA sequence fro 2.52e+02
	385	16	9.2	39192	18	CELZC123	Caenorhabditis elegans 2.52e+02
c	386	16	9.2	39534	18	CEZK643	Caenorhabditis elegans 2.52e+02
c	387	16	9.2	39565	18	CEK08E3	Caenorhabditis elegans 2.52e+02
	388	16	9.2	39626	18	CELF53A9	Caenorhabditis elegans 2.52e+02
c	389	16	9.2	39711	18	CELR13D11	Caenorhabditis elegans 2.52e+02
	390	16	9.2	39962	27	AC005616	Homo sapiens chromosom 2.52e+02
c	391	16	9.2	40852	18	CEF11C1	Caenorhabditis elegans 2.52e+02
c	392	16	9.2	41014	18	CEF01G10	Caenorhabditis elegans 2.52e+02
	393	16	9.2	41778	18	CELW04G7	Caenorhabditis elegans 2.52e+02
	394	16	9.2	42169	17	CET16G1	Caenorhabditis elegans 2.52e+02
c	395	16	9.2	42171	18	CELC18E3	Caenorhabditis elegans 2.52e+02
	396	16	9.2	42398	17	AC003095	*** SEQUENCING IN PROG 2.52e+02
	397	16	9.2	42493	27	HSCB1E7	Human DNA sequence fro 2.52e+02
c	398	16	9.2	42632	18	CEF52H3	Caenorhabditis elegans 2.52e+02
	399	16	9.2	42934	23	SPBC947	S.pombe chromosome II 2.52e+02

c	400	16	9.2	43910	18	CELC06G3	Caenorhabditis elegans 2.52e+02
c	401	16	9.2	44298	18	CELC07G1	Caenorhabditis elegans 2.52e+02
c	402	16	9.2	44307	18	CELC31B8	Caenorhabditis elegans 2.52e+02
	403	16	9.2	44594	19	RABBGLOB	Rabbit beta-like globi 2.52e+02
	404	16	9.2	45189	18	CELC34F11	Caenorhabditis elegans 2.52e+02
	405	16	9.2	46105	18	CELC55C3	Caenorhabditis elegans 2.52e+02
	406	16	9.2	46249	18	CELF53G2	Caenorhabditis elegans 2.52e+02
c	407	16	9.2	48645	27	AC005575	Homo sapiens chromosom 2.52e+02
	408	16	9.2	48985	17	CEW03G8	Caenorhabditis elegans 2.52e+02
	409	16	9.2	62257	17	AC005469	*** SEQUENCING IN PROG 2.52e+02
	410	16	9.2	62927	17	AB018119	*** SEQUENCING IN PROG 2.52e+02
c	411	16	9.2	69119	17	AC005418	*** SEQUENCING IN PROG 2.52e+02
c	412	16	9.2	70113	26	AC002478	Human BAC clone GS002F 2.52e+02
c	413	16	9.2	70315	26	AF035397	Homo sapiens cosmids Q 2.52e+02
c	414	16	9.2	70665	26	HS696H22	Human DNA sequence fro 2.52e+02
c	415	16	9.2	71803	17	AC005504	*** SEQUENCING IN PROG 2.52e+02
	416	16	9.2	72732	18	AC004715	Drosophila melanogaste 2.52e+02
c	417	16	9.2	75184	26	AC005218	Homo sapiens chromosom 2.52e+02
	418	16	9.2	75551	18	AC005111	Drosophila melanogaste 2.52e+02
	419	16	9.2	75709	23	AB010693	Arabidopsis thaliana g 2.52e+02
c	420	16	9.2	79899	23	AB011474	Arabidopsis thaliana g 2.52e+02
c	421	16	9.2	80124	18	AC002444	Drosophila melanogaste 2.52e+02
c	422	16	9.2	80315	23	AB005232	Arabidopsis thaliana g 2.52e+02
c	423	16	9.2	80423	18	DROABDB	Drosophila melanogaste 2.52e+02
c	424	16	9.2	80659	27	AC004553	Homo sapiens Xp22 bin 2.52e+02
c	425	16	9.2	81007	26	HSAF003627	Homo sapiens cosmids Q 2.52e+02
c	426	16	9.2	81421	17	CEC53C7	Caenorhabditis elegans 2.52e+02
	427	16	9.2	81741	26	AC005367	Homo sapiens chromosom 2.52e+02
c	428	16	9.2	82251	26	AC004506	Homo sapiens chromosom 2.52e+02
	429	16	9.2	82294	17	AC005125	*** SEQUENCING IN PROG 2.52e+02
	430	16	9.2	82594	17	AC005311	*** SEQUENCING IN PROG 2.52e+02
c	431	16	9.2	82594	17	AC005311	*** SEQUENCING IN PROG 2.52e+02
c	432	16	9.2	84553	17	AC005436	*** SEQUENCING IN PROG 2.52e+02
c	433	16	9.2	84583	17	AC004434	*** SEQUENCING IN PROG 2.52e+02
	434	16	9.2	85214	24	ATAC005396	Arabidopsis thaliana c 2.52e+02
	435	16	9.2	85338	18	AC004279	Drosophila melanogaste 2.52e+02
c	436	16	9.2	85590	26	HS453D15	Human DNA sequence fro 2.52e+02
	437	16	9.2	87585	18	AC004299	Drosophila melanogaste 2.52e+02
c	438	16	9.2	88882	26	HS448E20	Human DNA sequence fro 2.52e+02
	439	16	9.2	91017	26	HS162C6	Human DNA sequence fro 2.52e+02
c	440	16	9.2	91414	15	ECOUW85	E. coli genomic sequen 2.52e+02
c	441	16	9.2	91854	23	ATAC003680	Arabidopsis thaliana c 2.52e+02
c	442	16	9.2	93158	23	ATAC002334	Arabidopsis thaliana c 2.52e+02
	443	16	9.2	97789	23	AC004255	Arabidopsis thaliana B 2.52e+02
c	444	16	9.2	100000	26	AP000040	Homo sapiens genomic D 2.52e+02
c	445	16	9.2	101731	17	HS445N2	Human DNA sequence *** 2.52e+02
c	446	16	9.2	102004	24	ATAC005395	Arabidopsis thaliana c 2.52e+02
	447	16	9.2	102703	26	AC004910	Homo sapiens PAC clone 2.52e+02
c	448	16	9.2	102837	26	HS522O2	Human DNA sequence fro 2.52e+02
c	449	16	9.2	102846	27	AC002483	Homo sapiens PAC clone 2.52e+02
c	450	16	9.2	103391	17	HS892F13	Human DNA sequence *** 2.52e+02

c	451	16	9.2	104793	26	AF007262	Human Xq28 cosmid 3C3 2.52e+02
	452	16	9.2	104913	26	AC004242	Homo sapiens 12q13 PAC 2.52e+02
	453	16	9.2	105174	17	HS171H20	Human DNA sequence *** 2.52e+02
c	454	16	9.2	105178	17	HS470K1	Human DNA sequence *** 2.52e+02
	455	16	9.2	105580	25	HS36J3	Human DNA sequence fro 2.52e+02
	456	16	9.2	106233	26	HS43C13	Homo sapiens DNA seque 2.52e+02
c	457	16	9.2	106233	26	HS43C13	Homo sapiens DNA seque 2.52e+02
c	458	16	9.2	106374	17	AC005623	*** SEQUENCING IN PROG 2.52e+02
c	459	16	9.2	107527	23	F19P19	Sequence of BAC F19P19 2.52e+02
	460	16	9.2	108260	26	HS179P9	Human DNA sequence fro 2.52e+02
c	461	16	9.2	108642	26	AC004451	Homo sapiens PAC clone 2.52e+02
c	462	16	9.2	109201	26	AC000403	Genomic sequence from 2.52e+02
	463	16	9.2	110000	17	CEY54E2	Caenorhabditis elegans 2.52e+02
c	464	16	9.2	110000	17	CET37D8	Caenorhabditis elegans 2.52e+02
c	465	16	9.2	110000	17	HSY237C10	Human DNA sequence *** 2.52e+02
	466	16	9.2	110870	26	HS154P24	Human DNA sequence fro 2.52e+02
c	467	16	9.2	110946	23	IG002P16	Arabidopsis thaliana B 2.52e+02
c	468	16	9.2	111352	26	HS329E20	Homo sapiens DNA seque 2.52e+02
	469	16	9.2	111906	23	IG005I10	Arabidopsis thaliana B 2.52e+02
c	470	16	9.2	111945	23	ATF1C12	Arabidopsis thaliana D 2.52e+02
	471	16	9.2	112444	17	AC003022	*** SEQUENCING IN PROG 2.52e+02
	472	16	9.2	113567	23	AF058826	Arabidopsis thaliana B 2.52e+02
c	473	16	9.2	113950	24	ATAC004697	Arabidopsis thaliana c 2.52e+02
	474	16	9.2	114617	23	ATF13C5	Arabidopsis thaliana D 2.52e+02
c	475	16	9.2	115838	17	CEY24F12	Caenorhabditis elegans 2.52e+02
c	476	16	9.2	117478	17	CEY69H2	Caenorhabditis elegans 2.52e+02
c	477	16	9.2	117911	26	HS117P19	Human DNA sequence fro 2.52e+02
c	478	16	9.2	118387	17	AC005628	*** SEQUENCING IN PROG 2.52e+02
	479	16	9.2	119136	26	AC004222	Homo sapiens chromosom 2.52e+02
c	480	16	9.2	119566	25	HS274L7	Human DNA sequence fro 2.52e+02
c	481	16	9.2	119707	23	PINCPTRPG	Pinus thunbergii chlor 2.52e+02
	482	16	9.2	120000	17	HSAC000005	*** SEQUENCING IN PROG 2.52e+02
c	483	16	9.2	122532	17	HS232D4	Human DNA sequence *** 2.52e+02
c	484	16	9.2	123360	24	ATAC003040	Arabidopsis thaliana c 2.52e+02
c	485	16	9.2	123468	17	HS232N11	Human DNA sequence *** 2.52e+02
c	486	16	9.2	125927	27	AC005754	Homo sapiens chromosom 2.52e+02
	487	16	9.2	126147	26	HS432M19	Human DNA sequence fro 2.52e+02
	488	16	9.2	126198	26	AC003089	Human BAC clone RG180F 2.52e+02
	489	16	9.2	127514	26	AC004417	Homo sapiens BAC clone 2.52e+02
	490	16	9.2	128220	26	AC004457	Homo sapiens PAC clone 2.52e+02
	491	16	9.2	129281	17	AC004999	*** SEQUENCING IN PROG 2.52e+02
	492	16	9.2	129354	26	AC004080	Homo sapiens PAC clone 2.52e+02
c	493	16	9.2	130076	17	HS510D11	Human DNA sequence *** 2.52e+02
c	494	16	9.2	131047	17	AC002421	*** SEQUENCING IN PROG 2.52e+02
	495	16	9.2	131223	17	AC004853	*** SEQUENCING IN PROG 2.52e+02
	496	16	9.2	132012	17	AC005713	*** SEQUENCING IN PROG 2.52e+02
	497	16	9.2	132029	26	HS795G23	Human DNA sequence *** 2.52e+02
	498	16	9.2	134416	22	AF020713	Bacteriophage SPBc2 co 2.52e+02
c	499	16	9.2	134474	26	HS33F8	Human DNA sequence fro 2.52e+02
c	500	16	9.2	135448	17	AC004900	*** SEQUENCING IN PROG 2.52e+02
	501	16	9.2	136170	26	AC004595	Homo sapiens BAC clone 2.52e+02

c	502	16	9.2	136254	15	ECOUW82	E. coli; the region fr 2.52e+02
	503	16	9.2	136561	17	HS430A16	Human DNA sequence *** 2.52e+02
	504	16	9.2	136859	17	HS191N21	Human DNA sequence *** 2.52e+02
	505	16	9.2	137769	26	AC002090	Genomic sequence from 2.52e+02
c	506	16	9.2	138706	17	AC004946	*** SEQUENCING IN PROG 2.52e+02
c	507	16	9.2	139444	26	HS380C13	Human DNA sequence fro 2.52e+02
	508	16	9.2	140000	17	AC004047	*** SEQUENCING IN PROG 2.52e+02
	509	16	9.2	143172	23	CLEGCGA	E.gracilis chloroplast 2.52e+02
c	510	16	9.2	143172	23	CLEGCGA	E.gracilis chloroplast 2.52e+02
	511	16	9.2	145721	17	HS112L6	Human DNA sequence *** 2.52e+02
	512	16	9.2	145844	26	AC004931	Homo sapiens PAC clone 2.52e+02
	513	16	9.2	147476	17	AC004157	*** SEQUENCING IN PROG 2.52e+02
c	514	16	9.2	147705	17	AC005681	*** SEQUENCING IN PROG 2.52e+02
	515	16	9.2	148851	26	HS155D22	Human DNA sequence fro 2.52e+02
c	516	16	9.2	149169	17	HS107N3	Human DNA sequence *** 2.52e+02
	517	16	9.2	149571	17	CEY43C5	Caenorhabditis elegans 2.52e+02
c	518	16	9.2	149752	26	AC004616	Homo sapiens Xp22 BAC 2.52e+02
c	519	16	9.2	150228	26	AC003071	Human BAC clone BK085E 2.52e+02
	520	16	9.2	151166	17	AC005520	*** SEQUENCING IN PROG 2.52e+02
c	521	16	9.2	151429	26	AC005279	Homo sapiens chromosom 2.52e+02
c	522	16	9.2	151519	26	HS231L4	Human DNA sequence *** 2.52e+02
	523	16	9.2	154028	26	HSAC002087	Human BAC clone GS113D 2.52e+02
	524	16	9.2	154710	17	AC004859	*** SEQUENCING IN PROG 2.52e+02
c	525	16	9.2	157099	26	AC002451	Human BAC clone GS306C 2.52e+02
c	526	16	9.2	158293	26	AC004746	Homo sapiens BAC clone 2.52e+02
c	527	16	9.2	159479	17	AC005535	*** SEQUENCING IN PROG 2.52e+02
	528	16	9.2	160091	26	AC002422	Human Chromosome X, co 2.52e+02
	529	16	9.2	160531	17	HS127P18	Human DNA sequence *** 2.52e+02
	530	16	9.2	160850	17	AC004957	*** SEQUENCING IN PROG 2.52e+02
	531	16	9.2	160986	17	AC005505	*** SEQUENCING IN PROG 2.52e+02
c	532	16	9.2	160986	17	AC005505	*** SEQUENCING IN PROG 2.52e+02
c	533	16	9.2	161580	26	HUAC002307	Homo sapiens Chromosom 2.52e+02
	534	16	9.2	162701	26	AC004129	Homo sapiens BAC clone 2.52e+02
	535	16	9.2	162912	26	HSAC002067	Human BAC clone RG126M 2.52e+02
	536	16	9.2	163974	27	AB011399	Homo sapiens gene for 2.52e+02
	537	16	9.2	164135	17	AC004990	*** SEQUENCING IN PROG 2.52e+02
c	538	16	9.2	165590	17	HS782D21	Human DNA sequence *** 2.52e+02
c	539	16	9.2	167880	27	HUU91323	Human Chromosome 16 BA 2.52e+02
	540	16	9.2	168644	26	AC004544	Homo sapiens BAC clone 2.52e+02
c	541	16	9.2	172533	26	AC002381	Human BAC clone RG020D 2.52e+02
c	542	16	9.2	174424	27	HSU52112	Homo sapiens Xq28 geno 2.52e+02
c	543	16	9.2	175080	17	HS280I0	Human DNA sequence *** 2.52e+02
c	544	16	9.2	176154	17	AC005317	*** SEQUENCING IN PROG 2.52e+02
	545	16	9.2	176413	26	AC005189	Homo sapiens PAC clone 2.52e+02
	546	16	9.2	177726	26	AC004083	Homo sapiens BAC clone 2.52e+02
	547	16	9.2	178098	26	HS168L15	Homo sapiens DNA seque 2.52e+02
c	548	16	9.2	178615	17	HS742C19	Human DNA sequence *** 2.52e+02
	549	16	9.2	178844	17	HS1054A22	Human DNA sequence *** 2.52e+02
c	550	16	9.2	184490	26	HSU82828	Homo sapiens ataxia te 2.52e+02
	551	16	9.2	185893	26	AC005191	Homo sapiens PAC clone 2.52e+02
	552	16	9.2	186510	26	HS451B15	Human DNA sequence fro 2.52e+02

	553	16	9.2	186550	17	CEY44F5	Caenorhabditis elegans 2.52e+02
c	554	16	9.2	188682	17	AC005050	*** SEQUENCING IN PROG 2.52e+02
	555	16	9.2	191028	23	PPU38804	Porphyra purpurea chlo 2.52e+02
	556	16	9.2	193869	17	AC005043	*** SEQUENCING IN PROG 2.52e+02
c	557	16	9.2	197391	17	AC005140	*** SEQUENCING IN PROG 2.52e+02
	558	16	9.2	201964	28	MMMHC29H7	Mus musculus major his 2.52e+02
c	559	16	9.2	202540	26	AC002524	Homo sapiens Xp22 BAC 2.52e+02
c	560	16	9.2	207730	15	BSUB0011	Bacillus subtilis comp 2.52e+02
	561	16	9.2	210000	17	AC005674	*** SEQUENCING IN PROG 2.52e+02
c	562	16	9.2	210954	26	AC003684	Homo sapiens Xp22 BAC 2.52e+02
	563	16	9.2	211468	17	CEY1E3	Caenorhabditis elegans 2.52e+02
c	564	16	9.2	213420	15	BSUB0014	Bacillus subtilis comp 2.52e+02
c	565	16	9.2	215286	26	HSU91322	Human chromosome 16p13 2.52e+02
c	566	16	9.2	217873	26	HUAC004382	Homo sapiens Chromosom 2.52e+02
c	567	16	9.2	226063	27	AC005406	, complete sequence. 2.52e+02
c	568	16	9.2	227730	17	CEY56A3	Caenorhabditis elegans 2.52e+02
c	569	16	9.2	227892	17	AC005587	*** SEQUENCING IN PROG 2.52e+02
c	570	16	9.2	232650	26	U66061	Human germline T-cell 2.52e+02
c	571	16	9.2	242893	17	CEY53C12	Caenorhabditis elegans 2.52e+02
	572	16	9.2	244582	17	CEY40H7	Caenorhabditis elegans 2.52e+02
c	573	16	9.2	253516	17	CEY6B3	Caenorhabditis elegans 2.52e+02
	574	16	9.2	255479	17	HS503N11	Human DNA sequence *** 2.52e+02
c	575	16	9.2	261002	17	CEY41C4	Caenorhabditis elegans 2.52e+02
	576	16	9.2	281376	17	CEY39C12	Caenorhabditis elegans 2.52e+02
	577	16	9.2	282173	17	CEY40B1	Caenorhabditis elegans 2.52e+02
	578	16	9.2	323932	17	AC005682	*** SEQUENCING IN PROG 2.52e+02
c	579	16	9.2	323932	17	AC005682	*** SEQUENCING IN PROG 2.52e+02
	580	16	9.2	326569	17	AC004912	*** SEQUENCING IN PROG 2.52e+02
c	581	16	9.2	335867	17	CEY87G2	Caenorhabditis elegans 2.52e+02
c	582	16	9.2	338234	18	DMU31961	Drosophila melanogaste 2.52e+02
c	583	15	8.6	191	18	LTGRR3	L.tarentolae mRNA enco 8.98e+02
	584	15	8.6	230	30	G06189	human STS WI-6241. 8.98e+02
	585	15	8.6	276	33	HIVH224	HIV-1 DNA V3 region (h 8.98e+02
	586	15	8.6	279	33	AF071290	HIV-1 isolate I6086-r 8.98e+02
	587	15	8.6	284	23	DDIACT3	slime mold (D.discoide 8.98e+02
c	588	15	8.6	306	25	HSU79752	Human pro-alpha2(I) co 8.98e+02
c	589	15	8.6	337	30	G38686	SHGC-64097 Human Homo 8.98e+02
	590	15	8.6	343	19	MTBAN1GEN	B.acutorostrata mitoch 8.98e+02
c	591	15	8.6	347	30	G25747	human STS EST370860. 8.98e+02
c	592	15	8.6	373	30	HS191YC5	H.sapiens DNA segment 8.98e+02
	593	15	8.6	400	15	LLLCPJ	L.lactis DNA for lcp J 8.98e+02
	594	15	8.6	400	30	G16785	human STS SHGC-9407 cl 8.98e+02
	595	15	8.6	415	30	G09594	human STS CHLC.GATA30D 8.98e+02
	596	15	8.6	426	30	G22766	human STS WI-15959. 8.98e+02
	597	15	8.6	429	20	MTAPRC16S	A.porcatus mitochondri 8.98e+02
c	598	15	8.6	442	18	OSU20715	Ophraella sexvittata i 8.98e+02
	599	15	8.6	455	24	AF023760	Pohlia cruda tRNA-Leu 8.98e+02
c	600	15	8.6	469	25	HSPA16E9	H.sapiens flow-sorted 8.98e+02
c	601	15	8.6	472	26	HSASA21	H.Sapiens alpha satell 8.98e+02
	602	15	8.6	499	18	HPU16969	Hypera postica Eastern 8.98e+02
	603	15	8.6	504	18	AF008298	Glossus humanus cytoch 8.98e+02

c	604	15	8.6	518	30	HUM4S1020	Human chromosome 4 (cl 8.98e+02
c	605	15	8.6	577	18	U97508	Muscidifurax zaraptor 8.98e+02
	606	15	8.6	600	18	MTAMTRNAL	Apis mellifera mitocho 8.98e+02
	607	15	8.6	607	25	HSGLYCB	Human mRNA for erythro 8.98e+02
c	608	15	8.6	732	15	CTMIPB	C.trachomatis (B) gene 8.98e+02
	609	15	8.6	756	30	HUMUT6136	Human STS UT6136. 8.98e+02
	610	15	8.6	770	25	HUMGLYE6	Human glycophorin E ge 8.98e+02
c	611	15	8.6	774	25	HUMDHLP14	Human dihydrolipoamide 8.98e+02
	612	15	8.6	847	30	G40463	E6712 Zebrafish AB Dan 8.98e+02
	613	15	8.6	853	24	AF003619	Spermacoce confusa rsp 8.98e+02
c	614	15	8.6	854	23	MITGTRN6	Torulopsis glabrata mi 8.98e+02
c	615	15	8.6	881	26	HSU79244	Human clone 23786 mRNA 8.98e+02
	616	15	8.6	958	18	DROFB4R	D. melanogaster transp 8.98e+02
	617	15	8.6	961	23	CHEMRBCL	E.maackii chloroplast 8.98e+02
c	618	15	8.6	1011	18	U01012	Herpetomonas samuelpes 8.98e+02
	619	15	8.6	1020	16	AB015998	Serratia marcescens ge 8.98e+02
	620	15	8.6	1021	15	SKU57659	Spiroplasma kunkelii s 8.98e+02
c	621	15	8.6	1095	23	DDICSA	Dictyostelium discoide 8.98e+02
c	622	15	8.6	1121	23	ZMAFS01012	Zea mays GLOSSY15 (Gl1 8.98e+02
c	623	15	8.6	1143	16	CJUS8045	Campylobacter coli his 8.98e+02
c	624	15	8.6	1396	23	YSCRPHOM	Saccharomyces cerevisi 8.98e+02
	625	15	8.6	1402	23	BGRBCLGEN	B.glaberrima rbcL gene 8.98e+02
	626	15	8.6	1402	32	OGCPRBCL	O.goreensis rbcL gene. 8.98e+02
c	627	15	8.6	1509	25	HUMPA2I	Human alpha-2 type I c 8.98e+02
	628	15	8.6	1538	23	ATPATY7	A. thaliana genes for 8.98e+02
	629	15	8.6	1578	23	ATPATY1	A. thaliana genes for 8.98e+02
c	630	15	8.6	1605	24	SCYDL136W	S.cerevisiae chromosom 8.98e+02
c	631	15	8.6	1636	18	ABO02133	Mictyris brevidactylus 8.98e+02
	632	15	8.6	1637	21	E124454	cDNA encoding human ce 8.98e+02
c	633	15	8.6	1745	15	LYSPALPA	Lysobacter enzymogenes 8.98e+02
	634	15	8.6	1750	23	YSCCHL4GN	Saccharomyces cerevisi 8.98e+02
c	635	15	8.6	1828	21	E13589	cDNA encoding KF-1. 8.98e+02
c	636	15	8.6	1834	15	ECOF165A	Escherichia coli A-maj 8.98e+02
	637	15	8.6	1848	15	SESTPSMP	S.epidermidis (968) ge 8.98e+02
	638	15	8.6	1907	24	AF017252	Lycopersicon esculentu 8.98e+02
	639	15	8.6	1941	18	DMTHB1	Drosophila melanogaste 8.98e+02
c	640	15	8.6	1956	23	SCYDL137W	S.cerevisiae chromosom 8.98e+02
c	641	15	8.6	1988	15	ATTMR	Agrobacterium tumefaci 8.98e+02
c	642	15	8.6	2012	26	HUMXQB	Human mRNA, Xq termina 8.98e+02
c	643	15	8.6	2402	15	ECRHSFEX6	Escherichia coli exten 8.98e+02
c	644	15	8.6	2438	25	HUM2C9X01	Homo sapiens cytochrom 8.98e+02
	645	15	8.6	2576	27	HSU62733	Human carnitine palmit 8.98e+02
	646	15	8.6	2592	21	E12511	Human cDNA encoding a 8.98e+02
	647	15	8.6	2592	26	HSCPT1BI	H.sapiens mRNA for car 8.98e+02
	648	15	8.6	2659	19	CCITP5P	C.canis mRNA for inosi 8.98e+02
	649	15	8.6	2750	23	YSCPROT	Saccharomyces cerevisi 8.98e+02
c	650	15	8.6	2782	24	ATH010713	Arabidopsis thaliana f 8.98e+02
	651	15	8.6	2813	28	RATCPILP	Rat mRNA for carnitine 8.98e+02
c	652	15	8.6	3049	33	TSPCAPSID	Tospovirus nonstructur 8.98e+02
	653	15	8.6	3078	18	BMOSGH1	Silk worm mRNA for sil 8.98e+02
	654	15	8.6	3157	25	HUMHBBRT	Human beta globin regi 8.98e+02

	655	15	8.6	3163	25	CHPRSA	Chimpanzee Alu type DN 8.98e+02
	656	15	8.6	3175	25	ORAHBPSBD	Orangutan beta- and de 8.98e+02
c	657	15	8.6	3228	23	ALPLEGHEMA	Alfalfa leghemoglobin 8.98e+02
	658	15	8.6	3231	33	ASVBA71V	African swine fever (A 8.98e+02
	659	15	8.6	3255	15	RPFUSNUS	R.prowazekii fusA, trp 8.98e+02
c	660	15	8.6	3259	25	HUMPKCAMD	Human cAMP-dependent p 8.98e+02
c	661	15	8.6	3261	21	E13588	cDNA encoding KF-1. 8.98e+02
c	662	15	8.6	3263	28	D76445	Mouse mxf-1 mRNA, comp 8.98e+02
	663	15	8.6	3280	21	I28324	Sequence 4 from patent 8.98e+02
c	664	15	8.6	3353	18	MICPCOX3	Crithidia fasciculata 8.98e+02
c	665	15	8.6	3416	18	CSA7437	Cupiennius salei mRNA 8.98e+02
c	666	15	8.6	3539	33	AHV1DNA	Alcelaphine Herpesviru 8.98e+02
c	667	15	8.6	3555	24	FPZ99763	Flaveria pringlei gdcs 8.98e+02
	668	15	8.6	3694	26	HUMORF08	Human mRNA for KIAA002 8.98e+02
c	669	15	8.6	3700	18	DDGP80G	D.discoideum gp80 gene 8.98e+02
	670	15	8.6	3762	23	CCRHATPBE	C.caldarium rhodoplast 8.98e+02
	671	15	8.6	4089	18	DMTNFB	Transposable element F 8.98e+02
	672	15	8.6	4100	15	HEAHPAX	Haemophilus parainflue 8.98e+02
c	673	15	8.6	4300	28	RNU15138	Rattus norvegicus LIC- 8.98e+02
	674	15	8.6	4441	19	SSU09179	Sus scrofa dihydropyri 8.98e+02
	675	15	8.6	4487	23	AF021257	Hordeum vulgare J2 kDa 8.98e+02
c	676	15	8.6	4556	23	OSPCNAGEN	Rice (O. sativa) gene 8.98e+02
c	677	15	8.6	4584	23	GMENOD2B	Soybean ENOD2B gene fo 8.98e+02
	678	15	8.6	4601	18	DMU11584	Drosophila melanogaste 8.98e+02
	679	15	8.6	4740	28	MUSIL1TCR	Mouse interleukin-1 re 8.98e+02
c	680	15	8.6	4819	15	LLPEPPF12	L.lactis pepF1 & pepF2 8.98e+02
	681	15	8.6	4918	21	A37476	Sequence 1 from Patent 8.98e+02
	682	15	8.6	5127	18	PFSC03098	Plasmodium falciparum 8.98e+02
	683	15	8.6	5181	27	HSU59831	Human transcription fa 8.98e+02
	684	15	8.6	5212	26	ABO02387	Human mRNA for KIAA038 8.98e+02
	685	15	8.6	5212	28	MUSLBPA	Mouse adipocyte lipid 8.98e+02
c	686	15	8.6	5334	23	DCY15177	Daucus carota DNA frag 8.98e+02
	687	15	8.6	5647	15	BACKNB	Bacillus stearothermop 8.98e+02
c	688	15	8.6	5709	23	GMLOX3G	Soybean lox-3 gene for 8.98e+02
c	689	15	8.6	5798	28	RNU92079	Rattus norvegicus GLGF 8.98e+02
c	690	15	8.6	6264	23	MISC16	Yeast (S. cerevisiae) 8.98e+02
c	691	15	8.6	6478	25	HSU16812	Human Bak-2 gene, comp 8.98e+02
c	692	15	8.6	6812	23	SCSUP44GN	S.cerevisiae G2882, G2 8.98e+02
	693	15	8.6	6883	18	PFSC04010	Plasmodium falciparum 8.98e+02
	694	15	8.6	7732	33	JDVCG	Jembrana disease virus 8.98e+02
c	695	15	8.6	7964	18	PFSC04015	Plasmodium falciparum 8.98e+02
	696	15	8.6	8126	28	S66283	Spnb-1=beta-spectrin [8.98e+02
	697	15	8.6	8303	15	BBU03641	Borrelia burgdorferi I 8.98e+02
	698	15	8.6	8540	25	HUMLIMGP	Human lens membrane pr 8.98e+02
c	699	15	8.6	8797	33	AB006531	Plautia stali intestin 8.98e+02
c	700	15	8.6	9304	18	PFSC03105	Plasmodium falciparum 8.98e+02
	701	15	8.6	9400	15	D85752	Enterococcus faecalis 8.98e+02
c	702	15	8.6	9547	33	BYU47033	Bean yellow mosaic pot 8.98e+02
c	703	15	8.6	9603	28	CGU09104	Cricetulus griseus SIN 8.98e+02
	704	15	8.6	10240	28	AF029875	Rattus norvegicus musc 8.98e+02
	705	15	8.6	10277	15	AE001176	Borrelia burgdorferi (8.98e+02

	706	15	8.6	11198	26	AB003286	Homo sapiens DNA for c 8.98e+02
	707	15	8.6	11534	18	PFAARP1PR	P.falciparum mRNA for 8.98e+02
c	708	15	8.6	11700	23	RICSBE1	Rice gene for starch b 8.98e+02
c	709	15	8.6	12116	15	UJ9726	Mycoplasma genitalium 8.98e+02
	710	15	8.6	12222	19	BOVLYSOZMA	Bos taurus lysozyme ge 8.98e+02
	711	15	8.6	12971	15	U32819	Haemophilus influenzae 8.98e+02
	712	15	8.6	13036	33	ASVLIS57	African swine fever vi 8.98e+02
c	713	15	8.6	13701	15	U67559	Methanococcus jannasch 8.98e+02
c	714	15	8.6	13829	15	AE000722	Aquifex aeolicus aecti 8.98e+02
c	715	15	8.6	14524	18	CET04F3	Caenorhabditis elegans 8.98e+02
	716	15	8.6	14902	15	AE001164	Borrelia burgdorferi (8.98e+02
	717	15	8.6	15122	18	CEH40L08	Caenorhabditis elegans 8.98e+02
c	718	15	8.6	16116	18	CEH12C20	Caenorhabditis elegans 8.98e+02
c	719	15	8.6	16267	23	SCE325	S.cerevisiae chromosom 8.98e+02
	720	15	8.6	16407	18	CET51B9A	Caenorhabditis elegans 8.98e+02
	721	15	8.6	16959	15	ECAE000465	Escherichia coli K-12 8.98e+02
	722	15	8.6	16993	18	DM1433	D.melanogaster Jun and 8.98e+02
c	723	15	8.6	17237	18	CET11B7	Caenorhabditis elegans 8.98e+02
c	724	15	8.6	17340	16	SPAJ6986	Streptococcus pneumoni 8.98e+02
c	725	15	8.6	20759	18	PPSC03057	Plasmodium falciparum 8.98e+02
	726	15	8.6	20874	26	HSCFAT5	Human DNA sequence fro 8.98e+02
	727	15	8.6	21552	18	CEM04D8	Caenorhabditis elegans 8.98e+02
	728	15	8.6	22085	18	CELC18B10	Caenorhabditis elegans 8.98e+02
	729	15	8.6	22253	23	SCG3220X	S.cerevisiae chromosom 8.98e+02
	730	15	8.6	22516	18	CEC50C10	Caenorhabditis elegans 8.98e+02
c	731	15	8.6	22617	26	HSA005577	Homo sapiens pfkfb2 ge 8.98e+02
c	732	15	8.6	22795	18	CELF40H7	Caenorhabditis elegans 8.98e+02
c	733	15	8.6	22950	18	CEF55H2	Caenorhabditis elegans 8.98e+02
	734	15	8.6	24586	18	CEL2C262	C. elegans cosmid ZC26 8.98e+02
	735	15	8.6	25395	18	DMC118B3	Drosophila melanogaste 8.98e+02
c	736	15	8.6	25700	18	CELF27E11	Caenorhabditis elegans 8.98e+02
	737	15	8.6	26675	18	CELF22F4	Caenorhabditis elegans 8.98e+02
c	738	15	8.6	26756	23	SCCHRIVFY	S.cerevisiae DNA of co 8.98e+02
c	739	15	8.6	26923	18	CEMO3B6	Caenorhabditis elegans 8.98e+02
	740	15	8.6	27743	18	CELC30A5	C. elegans cosmid C30A 8.98e+02
	741	15	8.6	28065	26	AC005354	Homo sapiens chromosom 8.98e+02
	742	15	8.6	29095	18	CELF18C5	Caenorhabditis elegans 8.98e+02
c	743	15	8.6	29214	18	CELC50F2	Caenorhabditis elegans 8.98e+02
	744	15	8.6	29355	18	CEH13N06	Caenorhabditis elegans 8.98e+02
c	745	15	8.6	29820	18	CEF48F7	Caenorhabditis elegans 8.98e+02
	746	15	8.6	30320	18	CEC05B5	Caenorhabditis elegans 8.98e+02
c	747	15	8.6	30489	18	AC004421	Drosophila melanogaste 8.98e+02
	748	15	8.6	30567	18	CET28F4	Caenorhabditis elegans 8.98e+02
	749	15	8.6	30652	18	CELF59D6	Caenorhabditis elegans 8.98e+02
c	750	15	8.6	32258	18	CEF42D1	Caenorhabditis elegans 8.98e+02
	751	15	8.6	32403	18	CELK03B4	Caenorhabditis elegans 8.98e+02
	752	15	8.6	32480	18	AC004322	Drosophila melanogaste 8.98e+02
c	753	15	8.6	32480	18	CELC10E2	Caenorhabditis elegans 8.98e+02
c	754	15	8.6	32566	18	CEY94A7B	Caenorhabditis elegans 8.98e+02
	755	15	8.6	33086	18	CELT25B6	Caenorhabditis elegans 8.98e+02
c	756	15	8.6	33734	18	CEC05C9	Caenorhabditis elegans 8.98e+02

	757	15	8.6	33834	18	CELF53H1	Caenorhabditis elegans	8.98e+02
c	758	15	8.6	33837	18	CEF27C8	Caenorhabditis elegans	8.98e+02
	759	15	8.6	33860	18	CELT23E1	Caenorhabditis elegans	8.98e+02
	760	15	8.6	34581	18	CET01B4	Caenorhabditis elegans	8.98e+02
	761	15	8.6	34849	24	SPBC14F5	S.pombe chromosome II	8.98e+02
c	762	15	8.6	34944	18	TGU87145	Toxoplasma gondii chlo	8.98e+02
c	763	15	8.6	35561	23	SPBC2F6	S.pombe chromosome II	8.98e+02
c	764	15	8.6	36051	18	CET22A3	Caenorhabditis elegans	8.98e+02
	765	15	8.6	36143	18	CELT06C10	Caenorhabditis elegans	8.98e+02
c	766	15	8.6	36159	26	HSY09781	H.sapiens beta-sarcogl	8.98e+02
c	767	15	8.6	36262	18	CELR07E4	Caenorhabditis elegans	8.98e+02
c	768	15	8.6	36798	26	AC004164	Homo sapiens chromosom	8.98e+02
	769	15	8.6	36922	18	PFSC030Z6	Plasmodium falciparum	8.98e+02
	770	15	8.6	37003	18	CELF33D11	Caenorhabditis elegans	8.98e+02
c	771	15	8.6	37017	18	CELF54D8	Caenorhabditis elegans	8.98e+02
c	772	15	8.6	37310	18	CEZK1098	Caenorhabditis elegans	8.98e+02
c	773	15	8.6	37329	18	CELT19H12	Caenorhabditis elegans	8.98e+02
c	774	15	8.6	37797	18	CET21B4	Caenorhabditis elegans	8.98e+02
	775	15	8.6	37881	18	CELK09H9	Caenorhabditis elegans	8.98e+02
	776	15	8.6	38060	18	CET13F2	Caenorhabditis elegans	8.98e+02
	777	15	8.6	38269	23	SPAC23C4	S.pombe chromosome I c	8.98e+02
c	778	15	8.6	38343	18	CEF21D9	Caenorhabditis elegans	8.98e+02
	779	15	8.6	38464	18	CEK08E7	Caenorhabditis elegans	8.98e+02
	780	15	8.6	38468	25	HUMM73P4	Human cosmid M73P4, co	8.98e+02
c	781	15	8.6	38487	18	CELC01G8	Caenorhabditis elegans	8.98e+02
c	782	15	8.6	38636	27	AC005580	Homo sapiens chromosom	8.98e+02
	783	15	8.6	38644	26	AC002508	Human Cosmid g5129z171	8.98e+02
c	784	15	8.6	38671	18	CEC01G10	Caenorhabditis elegans	8.98e+02
c	785	15	8.6	38773	18	CET06E4	Caenorhabditis elegans	8.98e+02
c	786	15	8.6	39229	26	AC004755	Homo sapiens chromosom	8.98e+02
c	787	15	8.6	39395	26	AC004212	Homo sapiens clone UWG	8.98e+02
c	788	15	8.6	39582	18	DMC17A9	Drosophila melanogaste	8.98e+02
c	789	15	8.6	39779	26	AC002450	Human BAC clone GS244B	8.98e+02
	790	15	8.6	39977	18	CET08D2	Caenorhabditis elegans	8.98e+02
c	791	15	8.6	40233	26	D87010	Human (lambda) DNA for	8.98e+02
c	792	15	8.6	40293	18	CELC45G9	Caenorhabditis elegans	8.98e+02
	793	15	8.6	40437	18	CELT10B5	Caenorhabditis elegans	8.98e+02
	794	15	8.6	40879	18	CEF32H2	Caenorhabditis elegans	8.98e+02
	795	15	8.6	40951	18	CELB0205	Caenorhabditis elegans	8.98e+02
	796	15	8.6	41071	18	CELB0303	C.elegans cosmid B0303	8.98e+02
	797	15	8.6	41299	18	CELF55A4	Caenorhabditis elegans	8.98e+02
	798	15	8.6	41502	18	CEZK856	Caenorhabditis elegans	8.98e+02
c	799	15	8.6	41754	18	CELB0238	Caenorhabditis elegans	8.98e+02
	800	15	8.6	41947	26	HSAC000358	Human cosmid g1572c237	8.98e+02
c	801	15	8.6	42139	26	AC004654	Homo sapiens chromosom	8.98e+02
	802	15	8.6	42313	18	CELC05D10	Caenorhabditis elegans	8.98e+02
	803	15	8.6	42345	17	CEF52D11	Caenorhabditis elegans	8.98e+02
c	804	15	8.6	43468	23	SC8358	S.cerevisiae chromosom	8.98e+02
	805	15	8.6	43802	18	CEK07F5	Caenorhabditis elegans	8.98e+02
	806	15	8.6	44036	18	CELT28A11	Caenorhabditis elegans	8.98e+02
	807	15	8.6	44195	18	CELC15B12	Caenorhabditis elegans	8.98e+02

c	808	15	8.6	44237	18	CER05H10	Caenorhabditis elegans	8.98e+02
c	809	15	8.6	44548	26	AF020801	Homo sapiens cosmid c1	8.98e+02
	810	15	8.6	44726	18	CELF09E5	Caenorhabditis elegans	8.98e+02
	811	15	8.6	45857	18	DMC95B7	Drosophila melanogaste	8.98e+02
c	812	15	8.6	46270	17	CEY43F4	Caenorhabditis elegans	8.98e+02
c	813	15	8.6	55328	23	PWD02970	Prototheca wickerhamii	8.98e+02
	814	15	8.6	61380	24	AB017066	Arabidopsis thaliana g	8.98e+02
	815	15	8.6	68548	25	HSU70984	Human Xp22 cosmids U10	8.98e+02
c	816	15	8.6	70319	26	AC002453	Human BAC clone GS440B	8.98e+02
c	817	15	8.6	70475	17	AB018112	*** SEQUENCING IN PROG	8.98e+02
	818	15	8.6	71664	26	AC003692	Human Chromosome 11 pa	8.98e+02
	819	15	8.6	72722	18	AC004373	Drosophila melanogaste	8.98e+02
	820	15	8.6	72847	18	AC004266	Drosophila melanogaste	8.98e+02
	821	15	8.6	74877	23	AB006705	Arabidopsis thaliana g	8.98e+02
	822	15	8.6	76947	18	AC004326	Drosophila melanogaste	8.98e+02
c	823	15	8.6	78259	24	F20D22	Arabidopsis thaliana c	8.98e+02
	824	15	8.6	78360	17	AC005130	*** SEQUENCING IN PROG	8.98e+02
c	825	15	8.6	78844	23	AB006708	Arabidopsis thaliana g	8.98e+02
	826	15	8.6	80315	23	AB005232	Arabidopsis thaliana g	8.98e+02
c	827	15	8.6	80697	17	AC005455	*** SEQUENCING IN PROG	8.98e+02
	828	15	8.6	80776	17	AC005421	*** SEQUENCING IN PROG	8.98e+02
c	829	15	8.6	82503	24	AB016886	Arabidopsis thaliana g	8.98e+02
	830	15	8.6	83024	26	AC000389	Genomic sequence from	8.98e+02
	831	15	8.6	84621	26	HS244F1	Human DNA sequence fro	8.98e+02
c	832	15	8.6	85791	24	AB016888	Arabidopsis thaliana g	8.98e+02
c	833	15	8.6	86558	17	HS125M5	Human DNA sequence ***	8.98e+02
c	834	15	8.6	87325	18	AC004375	Drosophila melanogaste	8.98e+02
c	835	15	8.6	87552	17	HS217O16	Human DNA sequence ***	8.98e+02
c	836	15	8.6	87585	18	AC004299	Drosophila melanogaste	8.98e+02
c	837	15	8.6	87634	17	CEK01F6	*** SEQUENCING IN PROG	8.98e+02
c	838	15	8.6	88274	17	AC005693	*** SEQUENCING IN PROG	8.98e+02
c	839	15	8.6	88343	23	ATT19P19	Arabidopsis thaliana D	8.98e+02
c	840	15	8.6	90430	26	AC003100	Homo sapiens chromosom	8.98e+02
	841	15	8.6	91007	17	AC004374	*** SEQUENCING IN PROG	8.98e+02
	842	15	8.6	93008	17	AC005120	*** SEQUENCING IN PROG	8.98e+02
c	843	15	8.6	94718	18	AC004658	Drosophila melanogaste	8.98e+02
	844	15	8.6	95167	23	ATAC003105	Arabidopsis thaliana c	8.98e+02
	845	15	8.6	95272	18	AC005298	Drosophila melanogaste	8.98e+02
c	846	15	8.6	97451	23	F19K23	Sequence of BAC F19K23	8.98e+02
	847	15	8.6	98209	26	HS260J9	Human DNA sequence fro	8.98e+02
c	848	15	8.6	98695	18	AC005556	Drosophila melanogaste	8.98e+02
	849	15	8.6	100000	26	AP000009	Homo sapiens genomic D	8.98e+02
c	850	15	8.6	102230	26	HS234H5	Homo sapiens DNA seque	8.98e+02
c	851	15	8.6	104386	26	D87009	Human (lambda) DNA for	8.98e+02
c	852	15	8.6	105276	17	AC005700	*** SEQUENCING IN PROG	8.98e+02
c	853	15	8.6	105563	26	AC003983	Human PAC clone DJ0093	8.98e+02
c	854	15	8.6	107313	17	HS75N14	Human DNA sequence ***	8.98e+02
c	855	15	8.6	107331	17	CEY49A10	Caenorhabditis elegans	8.98e+02
c	856	15	8.6	108695	26	AC002119	Homo sapiens chromosom	8.98e+02
c	857	15	8.6	109279	26	HS394F12	Human DNA sequence fro	8.98e+02
c	858	15	8.6	110000	17	CEY53F4	Caenorhabditis elegans	8.98e+02

c	859	15	8.6	110000	15	ECOUW67	Escherichia coli K-12	8.98e+02
	860	15	8.6	110096	26	CH19HHR23	Homo sapiens DNA from	8.98e+02
c	861	15	8.6	110102	23	ATF2BJ12	Arabidopsis thaliana D	8.98e+02
	862	15	8.6	110174	26	D83253	Homo sapiens genomic D	8.98e+02
c	863	15	8.6	112242	26	AC004741	Homo sapiens PAC clone	8.98e+02
	864	15	8.6	112400	17	HS30A23	Human DNA sequence ***	8.98e+02
	865	15	8.6	112756	26	AF030876	Homo sapiens chromosom	8.98e+02
c	866	15	8.6	113457	26	AC003686	Homo sapiens 12q13.1 P	8.98e+02
	867	15	8.6	113950	24	ATAC004697	Arabidopsis thaliana c	8.98e+02
c	868	15	8.6	114956	26	HSAC002072	Human PAC clone DJ218B	8.98e+02
c	869	15	8.6	115355	17	CEY55D9	Caenorhabditis elegans	8.98e+02
	870	15	8.6	115984	26	HS292F10	Human DNA sequence fro	8.98e+02
	871	15	8.6	116884	17	PFMAL1P1	Plasmodium falciparum	8.98e+02
	872	15	8.6	117852	26	AC002426	Human Chromosome 15q26	8.98e+02
c	873	15	8.6	117994	17	HS409D10	Human DNA sequence ***	8.98e+02
c	874	15	8.6	118521	26	HUAC004514	Homo sapiens Chromosom	8.98e+02
c	875	15	8.6	121290	17	CEH27C14	Caenorhabditis elegans	8.98e+02
	876	15	8.6	123386	24	F12F1	Arabidopsis thaliana c	8.98e+02
c	877	15	8.6	123452	17	AC005079	*** SEQUENCING IN PROG	8.98e+02
c	878	15	8.6	123805	26	AC002464	Human BAC clone RG331P	8.98e+02
	879	15	8.6	124000	17	AC004060	*** SEQUENCING IN PROG	8.98e+02
c	880	15	8.6	125179	24	ATT4L20	Arabidopsis thaliana D	8.98e+02
	881	15	8.6	125958	17	HS796F18	Human DNA sequence ***	8.98e+02
c	882	15	8.6	126995	26	AC004773	Homo sapiens PAC clone	8.98e+02
c	883	15	8.6	127079	26	HS214K23	Human DNA sequence fro	8.98e+02
c	884	15	8.6	127356	17	AC005063	*** SEQUENCING IN PROG	8.98e+02
	885	15	8.6	127834	26	AC003958	Homo sapiens chromosom	8.98e+02
c	886	15	8.6	128019	17	AC005586	*** SEQUENCING IN PROG	8.98e+02
c	887	15	8.6	128340	17	AC002419	*** SEQUENCING IN PROG	8.98e+02
	888	15	8.6	129098	26	HS46H23	Human DNA sequence fro	8.98e+02
c	889	15	8.6	129515	17	HS886K2	Human DNA sequence ***	8.98e+02
	890	15	8.6	129722	26	HS1500S	Homo sapiens DNA seque	8.98e+02
c	891	15	8.6	129723	17	HS225E12	Human DNA sequence ***	8.98e+02
c	892	15	8.6	130594	26	HS63M23	Human DNA sequence fro	8.98e+02
	893	15	8.6	130632	26	AC004547	Homo sapiens PAC clone	8.98e+02
	894	15	8.6	132666	18	AC005270	Drosophila melanogaste	8.98e+02
c	895	15	8.6	132783	26	AC000378	Human Chromosome 11 pa	8.98e+02
	896	15	8.6	134970	26	AC002504	Human Xp22 PAC RPCI1-2	8.98e+02
c	897	15	8.6	135289	17	AC004961	*** SEQUENCING IN PROG	8.98e+02
c	898	15	8.6	135599	23	CPU30821	Cyanophora paradoxa cy	8.98e+02
	899	15	8.6	136037	26	AC004104	Homo sapiens 12q24 PAC	8.98e+02
c	900	15	8.6	136805	17	AC005507	*** SEQUENCING IN PROG	8.98e+02
	901	15	8.6	137496	27	AC005288	Homo sapiens chromosom	8.98e+02
	902	15	8.6	139171	26	HS354N19	Human DNA sequence ***	8.98e+02
c	903	15	8.6	139480	26	HUAC002331	Homo sapiens Chromosom	8.98e+02
	904	15	8.6	141120	26	AC002528	Human BAC clone GS542D	8.98e+02
c	905	15	8.6	141475	26	AC002980	Homo sapiens Xp22 BAC	8.98e+02
	906	15	8.6	141752	17	HS971N18	Human DNA sequence ***	8.98e+02
c	907	15	8.6	142761	17	HS566G5	Human DNA sequence ***	8.98e+02
	908	15	8.6	143092	18	CEY17G7B	Caenorhabditis elegans	8.98e+02
	909	15	8.6	143612	17	HS936P19	Human DNA sequence ***	8.98e+02

c	910	15	8.6	143819	26	AC002994	Homo sapiens chromosom	8.98e+02
	911	15	8.6	147542	17	HS754E20	Human DNA sequence ***	8.98e+02
c	912	15	8.6	148359	17	AC004811	*** SEQUENCING IN PROG	8.98e+02
c	913	15	8.6	148400	26	AC004768	Homo sapiens chromosom	8.98e+02
	914	15	8.6	149316	17	AC005008	*** SEQUENCING IN PROG	8.98e+02
	915	15	8.6	149322	17	HS1119D9	Human DNA sequence ***	8.98e+02
	916	15	8.6	150613	23	AB001684	Chlorella vulgaris C-2	8.98e+02
c	917	15	8.6	150641	18	CEY41E3	Caenorhabditis elegans	8.98e+02
	918	15	8.6	150894	15	D90904	Synechocystis sp. PCC6	8.98e+02
	919	15	8.6	151018	17	PFMAL1P6	Plasmodium falciparum	8.98e+02
	920	15	8.6	151779	17	HS431P23	Human DNA sequence ***	8.98e+02
	921	15	8.6	154821	17	AC005341	*** SEQUENCING IN PROG	8.98e+02
c	922	15	8.6	154909	17	AC005799	*** SEQUENCING IN PROG	8.98e+02
	923	15	8.6	155398	17	AC003020	*** SEQUENCING IN PROG	8.98e+02
c	924	15	8.6	155660	17	HS1177E19	Human DNA sequence ***	8.98e+02
	925	15	8.6	156050	17	HS111B22	Human DNA sequence ***	8.98e+02
c	926	15	8.6	156811	27	AC005548	Homo sapiens chromosom	8.98e+02
	927	15	8.6	156994	17	HS1018K9	Human DNA sequence ***	8.98e+02
c	928	15	8.6	157897	17	HS171N11	Human DNA sequence ***	8.98e+02
c	929	15	8.6	159245	17	AC004819	*** SEQUENCING IN PROG	8.98e+02
c	930	15	8.6	161453	27	AC004993	Homo sapiens PAC clone	8.98e+02
c	931	15	8.6	162075	26	HS127D3	Homo sapiens DNA seque	8.98e+02
c	932	15	8.6	162472	17	AC005730	*** SEQUENCING IN PROG	8.98e+02
	933	15	8.6	162575	17	AC004086	*** SEQUENCING IN PROG	8.98e+02
	934	15	8.6	163023	17	AC004815	*** SEQUENCING IN PROG	8.98e+02
c	935	15	8.6	163671	17	HS142F18	Human DNA sequence ***	8.98e+02
	936	15	8.6	163889	17	AC004958	*** SEQUENCING IN PROG	8.98e+02
c	937	15	8.6	164135	17	AC004990	*** SEQUENCING IN PROG	8.98e+02
	938	15	8.6	164237	17	AC004838	*** SEQUENCING IN PROG	8.98e+02
c	939	15	8.6	164511	27	AC004894	Homo sapiens PAC clone	8.98e+02
	940	15	8.6	164538	27	AC005774	Homo sapiens chromosom	8.98e+02
c	941	15	8.6	165608	26	AC004492	Homo sapiens BAC clone	8.98e+02
	942	15	8.6	167238	17	HS467L1	Human DNA sequence ***	8.98e+02
	943	15	8.6	167591	26	HS41A7	Human DNA sequence ***	8.98e+02
c	944	15	8.6	169045	27	AC005544	Homo sapiens chromosom	8.98e+02
	945	15	8.6	170000	17	AC004394	*** SEQUENCING IN PROG	8.98e+02
c	946	15	8.6	170658	26	AC004072	Human Chromosome X clo	8.98e+02
c	947	15	8.6	171147	17	HS1059H15	Human DNA sequence ***	8.98e+02
	948	15	8.6	172502	17	AC005039	*** SEQUENCING IN PROG	8.98e+02
c	949	15	8.6	172995	17	AC004926	*** SEQUENCING IN PROG	8.98e+02
	950	15	8.6	173321	17	HS242N11	Human DNA sequence ***	8.98e+02
	951	15	8.6	173513	26	HS125H2	Human DNA sequence fro	8.98e+02
c	952	15	8.6	174076	17	AC005743	*** SEQUENCING IN PROG	8.98e+02
	953	15	8.6	176296	17	AC004882	*** SEQUENCING IN PROG	8.98e+02
	954	15	8.6	177014	26	AC004063	Homo sapiens chromosom	8.98e+02
	955	15	8.6	177163	26	HS134E15	Human DNA sequence ***	8.98e+02
c	956	15	8.6	177638	17	AC004812	*** SEQUENCING IN PROG	8.98e+02
c	957	15	8.6	179854	17	CEY19D2	Caenorhabditis elegans	8.98e+02
	958	15	8.6	179933	17	AC005048	*** SEQUENCING IN PROG	8.98e+02
c	959	15	8.6	180114	17	HS52D1	Human DNA sequence ***	8.98e+02
	960	15	8.6	180388	25	HUMRETBLAS	Human retinoblastoma s	8.98e+02

```
  c  961     15     8.6 181221 17  AC004835     *** SEQUENCING IN PROG  8.98e+02
  c  962     15     8.6 181525 17  HS996D20     Human DNA sequence ***  8.98e+02
     963     15     8.6 181978 17  AC005058     *** SEQUENCING IN PROG  8.98e+02
  c  964     15     8.6 183180 17  AC004929     *** SEQUENCING IN PROG  8.98e+02
  c  965     15     8.6 185000 17  AC004064     *** SEQUENCING IN PROG  8.98e+02
  c  966     15     8.6 185035 26  AC004770     Homo sapiens chromosom  8.98e+02
     967     15     8.6 185095 17  AC005495     *** SEQUENCING IN PROG  8.98e+02
  c  968     15     8.6 188629 17  AC004816     *** SEQUENCING IN PROG  8.98e+02
     969     15     8.6 190665 17  HS495O10     Human DNA sequence ***  8.98e+02
     970     15     8.6 191233 17  AC005409     *** SEQUENCING IN PROG  8.98e+02
  c  971     15     8.6 191233 17  AC005409     *** SEQUENCING IN PROG  8.98e+02
  c  972     15     8.6 192037 17  AC004486     *** SEQUENCING IN PROG  8.98e+02
  c  973     15     8.6 195033 17  HS608E8      Human DNA sequence ***  8.98e+02
  c  974     15     8.6 197900 26  AC002531     Homo sapiens chromosom  8.98e+02
  c  975     15     8.6 198552 23  ATAP22       Arabidopsis thaliana D  8.98e+02
     976     15     8.6 199553 17  HS526I14     Human DNA sequence ***  8.98e+02
  c  977     15     8.6 199878 17  AC005027     *** SEQUENCING IN PROG  8.98e+02
  c  978     15     8.6 200134 17  AC004806     *** SEQUENCING IN PROG  8.98e+02
  c  979     15     8.6 200349 27  HSU85195     Homo sapiens BAC129, c  8.98e+02
     980     15     8.6 202004 26  HUAF001549   Human Chromosome 16 BA  8.98e+02
  c  981     15     8.6 206606 23  ATFCA1       Arabidopsis thaliana D  8.98e+02
  c  982     15     8.6 206784 26  HS93L7       Homo sapiens DNA seque  8.98e+02
  c  983     15     8.6 220000 17  AC004480     *** SEQUENCING IN PROG  8.98e+02
  c  984     15     8.6 224645 17  AC004153     *** SEQUENCING IN PROG  8.98e+02
  c  985     15     8.6 225676 17  HS199L16     Human DNA sequence ***  8.98e+02
  c  986     15     8.6 235134 17  AC004849     *** SEQUENCING IN PROG  8.98e+02
     987     15     8.6 235922 17  CEY7A5       Caenorhabditis elegans  8.98e+02
     988     15     8.6 242825 26  HSPEX        H.sapiens PEX gene.     8.98e+02
  c  989     15     8.6 242825 26  HSPEX        H.sapiens PEX gene.     8.98e+02
  c  990     15     8.6 250529 26  HUAE000658   Homo sapiens T-cell re  8.98e+02
     991     15     8.6 253812 17  CEY38H6      Caenorhabditis elegans  8.98e+02
  c  992     15     8.6 263313 17  CEY48C3      Caenorhabditis elegans  8.98e+02
  c  993     15     8.6 269078 17  CEY76A2      Caenorhabditis elegans  8.98e+02
     994     15     8.6 275138 17  CEY54E5      Caenorhabditis elegans  8.98e+02
  c  995     15     8.6 276261 17  HSAC000406   *** SEQUENCING IN PROG  8.98e+02
  c  996     15     8.6 286715 17  CEY45F3      Caenorhabditis elegans  8.98e+02
  c  997     15     8.6 300197 17  CEY54G9      Caenorhabditis elegans  8.98e+02
     998     15     8.6 305444 17  CEY67H2      Caenorhabditis elegans  8.98e+02
  c  999     15     8.6 316170 17  CEY37A1      Caenorhabditis elegans  8.98e+02
  c 1000     15     8.6 337546 17  CEY47D3      Caenorhabditis elegans  8.98e+02

                                ALIGNMENTS

 RESULT    1
 LOCUS       AF016833      6513 bp     mRNA             PRI       03-FEB-1998
 DEFINITION  Homo sapiens maltase-glucoamylase mRNA, complete cds.
 ACCESSION   AF016833
 NID         g2826520
 KEYWORDS    .
 SOURCE      human.
```



```
   ORGANISM  Homo sapiens
             Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
             Primates; Catarrhini; Hominidae; Homo.
 REFERENCE   1  (bases 1 to 6513)
   AUTHORS   Nichols,B.L., Eldering,J., Avery,S., Hahn,D., Quaroni,A. and
             Sterchi,E.
   TITLE     Human small intestinal maltase-glucoamylase cDNA cloning. Homology
             to sucrase-isomaltase
   JOURNAL   J. Biol. Chem. 273 (5), 3076-3081 (1998)
   MEDLINE   98112863
 REFERENCE   2  (bases 1 to 6513)
   AUTHORS   Nichols,B.L., Eldering,J.A., Avery,S.E., Hahn,D., Quaroni,A. and
             Sterchi,E.E.
   TITLE     Direct Submission
   JOURNAL   Submitted (31-JUL-1997) Pediatrics, Baylor College of Medicine,
             1100 Bates, CNRC 10066, Houston, TX 77030-2600, USA
 FEATURES             Location/Qualifiers
     source           1. .6513
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="7"
                      /tissue_type="small intestine"
                      /cell_type="enterocyte"
     CDS              55. .5628
                      /EC_number="3.2.1.20"
                      /EC_number="3.2.1.3"
                      /note="brush border hydrolase; glucosyl hydrolase;
                      1,4-O-alpha-D-glucanohydrolase; has similarity to
                      sucrase-isomaltase"
                      /codon_start=1
                      /product="maltase-glucoamylase"
                      /db_xref="PID:g2826521"
                      /translation="MARKKLKKFTTLEIVLSVLLLVLFIISIVLIVLLAKESLKSTAP
                      DPGTTGTPDPGTTGTPDPGTTGTTHARTTGPPDPGTTGTTPVSAECPVVNELERINCI
                      PDQPPTKATCDQRGCCWNPQGAVSVPWCYYSKNHSYHVEGNLVNTNAGPTARLKNLPS
                      SPVFGSNVDNVLLTAEYQTSNRFHFKLTDQTNNRFEVPHEHVQSFSGNAAASLTYQVE
                      ISRQPFSIKVTRRSNNRVLFDSSIGPLLFADQFLQLSTRLPSTNVYGLGEHVHQQYRH
                      DWDWKTWPIFNRDTTPNNGNSNLYGAQTFFLCLEDASGLSFGVFLMNSNAMEVVLQPA
                      PAITYRTIGGILDFYVFLGNTPEQVVQEYLELIGRPALPSYWALGFHLSRYEYGTLDN
                      MREVVERNRAAQLPYDVQHADIDYMDERRDFTYDSVDFKGFPEFVNELHNNGQKLVII
                      VDPAISNNSSSSKPYGPYDRGSDMKIWVNSSDGVTPLIGEVWPGQTVFPDYTNPNCAV
                      WWTKEFELFHNQVEFDGIWIDMNEVSNFVDGSVSGCSTNNLNNPPFTPRILDGYLPCK
                      TLCMDAVQHWGKQYDIHNLYGYSMAVATAEAAKTVFPNKRSFILTRSTFAGSGKFAAH
                      WLGDNTATWDDLRWSIPGVLEFNLFGIPMVGPDICGFALDTPEELCRRWMQLGAFYPF
                      SRNHNGQGYKDQDPASFGADSLLLNSSRHYLNIRYTLLPYLYTLFFRAHSRGDTVARP
                      LLHEFYEDNSTWDVHQQFLWGPGLLITPVLDEGAEKAMATVPDAVWYDYETGSQVRWR
                      KQKVEMELPGDKIGLHLRGGYIFPTQQPMTTTLASRKNPLGLIIALDENKEAKGELFW
                      DDGETKDTVANKVYLLCEFSVTQNRLEVNISQSTYKDPNNLAFNEIKILGTEEPSNVT
                      VKHNGVPSQTSPTVTYDSNLKVAIITDIDLLLGEAYTVEWSIKIRDEEKIDCYPDENG
                      ASAENCTARGCIWEASNSSGVPFCYFVNDLYSVSDVQYNSHGATADISLKSSVYANAF
                      PSTPVNPLRLDVTYHKNEMPQFKIYDPNKNRYEVPVPLNIPSMPSSTPEGQLYDVLIK
```



```
                      KNPPGIEIRRKSIGTIIWDSQLLGFTPSDMFIRISTRLPSKYLYGPGETEHRSYRRDL
                      EWHTWGMPSRDQPPGYKKNSYGVHPYYMGLEEDGSAHGVLLLNSNAMDVTFQPLPALT
                      YRTTGGVLDFYVFLGPTPELVTQQYTELIGRPVMVPYWSLGFQLCRYGYQNDSEIASL
                      YDEMVAAQIPYDVQYSDIDYMERQLDFTLSPKFAGFPALINRMKADGMRVILILDPAI
                      SGNETQPYPAFTRGVEDDVFIKYPNDGDIVWGKVWPDFPDVVVNGSLDWDSQVELYRA
                      YVAFPDFFRNSTAKWWKREIEELYNNPQNPERSLKFDGMWIDMNEPSSFVNGAVSPGC
                      RDASLNHPPYMPHLESRDRGLSSKTLCMESQQILPDGSLVQHYNVHNLYGWSQTRPTY
                      EAVQEVTGQRGVVITRSTFPSSGRWAGHWLGDNTAAWDQLKKSIIGMTEFSLFGISYT
                      GADICGFFQDAEYEMCVRWMQLGAFTPFSRNHNTIGTRRQDPVSWDVAFVNISRTVLQ
                      TRCTLLPYLYTLMHKAHTEGVTVVRPLLHEFVSDQVTWDIDSQFLLGPAFLVSPVLER
                      NARNVTAYFPRARWYDYFTGVDINARGEWKTLPAPLDHINLHVRGGYILPWQEPALNT
                      HLSRQKFMGFKIALDDEGTAGGWLFWDDGQSIDTYGKGLYYLASFSASQNTMQSHIIF
                      NNYITGTNPLKLGYIEIWGVGSVPVTSASISVSGMVITPSFNNDPTTQVLSIDVTDRN
                      ISLHNPTSLTWISTL"
     misc_feature     103. .165
                      /note="encodes membrane anchor region"
     misc_feature     166. .306
                      /note="encodes S-T rich stalk"
     misc_feature     322. .444
                      /note="encodes trefoil domain 1"
     misc_feature     1633. .1650
                      /note="encodes catalytic site 1"
     misc_feature     2950. .3042
                      /note="encodes trefoil domain 2"
     misc_feature     4306. .4323
                      /note="encodes catalytic site 2"
 BASE COUNT     1747 a   1508 c   1562 g   1696 t
 ORIGIN

   Query Match          100.0%;  Score 174;  DB 27;  Length 6513;
   Best Local Similarity 100.0%;  Pred. No. 1.51e-142;
   Matches   174;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

 Db    6313 GATCTGAGCAGGCAAAGCTCAAAAGAGAGTTTGGAGGTTAAAAATAATTTATTTTTGCAG 6372
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Qy       1 GATCTGAGCAGGCAAAGCTCAAAAGAGAGTTTGGAGGTTAAAAATAATTTATTTTTGCAG 60

 Db    6373 TAGTGTGCTTTGAAATGTGTAAATCTTATTTCTAATGTATACAACCACATTTCACATAAA 6432
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Qy      61 TAGTGTGCTTTGAAATGTGTAAATCTTATTTCTAATGTATACAACCACATTTCACATAAA 120

 Db    6433 AATATGCAATTTTATATGCCAGATAAAAATAAAACAAGTGAATTTGCAAGTGAAA 6486
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
 Qy     121 AATATGCAATTTTATATGCCAGATAAAAATAAAACAAGTGAATTTGCAAGTGAAA 174


 RESULT    2
 LOCUS       G25624        174 bp    DNA              STS       31-MAY-1996
 DEFINITION  human STS EST51018.
 ACCESSION   G25624
 NID         g1347856
```



```
 KEYWORDS    STS sequence; primer; sequence tagged site.
 SOURCE      human STSs derived from sequences in dbEST and the Unigene
             collection.
   ORGANISM  Homo sapiens
             Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
             Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE   1  (bases 1 to 174)
   AUTHORS   Hudson,T.
   TITLE     Whitehead Institute/MIT Center for Genome Research; Physically
             Mapped STSs
   JOURNAL   Unpublished (1995)
 COMMENT
             Contact: Thomas Hudson
             Whitehead Institute/MIT Center for Genome Research
             Whitehead Institute for Biomedical Research
             9 Cambridge Center, Cambridge MA 02142 USA
             Tel: 617 252 1900
             Fax: 617 252 1902
             Email: thudson@genome.wi.mit.edu

             Primer A: GATCTGAGCAGGCAAAGCTC
             Primer B: TTTTATCTGGCATATAAATTGCAT
             STS size: 147
             PCR Profile:
                  Presoak:
                  Denaturation:
                  Annealing: 56 degrees C
                  Polymerization:
                  PCR Cycles: 35
                  Thermal Cycler:
             Protocol:
                 Template: 10 ng
                 Primer: each 5 pM
                 dNTPs: each 4 nM
                 Taq Polymerase: 0.025 units/ul
                 Total Vol: 20 ul

             Buffer:
                 MgCl2: 1.5 mM
                 KCl: 50 mM
                 Tris-HCL: 10 mM
                 pH: 9.3

             Derived from dbEST (genbank accession D20391).
 FEATURES             Location/Qualifiers
     source           1. .174
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /map="643.0 cR from top of Chr7 linkage group"
     STS              1. .147
     primer_bind      1. .20
```

```
     primer_bind     complement(124. .147)
BASE COUNT       68 a      18 c      30 g      51 t       7 others
ORIGIN

  Query Match           96.0%;  Score 167;  DB 30;  Length 174;
  Best Local Similarity  96.0%;  Pred. No. 1.14e-135;
  Matches  167;  Conservative    0;  Mismatches   7;  Indels   0;  Gaps   0;

Db       1 GATCTGAGCAGGCAAAGCTCAAAAGAGAGTTTGGAGGNTAAAAATAATTTATTTTTGCAG 60
           |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Qy       1 GATCTGAGCAGGCAAAGCTCAAAAGAGAGTTTGGAGGTTAAAAATAATTTATTTTTGCAG 60

Db      61 TAGTGTGCTTTGAAATGTGTAAATNTTATTNNTAATGTATACAACCACATTTCACATAAA 120
           |||||||||||||||||||||||| |||||  |||||||||||||||||||||||||||||
Qy      61 TAGTGTGCTTTGAAATGTGTAAATCTTATTTCTAATGTATACAACCACATTTCACATAAA 120

Db     121 ANTATGCAATTTATATGCCAGATAAAANTAAAACAAGTGAATTTGCAAGNGAAA 174
           | ||||||||||||||||||||||||| |||||||||||||||||||| ||||
Qy     121 AATATGCAATTTATATGCCAGATAAAAATAAAACAAGTGAATTTGCAAGTGAAA 174


RESULT    3
LOCUS       G27635         174 bp    DNA             STS       28-JUN-1996
DEFINITION  human STS SHGC-32830.
ACCESSION   G27635
NID         g1396354
KEYWORDS    STS sequence; primer; sequence tagged site.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 174)
  AUTHORS   Myers,R.M.
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Richard M. Myers
            Stanford Human Genome Center (SHGC)
            Stanford University School of Medicine
            Department of Genetics, M-344, Stanford, CA 94305, USA
            Tel: 4157259687
            Fax: 4157259689
            Email: myers@shgc.stanford.edu

            Primer A: GATCTGAGCAGGCAAAGCTC
            Primer B: TTTTATCTGGCATATAAATTGCAT
            STS size: 147
            PCR Profile:
                    Initial incubation: 94 degrees C for 90 seconds

                    Denaturation:       94 degrees C for 15 seconds
                    Annealing:          62 degrees C for 23 seconds
```



```
                    Polymerization:     72 degrees C for 30 seconds
                    PCR Cycles:         30
                    Thermal Cycler:     Perkin Elmer 9600
            Protocol:
                    Template:           25 ng
                    Primer:             each 1 uM
                    dNTPs:              each 200 uM
                    Taq Polymerase:     0.05 units/ul
                    Total Vol:          10 ul

            Buffer:
                    MgCl2:              2.5 mM
                    KCl:                50  mM
                    Tris-HCl:           20  mM
                    pH:                 8.3

            Prepared with primer pairs provided by Sandoz, derived from D20391
            -- Washington University/Merck EST sequence.
FEATURES            Location/Qualifiers
     source         1. .174
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /map="7"
     STS            1. .147
     primer_bind    1. .20
     primer_bind    complement(124. .147)
BASE COUNT       68 a      18 c      30 g      51 t       7 others
ORIGIN

  Query Match           96.0%;  Score 167;  DB 30;  Length 174;
  Best Local Similarity  96.0%;  Pred. No. 1.14e-135;
  Matches  167;  Conservative    0;  Mismatches   7;  Indels   0;  Gaps   0;

Db       1 GATCTGAGCAGGCAAAGCTCAAAAGAGAGTTTGGAGGTTAAAAATAATTTATTTTTGCAG 60
           |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Qy       1 GATCTGAGCAGGCAAAGCTCAAAAGAGAGTTTGGAGGTTAAAAATAATTTATTTTTGCAG 60

Db      61 TAGTGTGCTTTGAAATGTGTAAATNTTATTNNTAATGTATACAACCACATTTCACATAAA 120
           |||||||||||||||||||||||| |||||  |||||||||||||||||||||||||||||
Qy      61 TAGTGTGCTTTGAAATGTGTAAATCTTATTTCTAATGTATACAACCACATTTCACATAAA 120

Db     121 ANTATGCAATTTATATGCCAGATAAAANTAAAACAAGTGAATTTGCAAGNGAAA 174
           | ||||||||||||||||||||||||| |||||||||||||||||||| ||||
Qy     121 AATATGCAATTTATATGCCAGATAAAAATAAAACAAGTGAATTTGCAAGTGAAA 174


RESULT    4
LOCUS       KMARS1G      1267 bp    DNA             PLN       16-DEC-1994
DEFINITION  K.marxianus (ATCC 12424) DNA with ARS1 and centromeric functions
            (1267bp).
ACCESSION   Z31562
```



```
NID         g557862
KEYWORDS    ARS1; autonomous replicating element; centromeric-like DNA.
SOURCE      Kluyveromyces marxianus.
  ORGANISM  Kluyveromyces marxianus
            Eukaryotae; mitochondrial eukaryotes; Fungi; Ascomycota;
            Hemiascomycetes; Saccharomycetales; Saccharomycetaceae;
            Kluyveromyces.
REFERENCE   1  (bases 1 to 1267)
  AUTHORS   Iborra,F. and Ball,M.M.
  TITLE     Kluyveromyces marxianus small DNA fragments contain both autonomous
            replicative and centromeric elements that also function in
            Kluyveromyces lactis
  JOURNAL   Yeast 10 (12), 1621-1629 (1994)
  MEDLINE   95242837
REFERENCE   2  (bases 1 to 1267)
  AUTHORS   Iborra,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-MAR-1994) Francois Iborra, Biologie Genetique et
            Moleculaire, Universite Paris-Sud Institut de Genetique et
            Microbiologie, Bat 400, ORSAY, 91405 ORSAY Cedex, FRANCE
FEATURES            Location/Qualifiers
     source         1. .1267
                    /organism="Kluyveromyces marxianus"
                    /strain="ATCC 12424"
                    /db_xref="taxon:4911"
     misc_feature   196. .389
                    /function="centromere A"
     misc_feature   881. .892
                    /function="ARS 1 core sequence"
BASE COUNT      451 a    150 c    158 g    508 t
ORIGIN

  Query Match           12.1%;  Score 21;  DB 23;  Length 1267;
  Best Local Similarity 100.0%;  Pred. No. 1.94e-01;
  Matches   21;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     412 CAAAAATAAATTATTTTTAAC 432
           |||||||||||||||||||||
Cp      57 CAAAAATAAATTATTTTTAAC 37


RESULT    5
LOCUS       AC003109     76850 bp    DNA             PRI       17-MAR-1998
DEFINITION  Human DNA from overlapping chromosome 7 PAC and P1 clones
            containing the XRCC2 gene, genomic sequence, complete sequence.
ACCESSION   AC003109
NID         g2961444
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
```



```
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 76850)
  AUTHORS   Liu,N., Lamerdin,J.E., Tebbs,R.S., Schild,D., Tucker,J.D., Shen,R.,
            Brookman,K.W., Siciliano,M.J., Walter,C.A., Fan,W., Narayana,L.S.,
            Zhou,Z.-Q., Adamson,A.W., Sorenson,K.J., Chen,D.J., Jones,N.J. and
            Thompson,L.H.
  TITLE     XRCC2 and XRCC3, New Human Rad51- Family Members, Promote
            Chromosome Stability and Protect Against DNA Crosslinks and Other
            Damages
  JOURNAL   Mol. Cell (1998) In press
REFERENCE   2  (bases 1 to 76850)
  AUTHORS   Lamerdin,J.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-NOV-1997) Human Genome Center, Lawrence Livermore
            National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
REFERENCE   3  (bases 1 to 76850)
  AUTHORS   Lamerdin,J.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAR-1998) Human Genome Center, Lawrence Livermore
            National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
COMMENT     Sequence derived from PAC clone 13620 from bases 1 to 10,382, and
            from P1 clone 7515 from bases 10,095 to 76,850. No polymorphisms or
            discrepancies were detected in the overlapping 288 base pairs.
FEATURES            Location/Qualifiers
     source         1. .76850
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="hsXRCC2"
                    /chromosome="7"
                    /map="7q36"
                    /note="P1 clone 7515 obtained from Genome Systems (library
                    constructed from male foreskin-derived fibroblast line).
                    PAC clone 13620 obtained from Genome Systems (library
                    constructed from male leukocytes by P. de Jong)."
     repeat_region  2. .208
                    /rpt_family="Alu"
     repeat_region  complement(209. .272)
                    /rpt_family="MER42"
     repeat_region  complement(380. .442)
                    /rpt_family="Alu"
     repeat_region  complement(497. .813)
                    /rpt_family="L1"
     repeat_region  complement(825. .1106)
                    /rpt_family="Alu"
     repeat_region  1178. .1464
                    /rpt_family="Alu"
     repeat_region  complement(1789. .1899)
                    /rpt_family="L1"
     repeat_region  2034. .2157
                    /rpt_family="Alu"
     repeat_region  2731. .3063
```

```
                     /rpt_family="Alu"
     repeat_region   complement(3549. .3789)
                     /rpt_family="Alu"
     mRNA            4236. .33903
                     /note="3'-UTR contains multiple repetitive elements."
                     /product="XRCC2"
     gene            4322. .31759
                     /note="X-ray repair cross-complementing DNA repair gene"
                     /gene="XRCC2"
     CDS             join(4322. .4360,19619. .19700,31038. .31759)
                     /gene="XRCC2"
                     /note="X-ray repair cross-complementing DNA repair
                     protein"
                     /codon_start=1
                     /product="Xrcc2"
                     /db_xref="PID:g2961445"
                     /translation="MCSAFHRAESGTELLARLEGRSSLKEIEPNLFADEDSPVHGDIL
                     EFHGPEGTGKTEMLYHLTARCILPKSEGGLEVEVLFIDTDYHFDMLRLVTILEHRLSQ
                     SSEEIIKYCLGRFFLVYCSSSTHLLLTLYSLESMPCSHPSLCLLILDSLSAFYWIDRV
                     NGGESVNLQESTLRKCSQCLEKLVNDYRLVLFATTQTIMQKASSSSEEPSHASRRLCD
                     VDIDYRPYLCKAWQQLVKHRMPFSKQDDSQSSNQFSLVSRCLKSNSLKKHFFIIGESG
                     VEFC"
     repeat_region   5618. .5774
                     /rpt_family="HSATI"
     repeat_region   complement(5783. .6067)
                     /rpt_family="Alu"
     repeat_region   complement(6510. .6810)
                     /rpt_family="Alu"
     repeat_region   6948. .7208
                     /rpt_family="Alu"
     repeat_region   complement(8375. .8630)
                     /rpt_family="Alu"
     repeat_region   8887. .9129
                     /rpt_family="MER42"
     repeat_region   complement(9181. .9456)
                     /rpt_family="Alu"
     repeat_region   9515. .9795
                     /rpt_family="Alu"
     repeat_region   complement(10534. .10574)
                     /rpt_family="Alu"
     repeat_region   10709. .10968
                     /rpt_family="Alu"
     repeat_region   12008. .12076
                     /rpt_family="MER44C"
     repeat_region   12109. .12326
                     /rpt_family="MER44C"
     repeat_region   complement(12359. .12632)
                     /rpt_family="Alu"
     repeat_region   complement(12978. .13268)
                     /rpt_family="Alu"
     repeat_region   13495. .13780
```



```
                     /rpt_family="Alu"
     repeat_region   13900. .14183
                     /rpt_family="Alu"
     repeat_region   14641. .14913
                     /rpt_family="Alu"
     repeat_region   complement(15901. .16104)
                     /rpt_family="MLT2B2"
     repeat_region   complement(16576. .16844)
                     /rpt_family="Alu"
     repeat_region   17529. .17829
                     /rpt_family="Alu"
     repeat_region   18140. .18430
                     /rpt_family="Alu"
     repeat_region   complement(19259. .19503)
                     /rpt_family="Alu"
     misc_feature    complement(20710. .20887)
                     /gene="XRCC2"
                     /note="BLASTX similarity to  (1088. .1146); match: 0.45,
                     score: 9.0e-06; database searched: nr; hypothetical
                     protein (L1H 3' region) - human"
     repeat_region   20942. .21055
                     /rpt_family="Alu"
     repeat_region   21253. .21526
                     /rpt_family="L1"
     repeat_region   21931. .22256
                     /rpt_family="Alu"
     repeat_region   22264. .22440
                     /rpt_family="Alu"
     repeat_region   22820. .23070
                     /rpt_family="Alu"
     repeat_region   complement(23153. .23436)
                     /rpt_family="Alu"
     repeat_region   complement(24031. .24148)
                     /rpt_family="Alu"
     repeat_region   complement(24205. .24387)
                     /rpt_family="L1"
     repeat_region   complement(24404. .24958)
                     /rpt_family="Alu"
     repeat_region   complement(26186. .26513)
                     /rpt_family="Alu"
     repeat_region   26621. .26930
                     /rpt_family="Alu"
     repeat_region   complement(27018. .27369)
                     /rpt_family="THE1"
     repeat_region   complement(28211. .28504)
                     /rpt_family="Alu"
     repeat_region   28597. .28774
                     /rpt_family="Alu"
     repeat_region   complement(28851. .29076)
                     /rpt_family="Alu"
     repeat_region   29170. .29452
```



```
     repeat_region   complement(29487. .29781)
                     /rpt_family="Alu"
     misc_feature    31037. .31372
                     /gene="XRCC2"
                     /note="BLASTN similarity to Y08837 (1. .336); match: 0.99,
                     score: 7.0e-130; database searched: nt; H.sapiens mRNA for
                     RAD51-like protein--Other overlapping matches:-BLASTX
                     similarity to PID (93. .149, 206. .243); match: 0.43, score:
                     9.7e-06; database searched: nr; (Y13144) Rad51 homologue
                     [Trypanosoma brucei]--BLASTX similarity to 2262211
                     (101. .162); match: 0.43, score: 1.8e-06; database
                     searched: nr; (U92068) RecA-like protein [Mus musculus]--"
     repeat_region   complement(31829. .32256)
                     /rpt_family="Alu"
     repeat_region   32533. .33613
                     /rpt_family="L1"
     misc_feature    32880. .33124
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 1, quality: good, score: 67.000"
     repeat_region   33635. .33888
                     /rpt_family="Alu"
     repeat_region   33903. .34050
                     /rpt_family="L1"
     repeat_region   complement(34480. .34609)
                     /rpt_family="Alu"
     repeat_region   35004. .35121
                     /rpt_family="Alu"
     misc_feature    36278. .36522
                     /note="DDS similarity to AA255626 zs31f06.s1 NCI_CGAP_GCB1
                     Homo sapiens cDNA clone 686819 3' similar to contains Alu
                     repetitive element;contains element LTR6 repetitive
                     element ;(186. .427); 96% identity.-"
     repeat_region   complement(36396. .36646)
                     /rpt_family="Alu"
     repeat_region   36742. .37029
                     /rpt_family="Alu"
     repeat_region   complement(37704. .38011)
...
Note: remainder of annotations omitted.

  Query Match           11.5%;  Score 20;  DB 26;  Length 76850;
  Best Local Similarity 100.0%;  Pred. No. 8.92e-01;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  63076 AAACAAGTGAATTTGCAAGT 63095
          ||||||||||||||||||||
Qy    151 AAACAAGTGAATTTGCAAGT 170


RESULT    6
```



```
LOCUS       AC004986   124752 bp    DNA             HTG       12-JUN-1998
DEFINITION  *** SEQUENCING IN PROGRESS *** Homo sapiens clone DJ1166A24; HTGS
            phase 1, 4 unordered pieces.
ACCESSION   AC004986
NID         g3213011
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 124752)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 124752)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-1998) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     ***                                                          ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                          ***
            * This sequence is unfinished. It consists of 4 contigs for
            * which the order is not known; their order in this record is
            * arbitrary. In some cases, the exact lengths of the gaps
            * between the contigs are also unknown; these gaps are  presented
            * as runs of N as a convenience only. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
            *       1     5883: contig of 5883 bp in length
            *    5884     5898: gap of unknown length
            *    5899    37984: contig of 32086 bp in length
            *   37985    37999: gap of unknown length
            *   38000    76330: contig of 38331 bp in length
            *   76331    76345: gap of unknown length
            *   76346   124752: contig of 48407 bp in length.
FEATURES              Location/Qualifiers
     source           1. .124752
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="DJ1166A24"
BASE COUNT    40728 a  24064 c  22602 g  37311 t     47 others
ORIGIN

  Query Match           11.5%;  Score 20;  DB 17;  Length 124752;
  Best Local Similarity 100.0%;  Pred. No. 8.92e-01;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  23799 AATGTATACAACCACATTTC 23818
          ||||||||||||||||||||
```

```
Qy      94 AATGTATACAACCACATTTC 113


RESULT    7
LOCUS       AC003037    213863 bp    DNA             PRI       06-OCT-1998
DEFINITION  Homo sapiens Xp22 BAC GSHB 526D21 (Genome Systems Human BAC
            library) complete sequence.
ACCESSION   AC003037
NID         g2920805
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 213863)
  AUTHORS   Muzny,D., Arenson,A.D., Brundage,E., Carvelli,K., Chen,E., Chen,J.,
            Di,W., Ding,Y., Dugan,S., Durbin,J., Forcum,J., Ganesh,R.,
            Garcia,C., Goodman,M., Gorrell,J.H., Haywood,M., Jackson,L.,
            Jin,S., Kampal,R., Karpathy,S., Leal,B., Li,Y., Liu,W., Logan,O.,
            Lu,J., Ly,T., Martinez,C., Oswal,G., Perez,L., Rashid,N.D.,
            Rowland,K., Savage,L., Scherer,S.S., Shen,H., Timms,K.M., Todd,J.,
            Vo,Q., Worley,K.C., Yu,W., Chinault,C., Nelson,D. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 213863)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-OCT-1997) Molecular and Human Genetics, Baylor
            College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 213863)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-FEB-1998) Molecular and Human Genetics, Baylor
            College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4  (bases 1 to 213863)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAR-1998) Molecular and Human Genetics, Baylor
            College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE   5  (bases 1 to 213863)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-1998) Molecular and Human Genetics, Baylor
            College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE   6  (bases 1 to 213863)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-OCT-1998) Molecular and Human Genetics, Baylor
            College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Mar 2, 1998 this sequence version replaced gi:2804324.
            Sequencing is completed to a minimum standard of double strand
```



```
            coverage with a minimum of 2 clones and 2 reads with no ambiguities
            or 2 chemistries with a minimum of 2 clones and 3 reads with no
            ambiguities. If the sequence quality does not meet this standard,
            it will be indicated in the annotation.

            The repeat regions shown were identified using RepeatMasker by
            Adrian Smit.

            Sequence similarities were identified using Powerblast by Jinghui
            Zhang.

            Exon/Intron boundaries of identified genes were chosen if there
            were canonical splice junctions that maintained sequence continuity
            across the splice junctions.
FEATURES              Location/Qualifiers
     source           1. .213863
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="GSHB-526D21"
                      /chromosome="X"
                      /clone_lib="Genome Systems Human BAC library"
                      /map="Xp22"
     misc_feature     1. .303
                      /note="Region: AA029104 zk09b01.r1 Soares pregnant uterus
                      NbHPU Homo sapiens cDNA clone 469993 5'"
     misc_feature     complement(380. .697)
                      /note="Region: AA043674 zk50a09.s1 Soares pregnant uterus
                      NbHPU Homo sapiens cDNA clone 486232 3'"
     repeat_region    complement(1102. .1291)
                      /rpt_family="MER3"
     repeat_region    complement(1346. .1387)
                      /rpt_family="POLY_A"
     misc_feature     complement(1738. .2143)
                      /note="Region: AA740623 ny97g07.s1 NCI_CGAP_GCB1 Homo
                      sapiens cDNA clone IMAGE:1286268"
     repeat_region    complement(2794. .2899)
                      /rpt_family="L2"
     repeat_region    3015. .3044
                      /rpt_family="AT_rich"
     repeat_region    complement(3128. .3423)
                      /rpt_family="AluJb"
     repeat_region    complement(3949. .4245)
                      /rpt_family="AluJb"
     repeat_region    complement(4247. .4607)
                      /rpt_family="L1MB7"
     repeat_region    6043. .6085
                      /rpt_family="L2"
     repeat_region    6389. .6517
                      /rpt_family="MIR"
     repeat_region    complement(6975. .7023)
                      /rpt_family="L2"
```



```
     repeat_region    complement(7007. .7092)
                      /rpt_family="MIR"
     repeat_region    complement(7787. .8093)
                      /rpt_family="AluSx"
     repeat_region    complement(8291. .8318)
                      /rpt_family="AT_rich"
     repeat_region    complement(9569. .9872)
                      /rpt_family="AluSx"
     repeat_region    complement(10250. .10411)
                      /rpt_family="MER3"
     repeat_region    11214. .11510
                      /rpt_family="AluJo"
     repeat_region    11602. .11660
                      /rpt_family="MIR"
     misc_feature     complement(12431. .12991)
                      /note="Region: AA563671 ng49e02.s1 NCI_CGAP_Co3 Homo
                      sapiens cDNA clone IMAGE:938138 similar to SW:RAB9_CANFA
                      P24408 RAS-RELATED PROTEIN RAB-9"
     repeat_region    13871. .14137
                      /rpt_family="AluSc"
     misc_feature     15421. .15925
                      /note="Region: AA826182 of33g12.s1 NCI_CGAP_Kid6 Homo
                      sapiens cDNA clone IMAGE:1422982"
     repeat_region    complement(16053. .16074)
                      /rpt_family="AT_rich"
     repeat_region    16136. .16293
                      /rpt_family="AluSp"
     repeat_region    16342. .16644
                      /rpt_family="AluSq"
     repeat_region    16710. .16999
                      /rpt_family="AluSg"
     repeat_region    17038. .17335
                      /rpt_family="AluSc"
     repeat_region    18365. .18654
                      /rpt_family="AluSx"
     repeat_region    20209. .20243
                      /rpt_family="POLY_A"
     repeat_region    20634. .20920
                      /rpt_family="AluSq"
     repeat_region    20946. .21375
                      /rpt_family="MLT1C"
     repeat_region    22048. .22349
                      /rpt_family="AluSx"
     repeat_region    23918. .24220
                      /rpt_family="AluJb"
     repeat_region    24370. .24596
                      /rpt_family="MIR"
     repeat_region    complement(24994. .25130)
                      /rpt_family="MER58A"
     repeat_region    25383. .25414
                      /rpt_family="POLY_A"
```



```
     repeat_region    25856. .25953
                      /rpt_family="MER3"
     repeat_region    25956. .26288
                      /rpt_family="AluY"
     repeat_region    complement(26289. .26373)
                      /rpt_family="MER33"
     repeat_region    complement(26397. .26641)
                      /rpt_family="AluSx"
     repeat_region    26965. .27292
                      /rpt_family="MLT1D"
     repeat_region    complement(27563. .27593)
                      /rpt_family="AT_rich"
     repeat_region    27807. .32800
                      /rpt_family="L1PA2"
     STS              32380. .32557
                      /standard_name="STS1-cSRL-31b6-uA/cSRL-31b6-uE, Chr. -,
                      Homo sapiens"
                      /db_xref="dbSTS:7821"
     repeat_region    32801. .33102
                      /rpt_family="AluSp"
     repeat_region    33106. .34245
                      /rpt_family="L1PA2"
     repeat_region    34748. .34864
                      /rpt_family="L2"
     STS              35687. .35752
                      /standard_name="HUMSWX1791, Chr. X, Homo sapiens"
                      /db_xref="dbSTS:13087"
     repeat_region    37266. .37472
                      /rpt_family="MER20"
     repeat_region    complement(40152. .40194)
                      /rpt_family="L2"
     repeat_region    complement(40214. .40291)
                      /rpt_family="MIR"
     repeat_region    complement(40560. .40761)
                      /rpt_family="L1MC1"
     repeat_region    40763. .40844
                      /rpt_family="AluSg/x"
     repeat_region    complement(41023. .41322)
                      /rpt_family="AluSx"
     repeat_region    complement(42296. .42322)
                      /rpt_family="AT_rich"
     repeat_region    42638. .42936
                      /rpt_family="AluSp"
     repeat_region    complement(43191. .43228)
                      /rpt_family="(TGAA)n"
     repeat_region    43415. .43591
                      /rpt_family="MER3"
     repeat_region    43685. .43982
                      /rpt_family="AluSx"
     repeat_region    44031. .44325
                      /rpt_family="AluSc"
```

...
Note: remainder of annotations omitted.

Query Match 11.5%; Score 20; DB 27; Length 213863;
Best Local Similarity 100.0%; Pred. No. 8.92e-01;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 62426 CACTTGTTTTATTTTTATCT 62445
||||||||||||||||||||
Cp 159 CACTTGTTTTATTTTTATCT 140

RESULT 8
LOCUS HUMPROS05 247 bp DNA PRI 08-JAN-1995
DEFINITION Human protein S alpha gene (PROS1), exon 6.
ACCESSION M36555 J02918
NID g190431
KEYWORDS S protein.
SEGMENT 5 of 14
SOURCE Human DNA
ORGANISM Homo sapiens
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 247)
AUTHORS Ploos van Amstel,H.K., Reitsma,P.H., van der Logt,C.P. and
Bertina,R.M.
TITLE Intron-exon organization of the active human protein S gene PS
alpha and its pseudogene PS beta: duplication and silencing during
primate evolution
JOURNAL Biochemistry 29 (34), 7853-7861 (1990)
MEDLINE 91084445
COMMENT Draft entry and computer-readable sequence for [Biochemistry 29,
7853-4861 (1990)] kindly submitted
by H.K.Ploos van Amstel, 13-JUL-1990.
FEATURES Location/Qualifiers
source 1..247
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="3p11-q11.2"
intron <1..48
/gene="PROS1"
/note="intron 5"
prim_transcript <1..>247
/gene="PROS1"
/note="protein S alpha mRNA and introns"
exon 49..180
/gene="PROS1"
/note="G00-120-721"
/number=6
intron 181..>247
/gene="PROS1"

/note="intron 6"
BASE COUNT 90 a 29 c 43 g 85 t
ORIGIN

Query Match 10.9%; Score 19; DB 25; Length 247;
Best Local Similarity 100.0%; Pred. No. 3.93e+00;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 18 TTAAAAATAATTTATTTTT 36
||||||||||||||||||||
Qy 38 TTAAAAATAATTTATTTTT 56

RESULT 9
LOCUS HUMPS05 511 bp DNA PRI 10-JAN-1995
DEFINITION Human S protein-alpha (PS-alpha) gene, exons 5 and 6.
ACCESSION M57844 J02917
NID g190538
KEYWORDS S protein; anticoagulant cofactor; vitamin K-dependent protein.
SEGMENT 5 of 14
SOURCE Human liver DNA.
ORGANISM Homo sapiens
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 511)
AUTHORS Schmidel,D.K., Tatro,A.V., Phelps,L.G., Tomczak,J.A. and Long,G.L.
TITLE Organization of the human protein S genes
JOURNAL Biochemistry 29 (34), 7845-7852 (1990)
MEDLINE 91084444
FEATURES Location/Qualifiers
source 1..511
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="liver"
/map="3p11-q11.2"
intron order(M57843:263..464,1..103)
/gene="PS-alpha"
exon 104..226
/gene="PROS1"
/note="G00-120-721"
/number=5
intron 227..331
/gene="PROS1"
/number=5
exon 332..463
/gene="PROS1"
/number=6
BASE COUNT 169 a 82 c 93 g 167 t
ORIGIN

Query Match 10.9%; Score 19; DB 25; Length 511;

Best Local Similarity 100.0%; Pred. No. 3.93e+00;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 301 TTAAAAATAATTTATTTTT 319
||||||||||||||||||||
Qy 38 TTAAAAATAATTTATTTTT 56

RESULT 10
LOCUS HUMPSPS02 540 bp DNA PRI 10-JAN-1995
DEFINITION Human protein S-beta pseudogene, exons 5 and 6.
ACCESSION M57855 J02917
NID g190676
KEYWORDS protein S; pseudogene.
SEGMENT 2 of 11
SOURCE Human liver DNA.
ORGANISM Homo sapiens
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 540)
AUTHORS Schmidel,D.K., Tatro,A.V., Phelps,L.G., Tomczak,J.A. and Long,G.L.
TITLE Organization of the human protein S genes
JOURNAL Biochemistry 29 (34), 7845-7852 (1990)
MEDLINE 91084444
FEATURES Location/Qualifiers
source 1..540
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="liver"
/map="3p21-cen"
intron order(M57854:155..194,1..113)
/gene="PS-beta"
/number=4
exon 114..236
/gene="PROS2"
/note="G00-120-757"
/number=5
intron 237..341
/gene="PROS2"
/number=5
exon 342..473
/gene="PROS2"
/number=6
BASE COUNT 171 a 85 c 99 g 185 t
ORIGIN

Query Match 10.9%; Score 19; DB 25; Length 540;
Best Local Similarity 100.0%; Pred. No. 3.93e+00;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 311 TTAAAAATAATTTATTTTT 329

||||||||||||||||||||
Qy 38 TTAAAAATAATTTATTTTT 56

RESULT 11
LOCUS SPAC1B9 19108 bp DNA PLN 10-APR-1996
DEFINITION S.pombe chromosome I cosmid c1B9.
ACCESSION Z70720
NID g1262147
KEYWORDS 5S rRNA; alpha-amylase; RTS1 homologue; signal recognition
particle; SRP54; transfer RNA-Lys.
SOURCE fission yeast.
ORGANISM Schizosaccharomyces pombe
Eukaryotae; mitochondrial eukaryotes; Fungi; Ascomycota;
Archaeascomycetes; Schizosaccharomycetales;
Schizosaccharomycetaceae; Schizosaccharomyces.
REFERENCE 1 (bases 1 to 19108)
AUTHORS Connor,R. and Churcher,C.M.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 19108)
AUTHORS Barrell,B.G., Rajandream,M.A. and Walsh,S.V.
TITLE Direct Submission
JOURNAL Submitted (09-APR-1996) Schizosaccharomyces pombe chromosome I
sequencing project, Sanger Centre, Hinxton Hall, Hinxton, Cambridge
CB10 1RQ E-mail: barrell@sanger.ac.uk
COMMENT Notes:
Details of yeast sequencing at the Sanger Centre are available on
the World Wide Web.
(URL, http://www.sanger.ac.uk/yeast/home.html)
Protein coding regions (CDS) have been predicted with the help of
computer analysis using the Genefinder program in PomBase (an ACEDB
database) with additional predictions for the branch-acceptor sites
supplied by the program Sp3splice. CAUTION: It is possible that for
any individual CDS we may have underestimated or overestimated the
number of introns/exons or we may not have chosen the correct
splice donor/acceptor sites.
CDS are numbered using the following system eg SPAC5H10.01c. SP (S.
pombe), A (chromosome 1), c5H10 (cosmid name), .01 (first CDS), c
(complementary strand).
The more significant matches with motifs in the PROSITE database
are also included but some of these may be fortuitous.
The length in codons is given for each CDS.
IMPORTANT: This sequence MAY NOT be the entire insert of the
sequenced clone. It may be shorter because we only sequence
overlapping sections once, or longer, because we arrange for a
small overlap between neighbouring submissions.
Cosmid c1B9 at the 3' end by cosmid c8A4 embl entry SPAC8A4
accession no. Z66569.
FEATURES Location/Qualifiers
source 1..19108
/organism="Schizosaccharomyces pombe"

```
                     /strain="972h-"
                     /db_xref="taxon:4896"
                     /chromosome="I"
                     /clone="cosmid c1B9"
                     /map="1L"
     gene            1. .750
                     /gene="SPAC1B9.01"
     CDS             <1. .750
                     /gene="SPAC1B9.01"
                     /note="SPAC1B9.01, unknown partial orf, len: > 349"
                     /codon_start=1
                     /product="unknown"
                     /db_xref="PID:e235398"
                     /db_xref="PID:g1262148"
                     /translation="DLGVYNDRYIVAPGQQVAIVKLTEERLMIDSRISLGSIVLQLIV
                     EGNEIAIADSIGRFTIMYFDGQKFIVVARYLPGENIVKAALYEGTVYIIATNSGLLKL
                     LRYNKDAKNFNDRFICESVYHLHDKVSKFQNFPITNTNSFLEPKMLFATEIGAIGSIV
                     SLKDKELELEELTRKIRKLKFSYLSSMDYESIEADLISPVPFIDGDLVIDVKRWASSE
                     LFRLCRSVEHRESLNSTQKVQALLEEIQSLC"
     gene            complement(1235. .2780)
                     /gene="SPAC1B9.02c"
     CDS             complement(join(1235. .1267,1311. .1754,1821. .2780))
                     /gene="SPAC1B9.02c"
                     /note="SPAC1B9.02c, len: 478, similar to alpha-amylases eg
                     SW:AMY1_SACFI P21567 (40.2% identity in 458 aa overlap)"
                     /codon_start=1
                     /product="unknown"
                     /db_xref="PID:e235477"
                     /db_xref="PID:g1262149"
                     /translation="MNLICSILPKKEPSYLKLWRKQVIYQVLTDRFALDEDNFYAKAS
                     GNLYLGGTWKGITRNLDYIKSLGCTAIWISPIVKNISETTDCGQAYHGYWAQDMTQLN
                     ENFGTEEDLKELVNAIHEKNMLCMVDIVVNHMGHAGSKPVNFLLTQPFNSGKTYHNWQ
                     FVQNTEDPHETITGWLGDSHVNLDDIRTEENKEVKRFFQNWVSDLIKRFQFDGIRLDTA
                     KHVEKSFFPFTFIEAANVFTTGEVFHGDPKYVGDTQKYLPSTLNFPLFPQLRETFLDFK
                     HSMFSFYDKAVLDVRHYFKDVTVLCNFLENHDFPRFFHETKDIALALNALTALIPMDG
                     IPIIYYGQEQMFDGGSDPDNREGLWKSKYNTSNPMFRHLSSMIRTRQNLVETTPEFTY
                     VLSFQLYIDDSVTVFTRPGVIIAISNEGGTSSFKVEIDLKEHWKEVPSSFTDILTQKT
                     IPCKDHKLKIKSKSGLPKILISSDPSPL"
     misc_feature    complement(1268. .1282)
                     /gene="SPAC1B9.02c"
                     /note="splice branch and acceptor sequence,
                     ctaacaatttttaag"
     misc_feature    complement(1305. .1310)
                     /gene="SPAC1B9.02c"
                     /note="splice donor sequence, gtaagc"
     misc_feature    complement(1755. .1769)
                     /gene="SPAC1B9.02c"
                     /note="splice branch and acceptor sequence,
                     ttaacatttatttag"
     misc_feature    complement(1815. .1820)
                     /gene="SPAC1B9.02c"
```



```
                     /note="splice donor sequence, gtaagt"
     misc_feature    complement(3713. .3831)
                     /gene="5sRNA"
                     /note="5s ribosomal RNA gene"
     gene            complement(3713. .3831)
                     /gene="5sRNA"
     misc_feature    4259. .4263
                     /note="AATAC sequence duplicated by tnl000 transposition
                     of cosmid c1B9, the tnl000 and the duplicated 5bp sequence
                     have been removed from this sequence"
     tRNA            4488. .4570
                     /gene="tRNA-Lys"
                     /note="tRNA-Lys, anticodon ctt, potential intron between
                     positions 4527-4534, len: 83"
     gene            4488. .4570
                     /gene="tRNA-Lys"
     gene            5231. .7432
                     /gene="SPAC1B9.03"
     CDS             join(5231. .5254,5298. .5393,5434. .5647,5979. .6733,
                     6786. .7084,7174. .7432)
                     /gene="SPAC1B9.03"
                     /note="SPAC1B9.03, len: 548, similar to SW:RTS1_YEAST
                     P38903 RTS1 protein PIR:S54620 (54.0% identity in 441 aa
                     overlap)"
                     /codon_start=1
                     /product="unknown"
                     /db_xref="PID:e235399"
                     /db_xref="PID:g1262150"
                     /translation="MKGIKSKMLSRGKSQDTQKSSKKKESKKSNSHDSSKAPKESPST
                     DPNGSVIGAQNDFLTVPKHSGKKVPIDTTPTPRDEILLENVRTVRKQRSSLYHISENR
                     NLVRLPSFTDVPVNKWHSLALEKLEQCCVVPDFNDPSTDLYGKEVKREALQDLIDLIS
                     VRKEAIDESLYPSIVHMFAVNVFRPLPPPSNPPGEIMDLEEDEPALEVAWPHLHLVTD
                     FFLRPFESPSLNTSVAKVYINQKPIRKLLVLFDSEDPRERDFLKTTLERITGKFLGLR
                     AFIRRSINNLFLQFVYENEQFNGIAELLEILGSIINGFALPLKEEHKIFLSRVLIPLH
                     KAKSLPLIYPQIAYGIVQFVERDSSVTEEVVLGLLRTWPKVNSSKEVLFLNEIEDIE
                     VMEPSEFLKIQVPLFHKLATSISSQNFQVAERALYFFNNDYFVHLVEENVDIILPIIY
                     PALFEISKSHWNRVIHSMVCNVLKLFMDINPSLFDEVDAETSESRRKKEDEEIIREER
                     WTILENIAKENAMKLKSQNPTTVHSTTERLKKLSLDYTNG"
     misc_feature    5255. .5260
                     /gene="SPAC1B9.03"
                     /note="splice donor sequence, gtaaga"
     misc_feature    5283. .5297
                     /gene="SPAC1B9.03"
                     /note="splice branch and acceptor sequence,
                     ttaaccagaatctag"
     misc_feature    5394. .5399
                     /gene="SPAC1B9.03"
                     /note="splice donor sequence, gtacgg"
     misc_feature    5419. .5433
                     /gene="SPAC1B9.03"
                     /note="splice branch and acceptor sequence,
```



```
                     ttaacattttcatag"
     misc_feature    5648. .5653
                     /gene="SPAC1B9.03"
                     /note="splice donor sequence gtatgt"
     misc_feature    5962. .5978
                     /gene="SPAC1B9.03"
                     /note="splice branch and acceptor sequence,
                     ctaacctgtctatttag"
     misc_feature    6734. .6739
                     /gene="SPAC1B9.03"
                     /note="splice donor sequence, gtatgt"
     misc_feature    6768. .6785
                     /gene="SPAC1B9.03"
                     /note="splice branch and acceptor sequence,
                     ctaacgatgacaatatag"
     misc_feature    7085. .7090
                     /gene="SPAC1B9.03"
                     /note="splice donor sequence, gtaagt"
     misc_feature    7154. .7173
                     /gene="SPAC1B9.03"
                     /note="splice branch and acceptor sequence,
                     ctgacgttttgttag"
     gene            7758. .10595
                     /gene="SPAC1B9.04"
     CDS             join(7758. .9671,9712. .9843,9904. .10132,10181. .10290,
                     10353. .10595)
                     /gene="SPAC1B9.04"
                     /note="SPAC1B9.04, unknown, len: 875, but splicing
                     uncertain at C-terminal region"
                     /codon_start=1
                     /product="unknown"
                     /db_xref="PID:e235478"
                     /db_xref="PID:g1262151"
                     /translation="MSCIQIISSSQTSIAGHRKLCNKLPTLRTQEGFETDILRALNII
                     LTVKKGNSNADRVLRFLVTFVNYLQQKDPEIDIVQPILKHILRGLDAKDRTVRYRCCQ
                     IIARVVNCVKEIDDDLYNTLKEKLLSRVLDRESIVRLEAVVALSRLQEDTGDEENDVR
                     NILLFLLQNDPSSEVRRSVLLNIEVSNSTLPFILERARDVDAANRKCVYARVLPKIGD
                     FRYLSIKKRVRILKWGLNDRDESVEKAAADMLAYQWIENADNNLLELLERLDVSNNSD
                     VAVLAIKKFFDVRVDSLSQLEFPEQFWLELTAESSLLARTFNEICIEKNYTDLLDKMP
                     EVVQLTYYIERQYVSLRDKSSYDESCFIIEQLLYIGLSQDMVDEIGRRKLLKSLTNSL
                     STMALPDSLISLHIELLRKLCSSENDFCSLLVEIITEVFPEQGHSQNQTEEQGNSNAPE
                     LNKNDYEGEEITVSQKSPSPSLPPNELNEPEPDDMDGYKEAFNELRCLSYVQCLFENI
                     TSSLNENLYMVDMLKTLIIPAVRSHDLPIREKGLECLSLVCLLNADLAPENVPLYLHC
                     YEKGSVVLKCTAIRTLTDMLIQHGKAKFTEYEDAISSILFEALGEFENAEQYLGAEA
                     IAKLLVILHYRDELFLKPLLICYFEPNTVDNHALRQVLGYFFPVYAFGAHEMQWRIAT
                     IFCDALLSLLEIYRDLDEDDVQLSIGHIAQQMLDWTDNEKLYERKTQTGDDYIALNHN
                     VHLHLANMIFPESLPNASEGKERKFMISLLGKLKIPTDLPSSDYQRTKRKLETYESHGF
                     TMDSTSLSILAKFERMLLQNEEARSKFEETEEERLMENAEENEHAGAEAISGEIIPDT
                     VEANMEDEEEVYVKQEEDL"
     misc_feature    9672. .9677
                     /gene="SPAC1B9.04"
```



```
                     /note="splice donor sequence, gtatgt"
...
Note: remainder of annotations omitted.

  Query Match           10.9%;  Score 19;  DB 23;  Length 19108;
  Best Local Similarity 100.0%;  Pred. No. 3.93e+00;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   6494 AATAATTTATTTTTGCAGT 6512
            |||||||||||||||||||
Qy     43 AATAATTTATTTTTGCAGT 61


RESULT   12
LOCUS       CEK01C8     24725 bp    DNA             INV       21-SEP-1998
DEFINITION  Caenorhabditis elegans cosmid K01C8, complete sequence.
ACCESSION   Z49068
NID         g780187
KEYWORDS    HTG; Aromatic-L-amino-acid decarboxylase; GTP-binding protein;
            Mitochondrial carrier protein; PDXJ; Pyridoxal biosynthetic
            protein; T-complex protein; Threonine dehydratase; Thyroliberin.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 24725)
  AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
  TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
  JOURNAL   Nature 368 (6466), 32-38 (1994)
  MEDLINE   94150718
REFERENCE   2  (bases 1 to 24725)
  AUTHORS   Sims,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-NOV-1996) Nematode Sequencing Project, Sanger Centre,
            Hinxton, Cambridge CB10 1RQ, England and Department of Genetics,
            Washington University, St. Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     Coding sequences below are predicted from computer analysis, using
            predictions from Genefinder (P. Green, U. Washington), and other
            available information.
```

```
            For a graphical representation of this sequence and its analysis
            see:-
            http://webace.sanger.ac.uk/cgi-
            bin/display?db=wormace&class=Sequence &object=K01C8
            Current sequence finishing criteria for the C. elegans genome
            sequencing consortium are that all bases are either sequenced
            unambiguously on both strands, or on a single strand with both a
            dye primer and dye terminator reaction, from distinct subclones.
            Exceptions are indicated by an explicit note.
            IMPORTANT:  This sequence is NOT necessarily the entire insert of
            the specified clone.  It may be shorter because we only sequence
            overlapping sections once, or longer because we arrange for a small
            overlap between neighbouring submissions.
            IMPORTANT: This sequence is not the entire insert of clone K01C8.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we arrange for a small overlap between
            neighbouring submissions.
            The true left end of clone K01C8 is at 15138 in
            sequence Z49966.
            The true right end of clone K01C8 is at 24725 in this sequence. The
            true right end of clone F35C11 is at 10525 in this sequence. The
            start of this sequence (1. .102) overlaps with the end of sequence
            Z49966.
            The end of this sequence (24626. .24725) overlaps with the start of
            sequence AL023831.
FEATURES    Location/Qualifiers
     source 1. .24725
            /organism="Caenorhabditis elegans"
            /db_xref="taxon:6239"
            /chromosome="II"
            /clone="K01C8"
     gene   4042. .6093
            /gene="K01C8.1"
     CDS    join(4042. .4092,4344. .4536,4590. .4708,4766. .5071,
            5128. .5385,5433. .5585,5635. .6093)
            /gene="K01C8.1"
            /note="similar to threonine dehydratase; cDNA EST
            EMBL:M89183 comes from this gene; cDNA EST yk195g4.5 comes
            from this gene"
            /codon_start=1
            /db_xref="PID:g780188"
            /db_xref="SPTREMBL:Q21080"
            /translation="MVAPSRKKSIIHHHDHSHESTPQKTHHHRVKQAERKRTASISKE
            DGMEYNKRVRKLHKAFLTHGAESPHVDDKQTSGSNSDTEAPLFDPDCDPENPKKLKFS
            DISSAAFNIKNGVQTTPCVRSLQLSSKCEMDLYFKKEYLQVTGSPKERGAHYALSKMA
            EQFKAGVIAASAGNHAIALSYHGQQMGIPVTVVMPVIAPLMKIQFCRSLGATVILKG
            ESIAVAKDFALRHAKENHLKYINGYDAIDILAGQGTIGLEILDQVPDVDTILVPVGGG
            GLIAGIATAVKTLKPDVHIYVSSPLIIANFKVTGIESETCPSFTEAYEAGHIITSQAK
            ASLADGLAVPTVGGNSLETAKGLVDEVITVKEESIALSILRLLEVEKAVVEGGGAVGL
            AAILEGKVPELKGKKVVSILSGGNIDTTVLGRSIERGLAVDGRLVRLEVVVSDRPGGI
            AELTTTIAHLGASIKDIFHERAWISTDVFHVKVKVVAETRGKEHVEEIETALKKIYDN
```



```
            VILH"
     gene   complement(6142. .8281)
            /gene="K01C8.2"
     CDS    complement(join(6142. .6174,6224. .6513,6586. .6808,
            6855. .7040,7092. .7368,8121. .8281))
            /gene="K01C8.2"
            /codon_start=1
            /db_xref="PID:g780189"
            /db_xref="SPTREMBL:Q21081"
            /translation="MYDKILFILSFYFSFICAQQPRFTCPNNANPLVSNGGTNLFCTS
            VGSNTDCPASSICVTAANGQGVLICCSQPSGVTPICPNNAVSQPSTIGYVECSQNNPT
            NCITGTQCVQSSNVASTFLCCSSSISTSSCPSDFTPAVGSNGGTVSCSPSVSTCPSGS
            SCMQSTLNSAFICCRSSSSQRICSNNQIALITNGALEMCTTPGTQCSSAGYTCQLSVL
            LATTVCCGQGSTGSTIGCADGRPVYQQIAGQTYTCEITSAITSCPSGYDCAPSDDPFV
            DVCCLTGSTPIPENLSCPTGWNSYKNEVDNAVRTCTAVLDTSCPIGYSCAPSNQVSQF
            LCCRLASSLICINGKTLLVNGAPKLCTPTTYSQCPYNYSCQQSVNPTVTVCCSNT"
     gene   9795. .11859
            /gene="K01C8.3"
     CDS    join(9795. .10109,10160. .10791,10837. .11173,11327. .11510,
            11558. .11859)
            /gene="K01C8.3"
            /note="similar to aromatic-L-amino-acid decarboxylase;
            cDNA EST EMBL:D68717 comes from this gene; cDNA EST
            EMBL:C12283 comes from this gene; cDNA EST yk279g7.3 comes
            from this gene; cDNA EST yk279g7.5 comes from this gene;
            cDNA EST yk296c11.3 comes from this gene; cDNA EST
            yk296c11.5 comes from this gene; cDNA EST yk303a5.5 comes
            from this gene; cDNA EST yk310g2.3 comes from this gene;
            cDNA EST yk310g2.5 comes from this gene; cDNA EST
            yk356d12.5 comes from this gene; cDNA EST yk374c1.3 comes
            from this gene; cDNA EST yk374c1.5 comes from this gene;
            cDNA EST yk410g3.3 comes from this gene; cDNA EST
            yk410g3.5 comes from this gene; cDNA EST yk424b3.3 comes
            from this gene; cDNA EST yk424b3.5 comes from this gene"
            /codon_start=1
            /db_xref="PID:g780197"
            /db_xref="SPTREMBL:Q21087"
            /translation="MTRDEPRQYGKETVDYIVDYLENIQKRRVVPAIEPGYLKDLIPS
            EAPNTPESPESVMEDFEKLIMPGITHWQHPRFHAYFPAGNSPPSIIADMLSDAIGCVG
            FSWAACPAMTELELIMLDWFGKMIGLPAEFLPLTENGKGGGVIQSSASECNFVTLLAA
            RFEVMKELRQRFPFVEEGLLLSKLIAYCSKEAHSSVEKACMIGMVKLRILETDSKFRL
            RGDTLRNAIQEDRNLGLIPFFVSTTLGTTSCCSFDVLSEIGPICKENELWLHVDAAYS
            GSAFICPEFRPLMNGIEYAMSFNTNPNKWLLINFDCSTMWVRDRFKLTQALVVDPLYL
            QHSWMGTEPALFQTIHFSVLDKSIDYRHWGIPLSRRFRSLKLWFVIRMYGIDGLQKYI
            REHVRLAKKMETLLRADAKFEIVNEVIMGLVCFRMKGDDELNQTLLTRLNASGRIHMV
            PASLGDRFVIRFCVCAETATDKDIEVAYEIIAQATQHVLHDSVKAVIAEEDEEAVALE
            EMVADLNITPTPEKCLTRQNSANAAESGQKLERQLSKEEILAQKQHESLAKKRYTPQP
            LDLALAPDSDPELSPCSPQILVS"
     gene   12146. .12381
            /gene="K01C8.4"
     CDS    join(12146. .12198,12249. .12381)
```



```
            /gene="K01C8.4"
            /note="cDNA EST yk303a5.3 comes from this gene; cDNA EST
            yk356d12.3 comes from this gene"
            /codon_start=1
            /db_xref="PID:g780190"
            /db_xref="SPTREMBL:Q21082"
            /translation="MANHRKMSQDLYRDRTLMQTISHSQRPNRLSQSPGSAGSAFFDD
            DDDRIVADVQTGLQTPI"
     gene   complement(12504. .14444)
            /gene="K01C8.5"
     CDS    complement(join(12504. .12643,12688. .13387,13435. .13564,
            13613. .13754,13802. .14444))
            /gene="K01C8.5"
            /note="weak similarity with thyroliberin (Swiss Prot
            accession number P01152); cDNA EST EMBL:T00107 comes from
            this gene; cDNA EST EMBL:T01656 comes from this gene; cDNA
            EST yk287g1.3 comes from this gene; cDNA EST yk352d1.3
            comes from this gene; cDNA EST yk420f9.3 comes from this
            gene"
            /codon_start=1
            /db_xref="PID:g1066993"
            /db_xref="SPTREMBL:Q21088"
            /translation="MFLNAKYSEAARPPIDEDRIRHTEDVIKRMLYKGRRKESMGYYP
            NRNHLMRDQIRQTVYNLGVPNGYRAPEQDHPEDDYDDQPGPSNRHYGVSSHPEYTRED
            RWRLAEDYEDVDKGDFSRDLNKEKYLKRRKQMIERESVDSEISEEEEEEPIYNKQTK
            RRRSQKTPIQHNKNRTESRLRQKEKSSSDEEVHEKSTSSRHSNVSRSSKSPYLNDDRA
            PSRGSKCTSSHHSTPKLLTPSPQYLKDSDDDDNWKNSIPFMSSSDGAGAGMRVRIGSA
            IASLPLRDDSKNSYSEVALMLLERSVKNLKNPQKMLIKEKPQTRIMNMSTSWAQQDPK
            KRNENDDDEIVPETEKAETSLMHDLSRIQHPQNATTVQMDQTLFTNSKLMKIDVAIAK
            NWLDTKMKKTLIPPPLFRSSPISEDVPKSSDRITMPSTIKGPESTAVGNRISSDGEEE
            DFATKQPDIDDYIARGASKIDPSKTITAPNTSFQEPSTSLFKDPTFFRQESKGNEMNI
            ESPSERTPSKNRTSQQMEFQISTSSISKTKTNASLDSIHRRKSKPQPMLEFDAETQKM
            FDDAFQSDKKSTKEKYPF"
     gene   15229. .16264
            /gene="K01C8.7"
     CDS    join(15229. .15340,15394. .15679,15729. .15945,15989. .16264)
            /gene="K01C8.7"
            /note="mitochondrial carrier protein"
            /codon_start=1
            /db_xref="PID:g780193"
            /db_xref="SPTREMBL:Q21084"
            /translation="MKIPDLTNYEHLIGGLCGGVTSTVVCHPFDLLKIRFSANEGSSL
            RPQYSSYADAVRKIVREVGVRGLYQGWTPSLIGASLSWGLYFQWYNSLRTKIYENFST
            GSKLANNLISGCISGSAIMCITNPIWLTKTRLCLQYENQQSKKYAGMMDCLKKTVKQE
            GFFGLYRGFVTGVIGTTHGAVQIAAYSWIIDKRCQSQGLPKDSFLSQTDYVIASSTSK
            VLATTIFFPYQVLRTRMQDHNTDSRGVWKTTLKTIHNEGIGGLWKGCLIANVRQLPAA
            VVTFLTYENVKRLVGMTKQI"
     gene   complement(16495. .17354)
            /gene="K01C8.6"
     CDS    complement(join(16495. .16826,16873. .17065,17117. .17222,
            17266. .17354))
```



```
            /gene="K01C8.6"
            /codon_start=1
            /db_xref="PID:g780192"
            /db_xref="SPTREMBL:Q21083"
            /translation="MLNQSLRPLVNVQVRSVSSKYPRPTPRHFERRLFEAALKPELPP
            TVNWCIPPGLIKEEHTGKSKYISDVELALSNLVRDWMVREEFRVNAVCQFLPVPGRTL
            WFAKNQLRSKNIEFRSYGNKILKKIFDKTPMSSLNTVLVGNNAILLAKDISAIKSILQ
            ETDKLNWIEPLVMMADGRIVDMVDARELAKLSSLEDLRAHTVQLLGQQVAQVTISLDT
            VTRHLPSLLDSHISTNQEAKK"
     gene   17526. .19875
...
Note: remainder of annotations omitted.

  Query Match          10.9%;  Score 19;  DB 18;  Length 24725;
  Best Local Similarity 100.0%;  Pred. No. 3.93e+00;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   6178 AAAATAAATTATTTTTAAC 6196
           ||||||||||||||||||
Cp     55 AAAATAAATTATTTTTAAC 37


RESULT   13
LOCUS       CELZK488     27885 bp    DNA             INV       10-SEP-1997
DEFINITION  Caenorhabditis elegans cosmid ZK488.
ACCESSION   AF022984
NID         g2384929
KEYWORDS    .
SOURCE      Caenorhabditis elegans strain=Bristol N2.
  ORGANISM  Caenorhabditis elegans
            Eukaryotae; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 27885)
  AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
  TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
  JOURNAL   Nature 368 (6466), 32-38 (1994)
  MEDLINE   94150718
REFERENCE   2  (bases 1 to 27885)
  AUTHORS   Wilson,R.
  TITLE     The sequence of C. elegans cosmid ZK488
```

```
  JOURNAL   Unpublished (1997)
REFERENCE   3  (bases 1 to 27885)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-SEP-1997) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     Submitted by:
                  Genome Sequencing Center
                  Department of Genetics, Washington University,
                  St. Louis, MO  63110, USA, and
                  Sanger Centre, Hinxton Hall
                  Cambridge CB10 1RQ, England
                  e-mail: rw@nematode.wustl.edu and jes@sanger.ac.uk


            NOTICE:  This sequence may not be the entire insert of this clone.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we provide a small overlap between
            neighboring submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded or sequenced with an alternate
            chemistry; an attempt was made to resolve all sequencing problems,
            such as compressions and repeats; all regions were covered by
            sequence from more than one subclone


                        NEIGHBORING COSMID INFORMATION:


            The 5' cosmid is K06H6;3' cosmid is F48G7.  Actual start of this
            cosmid is at base position 1 of CELZK488

            NOTES:

            Coding sequences below are predicted from computer analysis, using
            the program Genefinder(P. Green and L. Hillier, ms in preparation).
FEATURES            Location/Qualifiers
     source         1. .27885
                    /organism="Caenorhabditis elegans"
                    /strain="Bristol N2"
                    /db_xref="taxon:6239"
                    /chromosome="V"
                    /clone="ZK488"
     gene           1708. .3689
                    /gene="ZK488.6"
     CDS            join(1708. .1870,1922. .2075,2119. .2310,2362. .2848,
                    2896. .3005,3050. .3146,3224. .3402,3449. .3689)
                    /gene="ZK488.6"
                    /codon_start=1
                    /evidence=not_experimental
```



```
                    /db_xref="PID:g2384933"
                    /translation="MEREALLHTSSSPKNSELSSPTKRPSKSLLIIFALTNVAFLICF
                    SKIKFSTGGTEVTKIARPEVIVKVPTEISTTSTALSSTSQFISSTNAPAPRSPENKIF
                    PQNSTDCPYEQWNQVRTDFIPNMNFHKKLSKKWKSDFKYLYHTLPSAGAAFVHEDQIV
                    VTLTAENQVDDTVYCRYYDCRRREIMDPFKSTIFPRCTVPCPRRPGTKFITISKDLNE
                    TLEYGIPIVPRLESNIKTQISVLKPPKPISEPPHYPTVCMATLYGDEPKFLQIVDFIE
                    YHKLQGATFPHIYLRNVSDYDRVLLDNYVKTGDIELITLQDHFWRADYMWHNGQINDC
                    HHRSKHFSKWTALIDIDERLEMKSDKFKTVVDYLDSIQDDSISSLRFRVKWVMKHNYT
                    PARYENETQLQSEMIFHKYQNTSQLGRERDQPKCIIRPENVAMMSIHGPKDMYKGKKK
                    IVVPEDTGFIRHYRNVENKIFAHGYKKMMSHAPFNISPIDHDIDKDLSKKIMDRVKWT
                    YEIVQPTCEQKAKMYRVESTSPPCNNETKNLN"
     gene           4500. .6903
                    /gene="ZK488.5"
     CDS            join(4500. .4662,4716. .4863,4907. .5098,5147. .5331,
                    5381. .5631,5678. .5787,6260. .6356,6427. .6605,6663. .6903)
                    /gene="ZK488.5"
                    /codon_start=1
                    /evidence=not_experimental
                    /db_xref="PID:g2384932"
                    /translation="MEREALLHTSSFPKNSESSFPTKSISKSLLIIFALSNIAFFICF
                    SKIKFSIGGSEIVRPEVIVSVPTETSTTTSTSLSSTSQFISSTNAPALRSPENKIFPQ
                    NSTNCPYEEWNQVRTDFIPNTDLHKQWSKKWKSNFKYLYHTLPSVAAAFLHEDQIVVT
                    LTAENQADEIVYCRYYDCRRREIMDHPESTIFPRGTVYCARRPGAKFITVSKTLKKIL
                    EYSVPIVSRLGKPQHYFTVCMAPLYGDEPKFLQIVDFIEYHKLQGATFPHIYLRNVSD
                    YDRVLLDNYAKTGDIELITLQDHFWRADYMWHNGQINDCHRRNNYFSKWTALIDIDER
                    LEMKSDKFKIVADYLDSIQDDSIANLHFRVKWVMKHHDTPAKYENETQLKREMLFHKY
                    QNLSQLGAIWDQPKCIIRPENVAIMTIHGPREMYKGEKMTVVPENVGFIRHYRNVELR
                    IFHKAPERMMSHAPFNISPIDKDIDQDLSKKIMDRVKWIYEIVQPTCEQKAKMYRVES
                    TSPSCNNETKTLN"
     gene           11821. .13561
                    /gene="ZK488.4"
     CDS            join(11821. .11908,11954. .12019,12079. .12162,12210. .12439,
                    12852. .13078,13125. .13326,13379. .13453,13502. .13561)
                    /gene="ZK488.4"
                    /note="contains similarity to C4-type zinc fingers"
                    /codon_start=1
                    /evidence=not_experimental
                    /db_xref="PID:g2384931"
                    /translation="MACASFPRRTVSPGIRFLCKEGNRCPISQEIRFFCRSCRYDKCI
                    SVGMNRGQVQQKRDDNNVPKYVLDSRRLGIQEIVRGYTTSRTVPSEPAEPISPGSTGS
                    TGSTRSSSPETSPAEPSKKSEFSIILNTYPDDLLQYYVKQVEMSMIHKQLGLLKNTLL
                    VKSVDELVDISETLNDLSMEACMNCPGVDILRQRDVEVILKSFHFANVWMDSIWAYSR
                    HNESIVNKFPENNGPLSMFINQIKSTLGSSLSQLTFNIYEFSALKAFCVWKLSYHDST
                    ISINIMAEEQYIGICSALRKHYEATTNMTDVDIAMRIAEITLQIVPATTIYQDMIRFY
                    QQNGLIDNL"
     gene           complement(14891. .15431)
                    /gene="ZK488.7"
     CDS            complement(join(14891. .15139,15201. .15431))
                    /gene="ZK488.7"
                    /codon_start=1
                    /evidence=not_experimental
```



```
                    /db_xref="PID:g2384934"
                    /translation="MLEVQLASDFMKEQMNSILVFTALVAIVSAGSSWPMFLRDQPRN
                    VQQSYYTIASNSSLTQNQKDQQLQQWAQTNNLSNQYTDYTQQRDQQEQRMNQNTSQII
                    SQLSSVQNQLESILNNDMQTGEQQQQAIQSLTNQYPQEVPVLFYIRKHSMRQARNNN"
     gene           complement(15951. .17409)
                    /gene="ZK488.8"
     CDS            complement(join(15951. .16155,16249. .16371,16568. .16603,
                    16650. .16784,16833. .16976,17144. .17409))
                    /gene="ZK488.8"
                    /codon_start=1
                    /evidence=not_experimental
                    /db_xref="PID:g2384935"
                    /translation="MNSELYPDFIQKIQNVLSLIENDVFLVNFILSCVSLIVNIFHLS
                    IIPQKSMRISSINVLMIGMTICDIFSVLNTIYRYLELVDLEYPECVTSDSFLKAYMDI
                    TSWSLQDFLRRCSCWLGIFIATVRLIIIRNVSNNSNSDTATPRYGIILIFVFSCISGV
                    FQGMFQYAIKIVENHNNTLYSICAEHQDITKDPKIEKSRLLIQRNKEDNEKQFGLPTK
                    LITLLTITSFLADAPLGRLLTDLVVYFNILSTVVILLHPVMCLMMSSQYRITARKLFR
                    IKTITIALAENNIRVLPKRESPETGM"
     gene           complement(17643. .20170)
                    /gene="ZK488.9"
     CDS            complement(join(17643. .17790,18225. .18460,18505. .18628,
                    18866. .19009,19905. .20170))
                    /gene="ZK488.9"
                    /codon_start=1
                    /evidence=not_experimental
                    /db_xref="PID:g2384936"
                    /translation="MNSEIHFNFIQKFQNVLNLLEKYAFLVQFILAIISFILNLLHFV
                    IITRKSIRISSINCLMVGVTVCDICRMLTTIYRYLELEDLEYPECITVSSYIKAYLDI
                    TSWWLQDYFRRCSSWLDIFIANVRYIIMRKVSGASCAEHQDINKVYKTTINLRPIPTD
                    NKMLLIRTYIFLDVIFSHFLPSQAFPILTVLLLRKIQKMEKSRPVVRNNRVAENNEDK
                    HPLSTNLIIPLTISSFLADAPLGCIVMIKLFIPSGNRRIKKLNFFNYDIWSYLYHIRV
                    LGQFAALLHLQCQHKWPTTFLPTEMFQAD"
     gene           complement(20428. .20952)
                    /gene="ZK488.10"
     CDS            complement(join(20428. .20676,20731. .20952))
                    /gene="ZK488.10"
                    /codon_start=1
                    /evidence=not_experimental
                    /db_xref="PID:g2384937"
                    /translation="MQLSSDFLIENMNSILVLTALVAIVSAGSSWPMFLRDQPRNIQQ
                    SYYTIASNSSLTQNQKDQQLQQWAQTNNLSSQYSDYTQQRDQQEQQMSQNTSQIISQL
                    SSVQNQLESILNNDMQTGEQQQQAIQTLTNQYPQEVPVLFFIRKMSMRQARNNN"
     gene           21777. .23709
                    /gene="ZK488.2"
     CDS            join(21777. .21880,22098. .22644,22693. .22883,22932. .23202,
                    23596. .23709)
                    /gene="ZK488.2"
                    /note="contains similarity to C4-type zinc fingers"
                    /codon_start=1
                    /evidence=not_experimental
                    /db_xref="PID:g2384930"
```



```
                    /translation="MLVKSLQTCKICGAENTRGNHPGVQCCRACAVFFRRRAGTKYLR
                    LKCLSVHCGEAARFCKPCRLKRCYEAGMKIECNLFFSFSANNVLASDFQHNRNDSIKSS
                    SIAQIPRSFANIVGRPSLVIFCVPQDSYQKTFVDLNSLVGKASEIFTAGPESPYIGLT
                    QLKKLATPSNCSKEWAQTRFKTISHGEMSYFWEFYFLRTAKWLTYFDEFQRIPDEIKI
                    KLLSFWHVFARLDKLITTARARKLKLCSQQTWAMSNGLILDFDRTKVDFSDISNTFT
                    EDLFYFLNSITALDLPQVLESLMELFVSDMEFNFWLAQLTFSYAGKRPQGDILKICDR
                    FQEVLSNDLHEYYVKEMRTPRYSGRLGKIMKINNAIQNDIWKNRPRGELAAIFNVFKV
                    EFSHPEMFIDTGFVRN"
     gene           23871. .26134
                    /gene="ZK488.1"
     CDS            join(23871. .24100,24339. .24401,24737. .24827,25300. .25616,
                    25759. .26134)
                    /gene="ZK488.1"
                    /note="contains similarity to a C4-type zinc finger"
                    /codon_start=1
                    /evidence=not_experimental
                    /db_xref="PID:g2384938"
                    /translation="MAKLWRAHLFGGRCFRREILLNYPLRSFFFHLLVCYIDLVGTLV
                    MNELRPCEVCGAHGAHGNHFGATSCRACAAFFRRAANSKWSFQKCQTQNCSDLLLPSE
...
Note: remainder of annotations omitted.

  Query Match           10.9%;  Score 19;  DB 18;  Length 27885;
  Best Local Similarity 100.0%;  Pred. No. 3.93e+00;
  Matches    19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Db  23715 CAAAAATAAATTATTTTTA 23733
          ||||||||||||||||||
Cp     57 CAAAAATAAATTATTTTTA 39


RESULT   14
LOCUS       HS310H5     29718 bp    DNA             PRI       19-MAR-1997
DEFINITION  Human DNA sequence from cosmid 310H5 from a contig from the tip of
            the short arm of chromosome 16, spanning 2Mb of 16p13.3. Contains
            EST and CpG island.
ACCESSION   Z69705
NID         g1204102
KEYWORDS    16p13.3; CpG island.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 29718)
  AUTHORS   Lightning,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-1996) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1RQ, England. E-mail enquires: humquery@sanger.ac.uk
COMMENT     IMPORTANT:  This sequence is not the entire insert of clone 310H5.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we arrange for a small overlap between
```

```
            neighbouring submissions.
            The true left end of clone 310H5 is at 1 in this sequence. The true
            left end of clone 314G4 is at 20182.
            310H5 is from a 280kb clone contig extending from the telomere of
            16p.
            Higgs D.R., Flint J. unpublished. MRC Molecular Haematology Unit,
            Institute of Molecular Medicine, Oxford.
            310H5 came from the Los Alamos, flow sorted human Chromosome 16
            library.
FEATURES             Location/Qualifiers
     source          1. .29718
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /map="16p13.3"
                     /clone="310H5"
     repeat_region   1. .61
                     /partial
                     /note="Alu repeat: matches 62. .1 of consensus"
     repeat_region   186. .464
                     /note="Alu repeat: matches 1. .296 of consensus"
     repeat_region   472. .758
                     /partial
                     /note="Alu repeat: matches 308. .1 of consensus"
     repeat_region   779. .1069
                     /partial
                     /note="Alu repeat: matches 308. .1 of consensus"
     misc_feature    1012. .2069
                     /note="putative CpG island"
     repeat_region   2545. .2825
                     /note="Alu repeat: matches 1. .308 of consensus"
     repeat_region   2826. .3112
                     /note="Alu repeat: matches 1. .308 of consensus"
     repeat_region   3136. .3254
                     /partial
                     /note="Alu repeat: matches 124. .1 of consensus"
     repeat_region   3292. .3406
                     /partial
                     /note="Alu repeat: matches 116. .1 of consensus"
     repeat_region   3479. .3630
                     /note="MLT1F element fragment"
     repeat_region   3687. .3829
                     /note="MLT1F element fragment"
     repeat_region   3921. .4204
                     /partial
                     /note="Alu repeat: matches 308. .1 of consensus"
     repeat_region   4228. .4519
                     /partial
                     /note="Alu repeat: matches 308. .1 of consensus"
     repeat_region   4686. .4976
                     /note="Alu repeat: matches 1. .308 of consensus"
```



```
     repeat_region   5003. .5291
                     /note="Alu repeat: matches 1. .308 of consensus"
     repeat_region   5299. .5517
                     /partial
                     /note="Alu repeat: matches 84. .308 of consensus"
     repeat_region   5629. .5650
                     /note="11 copies of 2 mer 100 % conserved"
     repeat_region   5877. .6014
                     /partial
                     /note="Alu repeat: matches 148. .1 of consensus"
     repeat_region   6249. .6527
                     /note="Alu repeat: matches 1. .308 of consensus"
     repeat_region   6552. .6843
                     /note="Alu repeat: matches 1. .308 of consensus"
     repeat_region   7729. .8019
                     /partial
                     /note="Alu repeat: matches 308. .1 of consensus"
     repeat_region   8166. .8233
                     /partial
                     /note="Alu repeat: matches 308. .240 of consensus"
     repeat_region   8252. .8447
                     /partial
                     /note="Alu repeat: matches 202. .1 of consensus"
     repeat_region   8519. .8817
                     /note="Alu repeat: matches 1. .308 of consensus"
     repeat_region   8855. .9040
                     /partial
                     /note="Alu repeat: matches 1. .202 of consensus"
     repeat_region   9041. .9134
                     /partial
                     /note="Alu repeat: matches 214. .308 of consensus"
     repeat_region   9441. .9727
                     /partial
                     /note="Alu repeat: matches 308. .1 of consensus"
     repeat_region   9839. .10129
                     /partial
                     /note="Alu repeat: matches 308. .1 of consensus"
     repeat_region   10206. .10496
                     /partial
                     /note="Alu repeat: matches 308. .1 of consensus"
     misc_feature    complement(10765. .10912)
                     /note="match: EST T12196 clone A584"
     repeat_region   11195. .11485
                     /partial
                     /note="Alu repeat: matches 308. .1 of consensus"
     repeat_region   11914. .12209
                     /partial
                     /note="Alu repeat: matches 308. .1 of consensus"
     repeat_region   12317. .12564
                     /partial
                     /note="Alu repeat: matches 35. .300 of consensus"
```



```
     repeat_region   12795. .12981
                     /note="MER30 element fragment"
     repeat_region   13143. .13441
                     /note="Alu repeat: matches 1. .308 of consensus"
     repeat_region   13507. .13631
                     /partial
                     /note="Alu repeat: matches 308. .181 of consensus"
     repeat_region   13648. .13936
                     /partial
                     /note="Alu repeat: matches 306. .1 of consensus"
     repeat_region   14002. .14290
                     /partial
                     /note="Alu repeat: matches 308. .1 of consensus"
     repeat_region   14356. .14646
                     /partial
                     /note="Alu repeat: matches 308. .1 of consensus"
     repeat_region   14659. .14950
                     /partial
                     /note="Alu repeat: matches 308. .1 of consensus"
     repeat_region   15850. .16138
                     /partial
                     /note="Alu repeat: matches 305. .1 of consensus"
     repeat_region   17172. .17287
                     /partial
                     /note="Alu repeat: matches 308. .192 of consensus"
     repeat_region   17288. .17448
                     /partial
                     /note="Alu repeat: matches 177. .1 of consensus"
     repeat_region   17846. .18022
                     /note="L1 element fragment"
     repeat_region   18138. .18167
                     /note="L1 element fragment"
     repeat_region   18183. .18473
                     /partial
                     /note="Alu repeat: matches 308. .1 of consensus"
     repeat_region   18474. .18543
                     /note="L1 element fragment"
     repeat_region   18601. .18894
                     /partial
                     /note="Alu repeat: matches 308. .1 of consensus"
     repeat_region   18899. .18938
                     /partial
                     /note="Alu repeat: matches 148. .105 of consensus"
     repeat_region   18940. .19230
                     /note="Alu repeat: matches 1. .308 of consensus"
     repeat_region   19240. .19303
                     /partial
                     /note="Alu repeat: matches 298. .231 of consensus"
     repeat_region   19311. .19379
                     /partial
                     /note="Alu repeat: matches 70. .1 of consensus"
```



```
     repeat_region   19410. .19699
                     /partial
                     /note="Alu repeat: matches 308. .1 of consensus"
     repeat_region   19704. .19917
                     /note="L1 element fragment"
     repeat_region   20701. .20990
                     /partial
                     /note="Alu repeat: matches 308. .1 of consensus"
     repeat_region   22081. .22371
                     /partial
                     /note="Alu repeat: matches 308. .1 of consensus"
     repeat_region   22372. .22507
                     /partial
                     /note="Alu repeat: matches 152. .1 of consensus"
     repeat_region   22614. .22800
                     /partial
                     /note="Alu repeat: matches 105. .308 of consensus"
     repeat_region   22889. .23181
                     /partial
                     /note="Alu repeat: matches 308. .1 of consensus"
     repeat_region   23361. .23549
                     /note="L1 element fragment"
     repeat_region   23796. .23967
                     /partial
                     /note="Alu repeat: matches 308. .138 of consensus"
     repeat_region   23968. .24259
                     /partial
                     /note="Alu repeat: matches 308. .1 of consensus"
...
Note: remainder of annotations omitted.

  Query Match           10.5%;  Score 19;  DB 26;  Length 29718;
  Best Local Similarity 100.0%;  Pred. No. 3.93e+00;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  23764 TTAAAAATAATTTATTTTT 23782
          ||||||||||||||||||
Qy     38 TTAAAAATAATTTATTTTT 56


RESULT   15
LOCUS       CELC18F3    32793 bp    DNA             INV       02-APR-1996
DEFINITION  Caenorhabditis elegans cosmid C18F3.
ACCESSION   U50067
NID         g1208851
KEYWORDS    .
SOURCE      Caenorhabditis elegans strain=Bristol N2.
  ORGANISM  Caenorhabditis elegans
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Nematoda;
            Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea;
            Rhabditidae; Caenorhabditis.
```

```
REFERENCE   1  (bases 1 to 32793)
  AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kiraten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
  TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
  JOURNAL   Nature 368 (6466), 32-38 (1994)
  MEDLINE   94150718
REFERENCE   2  (bases 1 to 32793)
  AUTHORS   Du,Z and Gattung,S.
  TITLE     The sequence of C. elegans cosmid C18F3
  JOURNAL   Unpublished (1996)
REFERENCE   3  (bases 1 to 32793)
  AUTHORS   Wateraton,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-FEB-1996) Robert Waterston
COMMENT                 Submitted by:
                        Genome Sequencing Center
                        Department of Genetics, Washington University,
                        St. Louis, MO  63110, USA, and
                        Sanger Centre, Hinxton Hall
                        Cambridge CB10 IRQ, England
                        e-mail: rw@nematode.wustl.edu and jes@sanger.ac.uk
                        NEIGHBORING COSMID INFORMATION:
            The 5' end lies in a gap followed by the cosmid F53A12;3' cosmid is
            C26B2, 200 bp overlap.  Actual start of this cosmid is at base
            position 1 of CELC18F3; actual end is at 6916 of CELC26B2

            NOTES:
            Coding sequences below are predicted from computer analysis, using
            the program Genefinder(P. Green and L. Hillier, ms in preparation).
FEATURES             Location/Qualifiers
     source          1. .32793
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /map="IV"
     gene            complement(4411. .12474)
                     /gene="C18F3.2"
     CDS             complement(join(4411. .4484,4538. .4653,4839. .5160,
                     5569. .5739,8028. .8360,8600. .9404,9948. .10216,10431. .10721,
                     10969. .11253,12021. .12311,12468. .12474))
                     /gene="C18F3.2"
                     /note="coded for by C. elegans cDNA yk16f3.3; coded for by
```



```
                     C. elegans cDNA CEESX22F; coded for by C. elegans cDNA
                     yk5b1.3; coded for by C. elegans cDNA yk2b7.3; coded for
                     by C. elegans cDNA yk24c11.3; coded for by C. elegans cDNA
                     yk44e7.3; coded for by C. elegans cDNA yk35b3.3; coded for
                     by C. elegans cDNA yk35b3.5; coded for by C. elegans cDNA
                     yk16f3.5; coded for by C. elegans cDNA yk24c11.5; coded
                     for by C. elegans cDNA yk44e7.5; coded for by C. elegans
                     cDNA yk5b1.5;  similar to immunoglobulin superfamily,
                     contains C2-like and fibronectin type III-like domains"
                     /codon_start=1
                     /db_xref="PID:g1208852"
                     /translation="MGPAPYWPLGPPPNTNTSEGEKVIFDCTTYGKPTPKVTFYKNGV
                     EIESQKDDRYLIEGTRLTIFDVKKGTYGKGDNAVYQCKSENKHGWLWTNFYLNLLAFK
                     PQLLIDPGKEEVEAVAGKKVTLECKFFASPNAAVKWEAPMISGSKGNQIPADAYGVGK
                     LVFSEVTAAEDGFYECIGTNKVGQARGLITLKVRKPTIVQPFPRVEEVEMAGEEMRLA
                     CDATADNQLEVKFEWLVDGKSLPEDRISSGHYKIDDDHSLVISNPTQDDTAKYKCVVS
                     TKLDQVEKEIKIQFKDVPVAVHSAWVQTCDKNSLTAFIKFEHTYESIHTIAPIKEPWVQ
                     YQIDSETEGSQWRTHPVPSAAHPNDKIDNNLRHTTGDATVSLQPPGKTVFRVIARNSV
                     GDSAARLAKDQCETPAKQPDKNPDEVAAKGTSPENIIVQWKPMSREEWNGADFHYVVK
                     YRPKDEDQRVGDWKEVAVEDPPADRVTVNLDDEKDVKPFQPYEVQVQAVNSEGRTNVV
                     PETVEGRTGEGVPPSSIPSGLRVLEKSGTTVTLAWNGVDPQTANGNPTGYKITTWVDEA
                     DQDSDDSSEDDEDDEKRKFRWKRSIRVKRQSGIRKTVVFGPSATQGTLTDLKPATLNH
                     AYIQVTNGAHEGSASDTIDPQTDEGVPSPVRSLRAYPMNSKVGGEKGVVVLVWKKPRQ
                     TNGKLARYEVEYCKTQNGKLVEKSCPRKQIDADSKEIRITGLENETPYRFILRAMTSA
                     GEGDPNSSDATTLPETTAAGVDPGVPSLVENAIGDKYFNVSFRPAKYDDESRAPVGNT
                     YEVQYKPQDADEWETVTKPADDGLTVHVDGLSPGTKYDVRVAALQVDPEGGETTKTLSG
                     ISKITTTGTSSRERNVYLLYLLLLLLILLLLLIICCICCVVCRQRGQNYPVSQRERDQGR
                     EPILGKPDYKTDDDEKRSLTGSKAESETDSMAQYGDTDPGVFTEDGSPIGQYVPQKSL
                     MPAERPEKGSTSTFV"
     gene            complement(22699. .24704)
                     /gene="C18F3.3"
     CDS             complement(join(22699. .23144,23483. .23559,23896. .24073,
                     24394. .24505,24603. .24704))
                     /gene="C18F3.3"
                     /note="contains similarity to IG-like C2-type domain"
                     /codon_start=1
                     /evidence=not_experimental
                     /db_xref="PID:g1208853"
                     /translation="MNQILYENSPICTSSSTIIFLKIEGRFYFFKLIHDRKRVSNEVR
                     RDSRSRRKTTIMRSFIFLLLLTGHISSGSIDWSAIETPPQFVHDKNADTVYFKVEKSG
                     MSDESAKNTPGNLLEQTIHCYANGNPRPSYRWKKDGKTFLPSMPPEKVVQKPGEGSLV
                     FSRLDESDAGLYQCEAENSNGTAVDRPVRVQETWIRHFKTAEPEVVVVEVGDPIQRNC
                     SPPASNPYARVTWILMGKEPGHFETISSSHISSNEQVSDNSPFPHSFYRSFLVTSLCQ
                     ENFSHSLPVLCKPKSNNFHRIERFLEKD"
BASE COUNT    11102 a    5674 c    5931 g   10086 t
ORIGIN

  Query Match           10.9%;  Score 19;  DB 18;  Length 32793;
  Best Local Similarity 100.0%;  Pred. No. 3.93e+00;
  Matches    19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```



```
Db  21444 CAAAAATAAATTATTTTTA 21462
          ||||||||||||||||||
Cp     57 CAAAAATAAATTATTTTTA 39


RESULT   16
LOCUS       CER09A8     37145 bp    DNA             INV       21-SEP-1998
DEFINITION  Caenorhabditis elegans cosmid R09A8, complete sequence.
ACCESSION   Z68009
NID         g1070081
KEYWORDS    HTG; Canal protein like; Cuticle collagen.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 37145)
  AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
  TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
  JOURNAL   Nature 368 (6466), 32-38 (1994)
  MEDLINE   94150718
REFERENCE   2  (bases 1 to 37145)
  AUTHORS   Wilkinson,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-NOV-1996) Nematode Sequencing Project, Sanger Centre,
            Hinxton, Cambridge CB10 1RQ, England and Department of Genetics,
            Washington University, St. Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     Coding sequences below are predicted from computer analysis, using
            predictions from Genefinder (P. Green, U. Washington), and other
            available information.
            For a graphical representation of this sequence and its analysis
            see:-
            http://weface.sanger.ac.uk/cgi-
            bin/display?db=wormace&class=Sequence &object=R09A8
            Current sequence finishing criteria for the C. elegans genome
            sequencing consortium are that all bases are either sequenced
            unambiguously on both strands, or on a single strand with both a
            dye primer and dye terminator reaction, from distinct subclones.
            Exceptions are indicated by an explicit note.
            IMPORTANT:  This sequence is NOT necessarily the entire insert of
```



```
            the specified clone.  It may be shorter because we only sequence
            overlapping sections once, or longer because we arrange for a small
            overlap between neighbouring submissions.
            This sequence is the entire insert of clone R09A8. The true right
            end of clone T24D5 is at 957 in this sequence. The start of this
            sequence (1. .106) overlaps with the end of sequence Z68012.
            The end of this sequence (37042. .37145) overlaps with the start of
            sequence Z95309.
FEATURES             Location/Qualifiers
     source          1. .37145
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /chromosome="X"
                     /clone="R09A8"
     gene            8893. .14046
                     /gene="R09A8.1"
     CDS             join(8893. .9035,10097. .10518,10568. .10857,10902. .11087,
                     11136. .11300,11346. .11585,11632. .11814,11861. .12086,
                     12231. .12452,12509. .12742,12875. .13066,13211. .13905,
                     13957. .14046)
                     /gene="R09A8.1"
                     /codon_start=1
                     /db_xref="PID:g1070082"
                     /db_xref="SPTREMBL:Q21852"
                     /translation="MDQNGGRSMKGSVRPGWMGGLVDGLGQSEGIWNETTTKKDGRIR
                     KKNRQNTYEPVASKKSEAFSALASSCVGKKTNPPCEIGPVVSVYGRVVVEIDDKIKLR
                     RLMNHIPDHYSIPILRIVQFGEYNAGKKPRSLMSYRKRMTDCDYFSDALAIELLDFFE
                     MIHNDVERANYVSRILDTKAEYYGFKSYDDRVSDIRFNKTRPAPLNTLHVDISSHPNS
                     TTERFSNGSRNSSRNSAVLTFSTNVQSHAGLVRNDIDSNSYAGARSIEDSDFSKVDRK
                     METISTETQNGFRGNSALRGSSSLSDNNVQSSLMVGGNRSTDDRVNAQRFASTGFVEK
                     ECRRWDQLVEKKQKKEVNSDHKKANRITSHLEHNSRLESLPPQVDTRGLDNSHLAEKK
                     YIEGISEEVSNKQPNNLGSETSSAIENERNWMPTEHNIASRIAFPGLWLNVKSNDKYL
                     PPTISRTECWPYITDRKEKNRRHMEEMMKIAEQNEQLYQFPRPNQSCYSSTSSHRKDS
                     MTTDSKEITSSDMEYVYAIDEGKCFRLKRVNEWKDEKDSEHAVVSRRPSFRIYCEP
                     AINVIYFSSTQNASEEDPGNTVSSDVSQYSARTSNNASSKRKLSESSSDSDTGILLKK
                     ANVSRFVEEGVPVPWYQPQSSYAVAFYSSSSHMFFHILKSKRKFYILLHLYISINLV
                     KSFLVFNSQNSAPPIKPMKKEAGALLPPRAVVYKLPAIKGSMKDAISAHKLLKKSYAK
                     PIVDISSTSGDSDPEWWSDRELTADLNFKRELLSKKPRFSRKTFDERSSYIARFKESH
                     LGNYLNSDSDTVSNEKPKTLAQLKNFSKSNLVEYHPKRSRKKIRARVYRTKSAVTTA
                     CTSTSKQAARSDEPFFSQRDTRQHSYTQRVQVPIGAALHSSSSIAANVNDYRSKAPTR
                     MIHSSKVSTVNAAASKFAAPRCQAPNVSHMKNGQPSVIDRLLPSDSNTKPLLAPASNN
                     SWITANSRATGILKGRSGPVDQSQNQNTRRENASMSGRQPLDTSRKHMFLFCNKVCVI
                     TDEEIRRDKEQHRLERSQISAERKLEHRKAMRRQSRARLLSRLQTEYDNNDATTLFKF
                     RSSCLHS"
     gene            15638. .19470
                     /gene="R09A8.2"
     CDS             join(15638. .16226,16276. .16658,16710. .16868,16966. .17139,
                     17185. .17364,17426. .17699,17804. .17925,18125. .18433,
                     18682. .18925,19328. .19470)
                     /gene="R09A8.2"
                     /note="cDNA EST yk377g5.5 comes from this gene"
```

```
                     /codon_start=1
                     /db_xref="PID:g1070083"
                     /db_xref="SPTREMBL:Q21853"
                     /translation="MEDPKLGSRRMTLRSASRLNIERNGSDSSSDTHSKMRSNGYSSM
                     KLCDNESTFYFKLKSEDSIVSNVPDVCSKPVAGNGKETTSPCEIGRYSSFYARFEID
                     MDDKAELKKLMNSIPENYSIPILRIVQPGEYNDGKEPRSLLSYREKMADVDKFSDDLV
                     IELLDFFVMIRSDVEMANNKTEKLETNAEYIGFKSTDDSLYDIQPNTKRPAPLYRPEE
                     VIAFDPNWGTPNYPHVIEEDKTKKLASALQDYRTLRDATPDKTFYFGEIPTSPINVGT
                     HDDQLCNGMNSISGNTLNTSYEETRQLKNSDHVKQIDPQMKGNSVDAHKGLQFSMTVQ
                     NSSCLSSDGNTSTDDRVNAQCFASSGFVKSERRRLEQLMEQKLKEKNQCDYAKKTRNS
                     SALDHESSPGSCPLSDIHPEQSHFDISLLAENKKVEEVSQETAEKQEDKQISVTSLAR
                     MIERSSMPANPYSSADTQMEVHGFDSSVFVESEKQNSEQRCEQIKIGEKDREDSNRIV
                     SPEISLDVASDEKISSSTTTPNTLAEKCQATLRRAAFLYPVYPKTNSSACMNSHSSVD
                     VGYNEQERRQFEQALDEKKANVETNDDRSIQFSRVDTGLLKDSNNSTMNGREIEALGI
                     KAQNNQEEHQCTPTEYKSLLDKYTVETGPLIPDLEMLFGKRSTSTEATVSTSSRPLVP
                     DLGMLLGKRFKSIESTGSTSSCTDVQNIDVIAFETKAQLRESGQMKKHESSRDDAN
                     ENASFVHSLDVESNVDSCSSTGTQETSEKKRQEWVLAVVPYHPKNISTLVEDQSSNTL
                     KRMLVDTSSDLGKDQPPKKKRVPSSISFSNAEKTRMVLTTTSNIGSGSSSGEEILEKE
                     LI"
     gene            complement(21407. .25209)
                     /gene="R09A8.3"
     CDS             complement(join(21407. .21525,21574. .21674,21838. .22753,
                     22868. .23195,23280. .23423,23618. .23745,23792. .23916,
                     24244. .24439,24489. .24654,24818. .24955,25117. .25209))
                     /gene="R09A8.3"
                     /note="Similarity to Drosophila ring canal protein
                     (SW:KELC_DROME); cDNA EST EMBL:D27662 comes from this
                     gene; cDNA EST EMBL:D35096 comes from this gene; cDNA EST
                     EMBL:D34029 comes from this gene; cDNA EST EMBL:D36416
                     comes from this gene; cDNA EST EMBL:D37021 comes from this
                     gene; cDNA EST EMBL:D68294 comes from this gene; cDNA EST
                     EMBL:D67338 comes from this gene; cDNA EST EMBL:D67971
                     comes from this gene; cDNA EST EMBL:C11458 comes from this
                     gene; cDNA EST EMBL:C13577 comes from this gene; cDNA EST
                     EMBL:C11170 comes from this gene; cDNA EST EMBL:C13186
                     comes from this gene; cDNA EST EMBL:C11509 comes from this
                     gene; cDNA EST EMBL:C13736 comes from this gene; cDNA EST
                     yk411h12.3 comes from this gene; cDNA EST yk411h12.5 comes
                     from this gene; cDNA EST yk449a10.3 comes from this gene;
                     cDNA EST yk449a10.5 comes from this gene; cDNA EST
                     yk353g4.3 comes from this gene; cDNA EST yk353g4.5 comes
                     from this gene; cDNA EST yk321a5.3 comes from this gene;
                     cDNA EST yk321a5.5 comes from this gene; cDNA EST
                     yk286e2.3 comes from this gene; cDNA EST yk286e2.5 comes
                     from this gene; cDNA EST yk286a9.3 comes from this gene;
                     cDNA EST yk286a9.5 comes from this gene; cDNA EST
                     yk285d4.3 comes from this gene; cDNA EST yk285d4.5 comes
                     from this gene; cDNA EST yk271f12.3 comes from this gene;
                     cDNA EST yk271f12.5 comes from this gene; cDNA EST
                     yk244a1.3 comes from this gene; cDNA EST yk244a1.5 comes
                     from this gene; cDNA EST yk253c2.5 comes from this gene;
```



```
                     cDNA EST yk305b1.5 comes from this gene; cDNA EST
                     yk353g5.5 comes from this gene; cDNA EST yk524b11.5 comes
                     from this gene"
                     /codon_start=1
                     /db_xref="PID:g1070084"
                     /db_xref="SPTREMBL:Q21854"
                     /translation="MPLELNIDLTEERPKVPEIVAFAMDKPVLNKTNVLTLIPKRNRS
                     AVSMRADETDNWGQEALDEKEAYTLMLRDPSIKNDLLDRLNTFRRNRELCDAVLFVKE
                     REIFVHKVVLAAVSPALLDMFNEENESSQDEKSPTSSSPTNGHTNGITNGNGEVAHEN
                     GAPVSAQKPQKSSTTFYEFSQTDFECPEALVNFAITANLEISSRKVAELYKTAYALQM
                     IPVIRACAAYLAENLQLSNCIGIRRQANFHNDTFLMGKVDQFIVDNFDSIVNDSKEPT
                     QLPVIQVRIIVPAEDGKVTNNANNNGGLAEMALFYFQNNPHDRAEHSIDLLTCKTHIL
                     YMEENHLADCIEMDERSSVGSCDIIQDYKKSGKDRKDVAKAMTAQEPIINSVVQHRVN
                     GAVPVRLNASRLPNMNASNESLESAGTDDSDPQDTIEARLISTHQTAPQYWVALVVLY
                     RRLCVLSLQLTDNEELLRTKREPCPVDSQKAALLSRLISCTGKQQKPLECMSAPRCSI
                     GASFLNGKIFVCGGYDRGECLRSVEEYDVVNGKWRNVAKMKAERGRFDCTIHGGKVFA
                     VAGSNGNNDLKSCEIYDPKADVWAPLPNLATARCHNGCATIDNYVYCIGGSFDQKVLK
                     DCERLDTTKLGTEEAEWEPMASMEHPRYQAGVCTWKGLVVAAGGCDRWTCMDSVEAFD
                     PKTNAWRQLPKLRQARRGGAVAVVRDTLYVIGGHDGTQSLDTVEILDSPSSQWRVGPT
                     LTTPRANTHAVVTAGNVIFCIGGFNGVKFLDTIELLENEQIGWRNWQHCPQQILEEHE
                     EEDSSLMDDEATSTAPSSP"
     gene            complement(32319. .32777)
                     /gene="R09A8.4"
     CDS             complement(32319. .32777)
                     /gene="R09A8.4"
                     /note="similar to cuticle collagen"
                     /codon_start=1
                     /db_xref="PID:g1070085"
                     /db_xref="SPTREMBL:Q21855"
                     /translation="MPRLPTRTTGNNYIQGYPGQPGPQGLPGNQGDQGEIGREGSPGA
                     QGAPGDNGMMGEAGKEGEKGAPGREGVRGEKGPSGPPGLPGLVGPKGPSGNSGPPGSP
                     GEPGPEGEQGPAGKDGAPGYDGGFGNYGEAGQPGEDAGYCKCPTRKQRKH"
     gene            complement(34854. .36653)
                     /gene="R09A8.5"
     CDS             complement(join(34854. .34985,35320. .35458,35507. .35693,
                     36435. .36534,36585. .36653))
                     /gene="R09A8.5"
                     /codon_start=1
                     /db_xref="PID:g1070086"
                     /db_xref="SPTREMBL:Q21856"
                     /translation="MIYFFGIDEWLTSSGKTCNVSMKKSILCSVMFLLLFAAGIAQFF
                     FIDHFKLVVPSVGHAILFLFEMICFLVLINALFNERPLLFLPFVIAEVVRCVCLVGVI
                     FLYIIKMFQVQGHEDERPNIVLPALHYDNEHKTVKQMVKLDLHLIFIFFFHLILAAIG
                     YRCTQIYSRGTRRNRMNPADYRPTQPMHTVIVLEQPIEQAMYGKKPVN"
BASE COUNT    11689 a    6814 c    7314 g   11328 t
ORIGIN

  Query Match           10.9%;  Score 19;  DB 18;  Length 37145;
  Best Local Similarity 100.0%;  Pred. No. 3.93e+00;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```



```
  Db  33484 TTAAAAATAATTTATTTTT 33502
            ||||||||||||||||||
  Qy     38 TTAAAAATAATTTATTTTT 56


RESULT   17
LOCUS       AC004754    39188 bp    DNA              PRI       28-MAY-1998
DEFINITION  Homo sapiens chromosome 16, cosmid clone RT286 (LANL), complete
            sequence.
ACCESSION   AC004754
NID         g3165400
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 39188)
  AUTHORS   Ricke,D.O., Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E.,
            Robinson,D., Jones,M., Buckingham,J., Chasteen,L., Thompson,S.,
            Goodwin,L., Bryant,J., Tesmer,J., Meincke,L., Longmire,J.,
            White,S., Ueng,S., Tatum,O., Campbell,C., Pawcett,J., Maltbie,M.,
            Misra,M. and Deaven,L.
  TITLE     Sequencing of Human Chromosome 16p13.3
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 39188)
  AUTHORS   Ricke,D.O.
  TITLE     Large Scale Sequence Analysis and Annotation with the Sequence
            Comparison Analysis (SCAN) System
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 39188)
  AUTHORS   Ricke,D.O., Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E.,
            Robinson,D., Jones,M., Buckingham,J., Chasteen,L., Thompson,S.,
            Goodwin,L., Bryant,J., Tesmer,J., Meincke,L., Longmire,J.,
            White,S., Ueng,S., Tatum,O., Campbell,C., Pawcett,J., Maltbie,M.,
            Misra,M. and Deaven,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAY-1998) Center for Human Genome Studies, DOE Joint
            Genome Institute, Los Alamos National Laboratory, MS M888, Los
            Alamos, NM 87545, USA
FEATURES             Location/Qualifiers
     source          1. .39188
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RT286"
                     /chromosome="16"
                     /map="16p13.3"
     repeat_region   complement(1. .78)
                     /rpt_family="Alu"
     repeat_region   166. .459
                     /rpt_family="Alu"
     repeat_region   complement(452. .748)
```



```
                     /rpt_family="Alu"
     repeat_region   754. .783
                     /note="(T)30"
                     /rpt_type=tandem
                     /rpt_unit=T
     repeat_region   complement(766. .1078)
                     /rpt_family="Alu"
     repeat_region   2546. .2723
                     /rpt_family="Alu"
     repeat_region   complement(2720. .2930)
                     /rpt_family="Alu"
     repeat_region   3016. .3293
                     /rpt_family="Alu"
     repeat_region   3300. .3611
                     /rpt_family="Alu"
     repeat_region   complement(3632. .3751)
                     /rpt_family="Alu"
     repeat_region   complement(3743. .3880)
                     /rpt_family="Alu"
     repeat_region   complement(3929. .4293)
                     /rpt_family="MLT1f"
     misc_feature    complement(4269. .4394)
                     /note="66% protein identity GenPept:L24521
                     transformation-related protein"
     repeat_region   complement(4348. .4678)
                     /rpt_family="Alu"
     repeat_region   complement(4686. .4993)
                     /rpt_family="Alu"
     repeat_region   5160. .5461
                     /rpt_family="Alu"
     repeat_region   5441. .5461
                     /note="(A)21"
                     /rpt_type=tandem
                     /rpt_unit=A
     repeat_region   5478. .5774
                     /rpt_family="Alu"
     repeat_region   5772. .6311
                     /rpt_family="Alu"
     repeat_region   complement(7146. .7475)
                     /rpt_family="Alu"
     repeat_region   complement(7624. .7880)
                     /rpt_family="Alu"
     repeat_region   7934. .8275
                     /rpt_family="Alu"
     repeat_region   8311. .8611
                     /rpt_family="Alu"
     repeat_region   complement(8894. .9205)
                     /rpt_family="Alu"
     repeat_region   complement(9281. .9588)
                     /rpt_family="Alu"
     repeat_region   complement(9647. .9970)
```

```
                     /rpt_family="Alu"
     repeat_region   complement(10648. .10968)
                     /rpt_family="Alu"
     repeat_region   complement(11347. .11666)
                     /rpt_family="Alu"
     repeat_region   11358. .11380
                     /note="(T)23"
                     /rpt_type=tandem
                     /rpt_unit=T
     repeat_region   11770. .12033
                     /rpt_family="Alu"
     repeat_region   12252. .12443
                     /rpt_family="MER30"
     repeat_region   12601. .12915
                     /rpt_family="Alu"
     repeat_region   complement(12959. .13077)
                     /rpt_family="Alu"
     repeat_region   13093. .13114
                     /note="(T)22"
                     /rpt_type=tandem
                     /rpt_unit=T
     repeat_region   complement(13097. .13409)
                     /rpt_family="Alu"
     repeat_region   complement(13428. .13762)
                     /rpt_family="Alu"
     repeat_region   complement(13814. .14131)
                     /rpt_family="Alu"
     repeat_region   complement(14104. .14434)
                     /rpt_family="Alu"
     repeat_region   complement(15291. .15611)
                     /rpt_family="Alu"
     repeat_region   complement(16594. .16920)
                     /rpt_family="Alu"
     repeat_region   complement(17628. .17974)
                     /rpt_family="Alu"
     repeat_region   complement(17932. .18036)
                     /rpt_family="L1ME3A"
     repeat_region   complement(18045. .18381)
                     /rpt_family="Alu"
     repeat_region   18355. .18381
                     /note="(T)27"
                     /rpt_type=tandem
                     /rpt_unit=T
     repeat_region   18362. .18702
                     /rpt_family="Alu"
     repeat_region   complement(18705. .19159)
                     /rpt_family="Alu"
     repeat_region   complement(19163. .19374)
                     /rpt_family="L1"
     repeat_region   complement(20142. .20459)
                     /rpt_family="Alu"
```



```
     repeat_region   complement(20448. .20590)
                     /rpt_family="MIR"
     repeat_region   complement(21537. .21970)
                     /rpt_family="Alu"
     repeat_region   22071. .22260
                     /rpt_family="Alu"
     repeat_region   complement(22342. .22638)
                     /rpt_family="Alu"
     repeat_region   complement(22815. .22987)
                     /rpt_family="L1MB7"
     repeat_region   complement(23222. .23853)
                     /rpt_family="Alu"
     repeat_region   23912. .23937
                     /note="(T)26"
                     /rpt_type=tandem
                     /rpt_unit=T
     repeat_region   complement(23919. .24223)
                     /rpt_family="Alu"
     repeat_region   24380. .24502
                     /rpt_family="Alu"
     repeat_region   complement(24503. .24849)
                     /rpt_family="Alu"
     repeat_region   complement(24836. .25448)
                     /rpt_family="SVA"
     repeat_region   complement(24953. .25277)
                     /rpt_family="LTR5"
     repeat_region   complement(25455. .25777)
                     /rpt_family="Alu"
     prim_transcript join(26155. .26275,26643. .26764)
                     /standard_name="no74d10"
                     /note="99% identity dbEST:AA605082"
     misc_feature    26158. .26274
                     /note="GRAIL 2 excellent exon, frame 2"
     misc_feature    26671. .26835
                     /note="GRAIL 2 excellent exon, frame 2"
     misc_feature    29099. .29228
                     /note="GRAIL 2 excellent exon, frame 0"
     repeat_region   complement(29465. .29791)
                     /rpt_family="Alu"
     prim_transcript complement(join(29935. .30077,30373. .30454,30688. .30824,
                     32581. .32766))
                     /standard_name="yx60f07"
                     /note="97% identity dbEST:N30983, N21584"
     misc_feature    29937. .30077
                     /note="GRAIL 2 excellent exon, frame 0"
     prim_transcript join(30373. .30432,30688. .30816)
                     /standard_name="vy94f06"
                     /note="82% identity mouse AA914397"
     misc_feature    31289. .31391
                     /note="GRAIL 2 excellent exon, frame 1"
     misc_feature    31486. .31583
```



```
                     /note="GRAIL 2 excellent exon, frame 2"
     misc_feature    complement(36178. .36260)
...
Note: remainder of annotations omitted.

  Query Match           10.9%;  Score 19;  DB 26;  Length 39188;
  Best Local Similarity 100.0%;  Pred. No. 3.93e+00;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  23221 TTAAAAATAATTTATTTTT 23239
          ||||||||||||||||||| 
Qy     38 TTAAAAATAATTTATTTTT 56


RESULT   18
LOCUS       AC005334   111448 bp    DNA             INV       29-JUL-1998
DEFINITION  Drosophila melanogaster DNA sequence (P1s DS08221 (D130) and
            DS00738 (D271)), complete sequence.
ACCESSION   AC005334 AC004273 AC002643 AC003150 AC002644 AC003151 AC001752
            AC001753 AC003724 AC001754 AC003152 AC001755 AC002645 AC002646
            AC003725 AC003153 AC002647 AC001756 AC002648 AC001757 AC003154
            AC003726 AC001758 AC001759 AC002649 AC003727 AC003728 AC003155
            AC002650 AC002651 AC004569
NID         g1347819
KEYWORDS    HTG.
SOURCE      Drosophila melanogaster (Subclones in sac from P1 clones DS08221
            (D130) and DS00738 (D271)) DNA.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea;
            Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 111448)
  AUTHORS   Celniker,S.E., George,R.A., Galle,R.F., Hoskins,R.A.,
            Svirskas,R.R., Harris,N.L., Agbayani,A., Arcaina,T.T., Baxter,E.,
            Blasej,R.G., Chaves,C., Chew,M., Doyle,C.M., Farfan,D.E.,
            Flanagan,J., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,S.H., Lee,B., Lomotan,M.A., Mak,J., Mazda,P., Mok,H.S.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Punch,E., Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R.,
            Yee,A., Zhang,R., Zieran,L.L. and Kimmel,B.
  TITLE     Sequencing of Drosophila chromosome 2L, region 21B1-21B5
  JOURNAL   Unpublished (1997)
REFERENCE   2  (bases 1 to 111448)
  AUTHORS   Celniker,S.E., George,R.A., Galle,R.F., Hoskins,R.A.,
            Svirskas,R.R., Harris,N.L., Agbayani,A., Arcaina,T.T., Baxter,E.,
            Blasej,R.G., Chaves,C., Chew,M., Doyle,C.M., Farfan,D.E.,
            Flanagan,J., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,S.H., Lee,B., Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Punch,E., Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R.,
            Yee,A., Zhang,R., Zieran,L.L. and Kimmel,B.E.
```



```
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUL-1998) Berkeley Drosophila Genome Project, MS
            64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,
            Berkeley, CA 94720, US
COMMENT     Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://fruitfly.berkeley.edu/sequence/) or send
            email to drosophila@mhgc.lbl.gov.
            Library locations: 61-86, 66-8.
FEATURES    Location/Qualifiers
     source 1. .111448
            /organism="Drosophila melanogaster"
            /db_xref="taxon:7227"
            /chromosome="2L"
            /map="21B1-21B5"
            /clone="P1s DS08221 (D130) and DS00738 (D271)"
            /note="DS08221 (d130) extends from bp 1 to bp 79,931 and
            DS00738 (D271) is a partial bridge extending from bp
            76,285 to P1 end at bp 111,448."
BASE COUNT    31297 a  23397 c  23704 g  33050 t
ORIGIN

  Query Match           10.9%;  Score 19;  DB 18;  Length 111448;
  Best Local Similarity 100.0%;  Pred. No. 3.93e+00;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   4811 CAGATAAAAATAAAACAAG 4829
          ||||||||||||||||||| 
Qy    139 CAGATAAAAATAAAACAAG 157


RESULT   19
LOCUS       HS127B14   126263 bp    DNA             PRI       29-OCT-1997
DEFINITION  Human DNA sequence from PAC 127B14 on chromosome Xq22.
ACCESSION   Z93928
NID         g2578051
KEYWORDS    Xq22.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 126263)
  AUTHORS   Swan,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-OCT-1997) sanger.ac.uk/HGP/ChrX/) Sanger Centre,
            Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquires:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
```

```
COMMENT     On Oct 31, 1997 this sequence version replaced gi:1934872.
            IMPORTANT: This sequence is the entire insert of clone 127B14.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variations annotated may not be found in the sequence submission
            corresponding to the overlapping clone as we submit sequences with
            only a small overlap as described above.
            This sequence was generated from part of bacterial clone contigs of
            human chromosome X, constructed by the Sanger Centre chromosome X
            mapping group. Further information can be found at
            http://www.sanger.ac.uk/HGP/ChrX/
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            The true left end of clone 127B14 is at 1 in this sequence. The
            true right end of clone 127B14 is at 126263.
            127B14 is from the library RPCI1 constructed at the Roswell Park
            Cancer Institute by the group of Pieter de Jong.
            For further details see http://bacpac.med.buffalo.edu/.
FEATURES             Location/Qualifiers
     source          1. .126263
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="Xq22"
                     /clone="127B14"
                     /clone_lib="RPCI1"
     repeat_region   3421. .3781
                     /note="MLT1A1 repeat: matches 1. .365 of consensus"
     repeat_region   4271. .4439
                     /note="MIR repeat: matches 245. .76 of consensus"
     repeat_region   4454. .4755
                     /note="AluSq repeat: matches 1. .303 of consensus"
     repeat_region   4860. .4995
                     /note="MIR2 repeat: matches 5. .141 of consensus"
     repeat_region   7374. .7430
                     /note="MIR2 repeat: matches 146. .95 of consensus"
     repeat_region   7922. .8019
                     /note="MIR repeat: matches 262. .165 of consensus"
     repeat_region   8222. .8429
                     /note="MIR repeat: matches 231. .16 of consensus"
     repeat_region   8495. .8610
                     /note="MIR repeat: matches 72. .191 of consensus"
     repeat_region   8979. .9129
                     /note="MIR2 repeat: matches 1. .143 of consensus"
     repeat_region   10387. .10539
                     /note="MIR repeat: matches 208. .59 of consensus"
```



```
     repeat_region   12204. .12502
                     /note="AluSg repeat: matches 300. .1 of consensus"
     repeat_region   12556. .12854
                     /note="AluSx repeat: matches 296. .1 of consensus"
     repeat_region   13891. .14510
                     /note="L1MB7 repeat: matches 288. .917 of consensus"
     repeat_region   15051. .15384
                     /note="L1MB6 repeat: matches 565. .219 of consensus"
     repeat_region   15386. .15688
                     /note="AluSq repeat: matches 303. .3 of consensus"
     repeat_region   16689. .17042
                     /note="MLT1A2 repeat: matches 1. .374 of consensus"
     repeat_region   17511. .17642
                     /note="MIR repeat: matches 241. .108 of consensus"
     repeat_region   17872. .18117
                     /note="MIR repeat: matches 13. .253 of consensus"
     repeat_region   19966. .20001
                     /note="6 copies of 6 mer 92 % conserved"
     repeat_region   20557. .20694
                     /note="MIR repeat: matches 115. .253 of consensus"
     repeat_region   21340. .21623
                     /note="2 copies of 142 mer 86 % conserved"
     repeat_region   21518. .21723
                     /note="2 copies of 103 mer 82 % conserved"
     repeat_region   22343. .23893
                     /note="L1 repeat: matches 990. .2546 of consensus"
     repeat_region   23894. .24227
                     /note="THE1B-INTERNAL repeat: matches 2. .351 of consensus"
     repeat_region   24243. .25343
                     /note="THE1B-INTERNAL repeat: matches 427. .1580 of
                     consensus"
     repeat_region   25343. .25736
                     /note="THE1C repeat: matches 3. .371 of consensus"
     repeat_region   25736. .26013
                     /note="L1 repeat: matches 2541. .2822 of consensus"
     repeat_region   26015. .26315
                     /note="AluSx repeat: matches 1. .301 of consensus"
     repeat_region   26621. .26807
                     /note="L1 repeat: matches 2813. .3000 of consensus"
     repeat_region   26838. .26972
                     /note="L1PA13 repeat: matches 902. .768 of consensus"
     repeat_region   26976. .28913
                     /note="L1 repeat: matches 2987. .4947 of consensus"
     repeat_region   28921. .29281
                     /note="THE1B repeat: matches 364. .1 of consensus"
     repeat_region   29271. .29767
                     /note="L1 repeat: matches 4940. .5390 of consensus"
     repeat_region   29618. .29828
                     /note="L1PA8 repeat: matches 1. .192 of consensus"
     repeat_region   29829. .30198
                     /note="THE1C repeat: matches 1. .371 of consensus"
```



```
     repeat_region   30167. .30905
                     /note="L1PA14 repeat: matches 155. .896 of consensus"
     repeat_region   30967. .31194
                     /note="MIR repeat: matches 261. .8 of consensus"
     repeat_region   31267. .31368
                     /note="MIR2 repeat: matches 145. .36 of consensus"
     repeat_region   31378. .31681
                     /note="AluSx repeat: matches 1. .302 of consensus"
     unsure          32097. .32099
                     /note="GG or GGG"
     repeat_region   32505. .32695
                     /note="MIR repeat: matches 4. .221 of consensus"
     repeat_region   34231. .34331
                     /note="MIR2 repeat: matches 43. .144 of consensus"
     repeat_region   35633. .36270
                     /note="L1ME2 repeat: matches 676. .6 of consensus"
     repeat_region   36565. .36606
                     /note="7 copies of 6 mer 83 % conserved"
     repeat_region   37077. .37214
                     /note="L1PA13 repeat: matches 896. .759 of consensus"
     repeat_region   37215. .37363
                     /note="L1 repeat: matches 3432. .3286 of consensus"
     repeat_region   37364. .37663
                     /note="AluSx repeat: matches 300. .2 of consensus"
     repeat_region   37664. .37927
                     /note="L1 repeat: matches 3297. .3031 of consensus"
     repeat_region   37945. .39597
                     /note="L1 repeat: matches 2612. .947 of consensus"
     repeat_region   40028. .40072
                     /note="MSTD repeat: matches 351. .394 of consensus"
     repeat_region   40072. .40436
                     /note="L1 repeat: matches 4438. .4079 of consensus"
     repeat_region   40448. .41067
                     /note="L1 repeat: matches 4774. .5390 of consensus"
     repeat_region   40924. .41800
                     /note="L1MA5 repeat: matches 2. .886 of consensus"
     repeat_region   41801. .42094
                     /note="AluSp repeat: matches 1. .295 of consensus"
     repeat_region   42095. .42264
                     /note="L1MA7 repeat: matches 867. .1036 of consensus"
     repeat_region   43103. .43294
                     /note="MIR2 repeat: matches 36. .145 of consensus"
     repeat_region   43938. .44696
                     /note="L1 repeat: matches 4127. .4885 of consensus"
     repeat_region   44689. .44845
                     /note="L1PA4 repeat: matches 737. .893 of consensus"
     repeat_region   44847. .45384
                     /note="L1 repeat: matches 4858. .5390 of consensus"
     repeat_region   45235. .46107
                     /note="L1PA15 repeat: matches 1. .902 of consensus"
     repeat_region   46630. .46922
```



```
                     /note="AluY repeat: matches 1. .293 of consensus"
     repeat_region   51496. .51676
                     /note="MIR repeat: matches 6. .191 of consensus"
     repeat_region   51683. .51981
                     /note="AluY repeat: matches 3. .300 of consensus"
     repeat_region   51982. .52149
                     /note="AluSx repeat: matches 134. .301 of consensus;
                     incomplete repeat"
     repeat_region   52292. .52358
                     /note="MIR repeat: matches 193. .262 of consensus"
     repeat_region   52528. .53095
                     /note="MLT2D repeat: matches 553. .1 of consensus"
     repeat_region   53954. .54095
                     /note="MIR2 repeat: matches 1. .146 of consensus"
     repeat_region   54115. .54298
                     /note="MIR repeat: matches 220. .20 of consensus"
     repeat_region   54328. .54471
                     /note="MIR2 repeat: matches 146. .1 of consensus"
     repeat_region   54547. .54770
                     /note="MIR repeat: matches 29. .262 of consensus"
     repeat_region   57723. .57886
                     /note="MER5A repeat: matches 189. .1 of consensus"
     repeat_region   57952. .58043
                     /note="MIR repeat: matches 143. .45 of consensus"
     repeat_region   58158. .58305
                     /note="MIR repeat: matches 123. .261 of consensus"
     repeat_region   58964. .59095
                     /note="L1ME3A repeat: matches 905. .763 of consensus"
     repeat_region   60202. .60348
                     /note="MIR2 repeat: matches 1. .146 of consensus"
     repeat_region   60420. .60549
                     /note="MIR repeat: matches 21. .150 of consensus"
...
Note: remainder of annotations omitted.

  Query Match           10.9%;  Score 19;  DB 26;  Length 126263;
  Best Local Similarity 100.0%;  Pred. No. 3.93e+00;
  Matches    19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Db 101275 AATAAATTATTTTTAACCT 101293
          ||||||||||||||||||
Cp     53 AATAAATTATTTTTAACCT 35


RESULT   20
LOCUS       AC004740   127270 bp    DNA             PRI       23-MAY-1998
DEFINITION  Homo sapiens PAC clone DJ0631B17 from 7p21-p22, complete sequence.
ACCESSION   AC004740
NID         g3152633
KEYWORDS    HTG.
SOURCE      human.
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 127270)
  AUTHORS   Leonard,S. and Yoakum,M.
  TITLE     The sequence of Homo sapiens PAC clone DJ0631B17
  JOURNAL   Unpublished (1998)
REFERENCE   2  (bases 1 to 127270)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-MAY-1998) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     SUBMITTED BY:   WUGSC
            Genome Sequencing Center
            Department of Genetics
            Washington University
            St. Louis MO 63108, USA
            http://genome.wustl.edu/gsc
            mailto:sapiens@watson.wustl.edu

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded or sequenced with an alternate
            chemistry; an attempt was made to resolve all sequencing problems,
            such as compressions and repeats; all regions were covered by
            sequence from more than one subclone; and the assembly was
            confirmed by restriction digest.

            MAPPING INFORMATION:
            The sequence of this clone was established as part of a mapping and
            sequencing collaboration between the NHGRI Chromosome 7 Mapping
            Project (Eric D. Green, Director), John D. McPherson in the
            Department of Genetics (Washington University), and the Washington
            University Genome Sequencing Center.  For additional information
            about the map position of this sequence, see
            http://www.nhgri.nih.gov/DIR/GTB/CHR7 , send
            mailto:egreen@nhgri.nih.gov , or see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            This clone was derived from human PAC library RPCI-4, prepared by
            Pieter de Jong and coworkers at Roswell Park Cancer Institute,
            using the method described by Ioannou et al., Nature Genetics
            6:84-9 (1994).  The library is from one male donor.  For further
            details, see http://bacpac.med.buffalo.edu/
            The clone is available from Genome Systems, Inc.
            (http://www.genomesystems.com).
            VECTOR:  pCYPAC2
```



```
            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is DJ0651K02, 200 bp overlap. The
            actual start of this clone is at base position 118942 of DJ0651K02;
            actual end is at 127270 of DJ0631B17.

            This clone contains STS sWSS1729 (NID:g1113193).
FEATURES            Location/Qualifiers
     source         1. .127270
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="7"
                    /map="7p21-p22"
                    /clone_lib="RPCI-4"
                    /clone="DJ0631B17"
     repeat_region  250. .426
                    /rpt_family="Retroviral"
     repeat_region  526. .1804
                    /rpt_family="L1"
     repeat_region  1807. .2557
                    /rpt_family="L1"
     repeat_region  3241. .3389
                    /rpt_family="MIR"
     repeat_region  4444. .4502
                    /rpt_family="L2"
     repeat_region  4944. .5244
                    /rpt_family="Alu"
     repeat_region  7957. .8130
                    /rpt_family="MIR"
     repeat_region  9978. .13852
                    /rpt_family="L1"
     repeat_region  14173. .14469
                    /rpt_family="Alu"
     repeat_region  17438. .17553
                    /rpt_family="L2"
     repeat_region  17947. .18279
                    /rpt_family="MaLR"
     repeat_region  19814. .20161
                    /rpt_family="Retroviral"
     repeat_region  20344. .20458
                    /rpt_family="L2"
     repeat_region  21945. .22462
                    /rpt_family="L1"
     repeat_region  25199. .27816
                    /rpt_family="L1"
     repeat_region  28354. .28410
                    /rpt_family="(GA)n"
     repeat_region  28436. .28697
                    /rpt_family="L1"
     repeat_region  28706. .29154
                    /rpt_family="MaLR"
     repeat_region  29186. .29368
```



```
                    /rpt_family="L1"
     repeat_region  29387. .30250
                    /rpt_family="L1"
     repeat_region  30353. .30473
                    /rpt_family="L1"
     repeat_region  30490. .30599
                    /rpt_family="L1"
     repeat_region  30651. .31456
                    /rpt_family="L1"
     repeat_region  32934. .33026
                    /rpt_family="L1"
     repeat_region  33368. .33446
                    /rpt_family="Mariner"
     repeat_region  36761. .37015
                    /rpt_family="L1"
     repeat_region  37092. .37413
                    /rpt_family="L1"
     repeat_region  37414. .37959
                    /rpt_family="L1"
     repeat_region  38204. .38523
                    /rpt_family="MER2_type"
     repeat_region  38550. .38916
                    /rpt_family="L1"
     repeat_region  39067. .39360
                    /rpt_family="Alu"
     repeat_region  39723. .39987
                    /rpt_family="L1"
     repeat_region  39989. .40077
                    /rpt_family="MaLR"
     repeat_region  40107. .40243
                    /rpt_family="MaLR"
     repeat_region  40245. .40873
                    /rpt_family="L1"
     repeat_region  40924. .41000
                    /rpt_family="MaLR"
     repeat_region  41003. .42370
                    /rpt_family="L1"
     repeat_region  43777. .43852
                    /rpt_family="Retroviral"
     repeat_region  43890. .44039
                    /rpt_family="Retroviral"
     repeat_region  45517. .45737
                    /rpt_family="MER1_type"
     repeat_region  45865. .46250
                    /rpt_family="L1"
     repeat_region  47281. .47523
                    /rpt_family="L1"
     repeat_region  47537. .47651
                    /rpt_family="L1"
     repeat_region  47762. .47823
                    /rpt_family="MaLR"
```



```
     repeat_region  48224. .48572
                    /rpt_family="Retroviral"
     repeat_region  49184. .49244
                    /rpt_family="L1"
     repeat_region  50479. .50634
                    /rpt_family="MIR"
     repeat_region  51016. .51111
                    /rpt_family="MER1_type"
     repeat_region  52365. .52664
                    /rpt_family="Alu"
     repeat_region  52765. .52850
                    /rpt_family="MIR"
     repeat_region  54446. .54712
                    /rpt_family="MER1_type"
     repeat_region  57063. .57299
                    /rpt_family="MIR"
     repeat_region  57609. .57664
                    /rpt_family="MER1_type"
     repeat_region  58292. .58651
                    /rpt_family="MaLR"
     repeat_region  58655. .60140
                    /rpt_family="MaLR"
     repeat_region  60142. .60515
                    /rpt_family="MaLR"
     repeat_region  62138. .63844
                    /rpt_family="L1"
     repeat_region  63836. .64162
                    /rpt_family="L1"
     repeat_region  65343. .65514
                    /rpt_family="L1"
     repeat_region  65509. .66765
                    /rpt_family="L1"
     repeat_region  66812. .69075
                    /rpt_family="L1"
     repeat_region  70113. .70214
                    /rpt_family="MIR"
     repeat_region  70425. .70653
                    /rpt_family="L2"
     repeat_region  70921. .71370
                    /rpt_family="Retroviral"
     repeat_region  72456. .72743
...
Note: remainder of annotations omitted.

  Query Match           10.9%;  Score 19;  DB 26;  Length 127270;
  Best Local Similarity 100.0%;  Pred. No. 3.93e+00;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db 117716 ATTTACACATTTCAAAGCA 117734
          |||||||||||||||||||
Cp     84 ATTTACACATTTCAAAGCA 66
```

```
RESULT   21
LOCUS       AC004962   147709 bp    DNA             HTG       12-JUN-1998
DEFINITION  *** SEQUENCING IN PROGRESS *** Homo sapiens clone DJ1099N07; HTGS
            phase 1, 4 unordered pieces.
ACCESSION   AC004962
NID         g3213035
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 147709)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 147709)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-1998) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     ***                                                        ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                        ***
            * This sequence is unfinished. It consists of 4 contigs for
            * which the order is not known; their order in this record is
            * arbitrary. In some cases, the exact lengths of the gaps
            * between the contigs are also unknown; these gaps are  presented
            * as runs of N as a convenience only. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
            *     1     1571: contig of 1571 bp in length
            *  1572     1589: gap of unknown length
            *  1590    17664: contig of 16075 bp in length
            *  17665   17682: gap of unknown length
            *  17683   61881: contig of 44199 bp in length
            *  61882   61899: gap of unknown length
            *  61900   147709: contig of 85810 bp in length.
FEATURES             Location/Qualifiers
     source          1. .147709
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="DJ1099N07"
BASE COUNT    39547 a  35893 c  35502 g  36713 t     54 others
ORIGIN

  Query Match           10.9%;  Score 19;  DB 17;  Length 147709;
  Best Local Similarity 100.0%;  Pred. No. 3.93e+00;
  Matches    19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```



```
Db  32222 AAAAATAAATTATTTTTAA 32240
          ||||||||||||||||||
Cp     56 AAAAATAAATTATTTTTAA 38


RESULT   22
LOCUS       AC003081   149566 bp    DNA             PRI       06-NOV-1997
DEFINITION  Human BAC clone RG009H02 from 7q31, complete sequence.
ACCESSION   AC003081
NID         g2588626
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 149566)
  AUTHORS   Murray,J, Rohlfing,T, David,M and Ahrens,C.
  TITLE     The sequence of H. sapiens BAC clone RG009H02
  JOURNAL   Unpublished (1997)
REFERENCE   2  (bases 1 to 149566)
  AUTHORS   Wateraton,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-NOV-1997) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     SUBMITTED BY:
            Genome Sequencing Center
            Department of Genetics
            Washington University
            St. Louis MO 63108, USA
            http://genome.wustl.edu/gsc
            mailto:sapiens@watson.wustl.edu

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded or sequenced with an alternate
            chemistry; an attempt was made to resolve all sequencing problems,
            such as compressions and repeats; all regions were covered by
            sequence from more than one subclone; and the assembly was
            confirmed by restriction digest.

            MAPPING INFORMATION:
            The sequence of this clone was established as part of a mapping and
            sequencing collaboration between the NHGRI Chromosome 7 Mapping
            Project (Eric D. Green, Director), John D. McPherson in the
            Department of Genetics (Washington University), and the Washington
            University Genome Sequencing Center.  For additional information
```



```
            about the map position of this sequence, see
            http://www.nhgri.nih.gov/DIR/GTB/CHR7 or send
            mailto:egreen@nhgri.nih.gov

            SOURCE INFORMATION:
            Clone RG009H02 is from the first release of the human BAC library
            CITB-978SK-B.  The library contains cloned DNA from the male
            fibroblast cell line 978SK.  See:  Shizuya et al., Proc. Natl.
            Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8
            (1996).  This clone is available from Research Genetics, Inc.
            (http://www.resgen.com).
            VECTOR:  pBeloBAC11
            Selection:  chloramphenicol


            NEIGHBORING SEQUENCE INFORMATION:
            The actual start of this clone is at base position 1 of RG009H02;
            actual end is at 149566 of RG009H02. The orientation of this clone
            is unknown.

            This clone contains STS sWSS1764 (NID:g1254731).
FEATURES             Location/Qualifiers
     source          1. .149566
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /clone="RG009H02"
                     /clone_lib="CITB-978SK-B"
                     /map="7q31"
     repeat_region   complement(2582. .2612)
                     /rpt_family="L1"
     repeat_region   complement(5343. .5372)
                     /rpt_family="L1"
     repeat_region   complement(6507. .6556)
                     /rpt_family="L1"
     repeat_region   6707. .6998
                     /rpt_family="ALU"
     repeat_region   complement(8773. .8991)
                     /rpt_family="L1"
     repeat_region   8783. .8991
                     /rpt_family="L1"
     repeat_region   9441. .9463
                     /rpt_family="L1"
     repeat_region   10532. .11056
                     /rpt_family="L1"
     repeat_region   11111. .11349
                     /rpt_family="L1"
     repeat_region   complement(11292. .11315)
                     /rpt_family="L1"
     repeat_region   11371. .11787
                     /rpt_family="L1"
```



```
     repeat_region   complement(11859. .12046)
                     /rpt_family="MER"
     repeat_region   complement(12201. .12380)
                     /rpt_family="MER"
     repeat_region   complement(12677. .12888)
                     /rpt_family="MER"
     repeat_region   12822. .13274
                     /rpt_family="MER"
     repeat_region   13488. .13664
                     /rpt_family="L1"
     repeat_region   13674. .13963
                     /rpt_family="ALU"
     repeat_region   13994. .14288
                     /rpt_family="MER"
     repeat_region   complement(14069. .14359)
                     /rpt_family="MER"
     repeat_region   14702. .15145
                     /rpt_family="L1"
     repeat_region   15570. .15961
                     /rpt_family="L1"
     repeat_region   16719. .16752
                     /rpt_family="L1"
     repeat_region   17401. .17439
                     /rpt_family="L1"
     repeat_region   20299. .20358
                     /rpt_family="L1"
     repeat_region   complement(27732. .27944)
                     /rpt_family="MER"
     repeat_region   complement(28963. .29390)
                     /rpt_family="L1"
     repeat_region   complement(29504. .29800)
                     /rpt_family="ALU"
     repeat_region   complement(32060. .32091)
                     /rpt_family="L1"
     repeat_region   32138. .32429
                     /rpt_family="ALU"
     repeat_region   32466. .32496
                     /rpt_family="L1"
     repeat_region   complement(33292. .33521)
                     /rpt_family="L1"
     repeat_region   complement(33601. .33648)
                     /rpt_family="L1"
     repeat_region   complement(33687. .34068)
                     /rpt_family="L1"
     repeat_region   34078. .34773
                     /rpt_family="L1"
     repeat_region   34872. .34931
                     /rpt_family="L1"
     repeat_region   35386. .35438
                     /rpt_family="L1"
     repeat_region   35648. .35794
```

```
                     /rpt_family="L1"
     repeat_region   36975. .37345
                     /rpt_family="MER"
     repeat_region   37366. .37657
                     /rpt_family="ALU"
     repeat_region   38419. .38821
                     /rpt_family="MER"
     misc_feature    38870. .38918
                     /note="similar to EST AA508510 (NID:g2246013) nh69g01.s1"
     repeat_region   complement(39706. .40889)
                     /rpt_family="L1"
     repeat_region   40527. .40552
                     /rpt_family="L1"
     repeat_region   40875. .41563
                     /rpt_family="L1"
     repeat_region   44359. .44419
                     /rpt_family="L1"
     repeat_region   46973. .47224
                     /rpt_family="ALU"
     repeat_region   48243. .48667
                     /rpt_family="ALU"
     repeat_region   complement(49290. .49348)
                     /rpt_family="L1"
     repeat_region   50619. .50649
                     /rpt_family="L1"
     repeat_region   complement(51938. .51969)
                     /rpt_family="L1"
     repeat_region   complement(52878. .53176)
                     /rpt_family="ALU"
     repeat_region   complement(54780. .54834)
                     /rpt_family="L1"
     repeat_region   complement(58438. .58521)
                     /rpt_family="L1"
     repeat_region   complement(58593. .58905)
                     /rpt_family="ALU"
     repeat_region   complement(59325. .59599)
                     /rpt_family="ALU"
     repeat_region   59712. .60313
                     /rpt_family="L1"
     repeat_region   complement(60420. .60741)
                     /rpt_family="MER"
     repeat_region   60497. .60741
                     /rpt_family="MER"
     repeat_region   complement(60770. .60990)
                     /rpt_family="MER"
     repeat_region   60813. .61117
                     /rpt_family="MER"
     repeat_region   61307. .61419
                     /rpt_family="MER"
     repeat_region   complement(61423. .61440)
                     /rpt_family="L1"
```



```
     repeat_region   complement(61441. .61730)
                     /rpt_family="ALU"
     repeat_region   61732. .61831
...
Note: remainder of annotations omitted.

  Query Match           10.9%;  Score 19;  DB 26;  Length 149566;
  Best Local Similarity 100.0%;  Pred. No. 3.93e+00;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  55055 AGATAAAAATAAAACAAGT 55073
          ||||||||||||||||||
Qy    140 AGATAAAAATAAAACAAGT 158


RESULT   23
LOCUS       HS404K8    152115 bp    DNA              HTG       06-OCT-1998
DEFINITION  Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone 404K8;
            HTGS phase 1.
ACCESSION   AL023883
NID         g3676200
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 152115)
  AUTHORS   Nickerson,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-OCT-1998) Wellcome Trust Genome Campus, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquires:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 30, 1998 this sequence version replaced gi:3646081.
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and the release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc. Order of segments is not known; 800 n's separate
            segments. Unfinished: dJ404K8  Contig_ID: 02519  acc=AL023883
            Length: 152115 bp.
            ***                                                        ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                        ***
            * This sequence is unfinished. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
FEATURES              Location/Qualifiers
     source           1. .152115
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
```



```
                     /chromosome="6"
                     /clone="404K8"
BASE COUNT    49681 a  30018 c  28464 g  43952 t
ORIGIN

  Query Match           10.9%;  Score 19;  DB 17;  Length 152115;
  Best Local Similarity 100.0%;  Pred. No. 3.93e+00;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  21026 TATTTTTATGTGAAATGTG 21044
          ||||||||||||||||||
Cp    124 TATTTTTATGTGAAATGTG 106


RESULT   24
LOCUS       HS7706     167034 bp    DNA              HTG       29-OCT-1997
DEFINITION  Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone 7706;
            HTGS phase 1.
ACCESSION   Z96804
NID         g2467168
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 167034)
  AUTHORS   Mistry,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-OCT-1997) Wellcome Trust Genome Campus, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquires:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 7, 1997 this sequence version replaced gi:2465072.
            Order of segments is not known; 800 n's separate segments.
            Unfinished sequence: dJ7706  Contig_ID: 01827  Length: 85903 bp
            Unfinished sequence: dJ7706  Contig_ID: 00967  Length: 80331 bp
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and the release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc.
            ***                                                        ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                        ***
            * This sequence is unfinished. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
FEATURES              Location/Qualifiers
     source           1. .167034
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
```



```
                     /chromosome="X"
                     /clone="7706"
BASE COUNT    47631 a  34531 c  33968 g  50102 t    802 others
ORIGIN

  Query Match           10.9%;  Score 19;  DB 17;  Length 167034;
  Best Local Similarity 100.0%;  Pred. No. 3.93e+00;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  27934 ATAAAAATAAAACAAGTGA 27952
          ||||||||||||||||||
Qy    142 ATAAAAATAAAACAAGTGA 160


RESULT   25
LOCUS       CEY1E3     211468 bp    DNA              HTG       20-JAN-1998
DEFINITION  Caenorhabditis elegans DNA *** SEQUENCING IN PROGRESS *** from
            clone Y1E3; HTGS phase 1.
ACCESSION   AL021388
NID         g2808481
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryotae; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 211468)
  AUTHORS   Wallis,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JAN-1998) Nematode Sequencing Project, Sanger Centre,
            Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics,
            Washington University, St. Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and the release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc. Order of segments is not known; 800 n's separate
            segments.
            ***                                                        ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                        ***
            * This sequence is unfinished. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
FEATURES              Location/Qualifiers
     source           1. .211468
                      /organism="Caenorhabditis elegans"
                      /db_xref="taxon:6239"
                      /chromosome="II"
                      /clone="Y1E3"
```

```
  BASE COUNT    67751 a  37255 c  37680 g  67711 t   1071 others
  ORIGIN

    Query Match           10.9%;  Score 19;  DB 17;  Length 211468;
    Best Local Similarity 100.0%;  Pred. No. 3.93e+00;
    Matches    19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

  Db 101139 AAAATAAATTATTTTTAAC 101157
            ||||||||||||||||||
  Cp     55 AAAATAAATTATTTTTAAC 37


  RESULT   26
  LOCUS       G14255         400 bp    DNA                STS       22-DEC-1995
  DEFINITION  human STS SHGC-7426 clone pG-3535.
  ACCESSION   G14255
  NID         g1129994
  KEYWORDS    STS sequence; primer; sequence tagged site.
  SOURCE      human.
    ORGANISM  Homo sapiens
              Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
              Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Choanata;
              Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates;
              Catarrhini; Hominidae; Homo.
  REFERENCE   1  (bases 1 to 400)
    AUTHORS   Myers,R.M.
    JOURNAL   Unpublished (1995)
  COMMENT
              Contact: Richard M. Myers
              Stanford Human Genome Center (SHGC)
              Stanford University School of Medicine
              Department of Genetics, M-344, Stanford, CA 94305, USA
              Tel: 4157259687
              Fax: 4157259689
              Email: myers@shgc.stanford.edu

              Primer A: GACTGGCTGCTTTCACTTCAC
              Primer B: GGGTTTCAGCGTAAGCTATCC
              STS size: 253
              PCR Profile:
                     Initial incubation: 94 degrees C for 90 seconds

                     Denaturation:       94 degrees C for 15 seconds
                     Annealing:          62 degrees C for 23 seconds
                     Polymerization:     72 degrees C for 30 seconds
                     PCR Cycles:         30
                     Thermal Cycler:     Perkin Elmer 9600
              Protocol:
                     Template:       25 ng
                     Primer:         each 1 uM
                     dNTPs:          each 200 uM
```



```
                     Taq Polymerase: 0.05 units/ul
                     Total Vol:      10 ul

              Buffer:
                     MgCl2:          2.5 mM
                     KCl:            50  mM
                     Tris-HCl:       20  mM
                     pH:             8.3

              Plasmid clones, generated from a lymphoblastoid cell line from a
              human male. Localized to human chromosome 5 by analysis on the
              NIGMS Human/Rodent Somatic Cell Hybrid Panel #1, Coriell Institue
              for Medical Research, Camden, NJ 08103.
  FEATURES             Location/Qualifiers
       source          1. .400
                       /organism="Homo sapiens"
       STS             28. .280
       primer_bind     28. .48
       primer_bind     complement(260. .280)
  BASE COUNT      128 a     59 c     70 g    133 t     10 others
  ORIGIN

    Query Match           10.3%;  Score 18;  DB 30;  Length 400;
    Best Local Similarity 100.0%;  Pred. No. 1.66e+01;
    Matches    18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

  Db    206 TATGCCAGATAAAAATAA 223
            ||||||||||||||||||
  Qy    134 TATGCCAGATAAAAATAA 151


  RESULT   27
  LOCUS       BOU17773       436 bp    rRNA               INV       04-APR-1995
  DEFINITION  Blatta orientalis mitochondrion 12S ribosomal RNA, partial
              sequence.
  ACCESSION   U17773
  NID         g623271
  KEYWORDS    .
  SOURCE      oriental cockroach.
    ORGANISM  Mitochondrion Blatta orientalis
              Eukaryotae; mitochondrial eukaryotes; Metazoa; Arthropoda;
              Tracheata; Insecta; Pterygota; Orthopterodea; Blattaria; Blattidae;
              Blatta.
  REFERENCE   1  (bases 1 to 436)
    AUTHORS   Kambhampati,S.
    TITLE     A phylogeny of cockroaches and related insects based on DNA
              sequence of mitochondrial ribosomal RNA genes
    JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 92 (6), 2017-2020 (1995)
    MEDLINE   95199284
  REFERENCE   2  (bases 1 to 436)
    AUTHORS   Kambhampati,S.
```



```
    TITLE     Direct Submission
    JOURNAL   Submitted (29-NOV-1994) Srinivas Kambhampati, Entomology, Kansas
              State University, Waters Hall, Manhattan, KS 66506, USA
  FEATURES             Location/Qualifiers
       source          1. .436
                       /organism="Blatta orientalis"
                       /mitochondrion
                       /db_xref="taxon:6976"
                       /sex="female"
                       /tissue_type="fat body"
                       /dev_stage="adult"
       rRNA            <1. .>436
                       /product="12S ribosomal RNA"
  BASE COUNT      181 a     69 c     47 g    139 t
  ORIGIN

    Query Match           10.3%;  Score 18;  DB 18;  Length 436;
    Best Local Similarity 100.0%;  Pred. No. 1.66e+01;
    Matches    18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

  Db    385 AGGTTAAAAATAATTTAT 402
            ||||||||||||||||||
  Qy     35 AGGTTAAAAATAATTTAT 52


  RESULT   28
  LOCUS       HUMZD55D10     662 bp    mRNA               PRI       29-AUG-1998
  DEFINITION  Homo sapiens full length insert cDNA clone ZD55D10.
  ACCESSION   AF086334
  NID         g3483679
  KEYWORDS    FLI_CDNA.
  SOURCE      human.
    ORGANISM  Homo sapiens
              Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
              Primates; Catarrhini; Hominidae; Homo.
  REFERENCE   1  (bases 1 to 662)
    AUTHORS   Woessner,J., Tan,F., Marra,M., Kucaba,T., Yandell,M., Martin,J.,
              Marth,G., Bowles,L., Wylie,T., Bowers,Y., Steptoe,M., Theising,B.,
              Geisel,S., Allen,M., Underwood,K., Chappell,J., Person,B.,
              Gibbons,M., Harvey,N., Pape,D., Chamberlain,A., Morales,R.,
              Schurk,R., Ritter,E., Kohn,S., Swaller,T., Behymer,K., Hillier,L.,
              Wilson,R. and Waterston,R.
    TITLE     Full Clone Sequencing of the Longest Available Member from Each
              Unigene Cluster
    JOURNAL   Unpublished
  REFERENCE   2  (bases 1 to 662)
    AUTHORS   Waterston,R.
    TITLE     Direct Submission
    JOURNAL   Submitted (24-AUG-1998) Department of Genetics, Washington
              University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
  COMMENT     SUBMITTED BY:
```



```
              Genome Sequencing Center
              Department of Genetics
              Washington University
              St. Louis MO 63108, USA
              http://genome.wustl.edu/gsc
              mailto:est@watson.wustl.edu

              NOTICE:  This sequence represents the full insert of this cDNA.  No
              attempt has been made to verify whether this corresponds to the
              full-length of the original mRNA from which it was derived.  We
              have tried to obtain double-stranded, or double chemistry sequence
              across the entire clone, but potentially, there are areas in the
              sequence where this level of coverage was not achieved.
              Nevertheless, we are confident of the accuracy of this sequence as
              all regions of low quality, as defined by PHRAP (P. Green, in
              preparation), were visually inspected and edited accordingly.  The
              consensus quality values for this sequence have been submitted
              separately.

              The location of this clone is unknown.
  FEATURES             Location/Qualifiers
       source          1. .662
                       /organism="Homo sapiens"
                       /db_xref="taxon:9606"
                       /clone_lib="Soares_fetal_heart_NbHH19W"
                       /clone="IMAGE:344563"
  BASE COUNT      209 a     94 c    130 g    229 t
  ORIGIN

    Query Match           10.3%;  Score 18;  DB 27;  Length 662;
    Best Local Similarity 100.0%;  Pred. No. 1.66e+01;
    Matches    18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

  Db    417 AAAAATAAATTATTTTTA 434
            ||||||||||||||||||
  Cp     56 AAAAATAAATTATTTTTA 39


  RESULT   29
  LOCUS       DDU41289      4037 bp    mRNA               INV       10-SEP-1996
  DEFINITION  Dictyostelium discoideum K7 kinesin-like protein mRNA, complete
              cds.
  ACCESSION   U41289
  NID         g1526990
  KEYWORDS    .
  SOURCE      slime mold strain=AX2.
    ORGANISM  Dictyostelium discoideum
              Eukaryotae; mitochondrial eukaryotes; Dictyosteliida;
              Dictyostelium.
  REFERENCE   1  (bases 1 to 4037)
    AUTHORS   deHostos,E.L., McCaffrey,G. and Vale,R.D.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (24-NOV-1995) Eugenio L. deHostos, Biochemistry, Rice
            University, MS 140, 6100 Main St., Houston, TX 77005-1892, USA
COMMENT     On Sep 9, 1996 this sequence version replaced gi:1145767.
FEATURES             Location/Qualifiers
     source          1. .4037
                     /organism="Dictyostelium discoideum"
                     /strain="AX2 and AX3"
                     /db_xref="taxon:44689"
     CDS             208. .3972
                     /codon_start=1
                     /product="K7 kinesin-like protein"
                     /db_xref="PID:g1526991"
                     /translation="MESPVVEGNSGEVATPTLPQPPTPVSSNIRVVCRVRPLTELEKG
                     RNEHSIVHFPDSKSISIRANGPQPTFDRIFGYQETQSQIFEDVAEPIVWDFLDGYHGT
                     IIAYGQTASGKTFTMVGDPDSHGIIPRVIESIFVGISKMREKDTSLSLAPCLKISALE
                     LYNEKLYDLYIASKSNLNIREHKQNGIYVEGISEIVITSIEEAYNFLNISNNNRAIAS
                     TKMSAASSRSHSVLMIELSQQNLSMESSKISKLFLVDLAGSDISHKTGAEGDRMQEAK
                     NINLSLSALGKVINALTCGVNYVPYRDSKLTRVLQDSLGGNSKTSLIINCSPSNNNEH
                     ETITTLQFGTRAKTIKNQPKINKKITYHELELFIIKLAKDLEKSRRECEEITRSKNLE
                     INNLLIQLENNQKMVVESNQKLELLNSQISSNHSFDNTPKEIENTCENSKIIFDDLND
                     HINNNNNVDENNNTNNNDNNNNDNNNNNQYQEESNQYQQENNQKDGDQNNSSFDSIKV
                     EDLRDLDDEPDIEDIILNSTLGNISDDDDDDDDHHSNNNNVDDNNNGEINNDSDGYLN
                     RSLKDIKIPEISDLNDHNINNNNNNNNNNINNDNNSNSGGLRVSTSYITSSPNLSIKEY
                     GCHRNSPPLFSYFKTKDPPPSSDENDKFFNDLIAKCENEQQQQQQQHNDDDEDIKSTT
                     SNATTTTITTIDMNASHPSGIDDPIEFTIIKSDKTITSTIRETIQPSSLSNSTSLLD
                     IETVESSTLPAPPPVTTTTTLTTVTTTKLTKTTNIPSNTNDINSIDDPGPSKIEEEGS
                     SSNRKPNDTAILSPGDDDDEENEDNENEDVIVDSDEDTHSGKNNLLNTFKNDHHRGDF
                     GATPTKSIFNKNGNITIKEFETPQQQQQQQQQQQQQQQQQQQQQQQPLILQTTSTNPTII
                     SIKSNKEPSPSSSTTTSIKKKNFNKRRSWIIFTIILTITLVSSSLLCLYLPEYKERLV
                     QRRGYLNKLGIYSDIPTNEKISLAQHNQISLAKELYGGNSKQYYDEMSSFNTAYNHLI
                     EMNHLETAVSKIPGSAIDLRPSGDDVINSIECKRAIHKLKTNNYVNGDLDQQQQQHNY
                     ITKIDQLSEQSKEQNQLIENFKLDLKNKTSEIEKLEKEIKQKDNKIKEKEEKIELIES
                     RVLNEEKGGEKVLEDQIISLRNDNTLSTQILNLEGDKKSLGVLVIKLNSDKTEIQNE
                     VKELKREVQELEDAPIALIPNPFVKVKSFTVEKKSWFVENIYKIWNWFK"
BASE COUNT     1747 a     573 c     541 g    1176 t
ORIGIN

  Query Match           10.3%;  Score 18;  DB 18;  Length 4037;
  Best Local Similarity 100.0%;  Pred. No. 1.66e+01;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   3547 AAAAATAAAACAAGTGAA 3564
           ||||||||||||||||||
Qy    144 AAAAATAAAACAAGTGAA 161


RESULT   30
LOCUS       MMYPT14      4678 bp    DNA             ROD       05-APR-1995
DEFINITION  Mouse yptl gene for ras-related GTP-binding protein, exons 4-6.
ACCESSION   X15747
```



```
NID         g55463
KEYWORDS    GTP-binding protein; ras-related protein; yptl gene.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;
            Murinae; Mus.
REFERENCE   1  (bases 1 to 4678)
  AUTHORS   Wichmann,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUL-1989) Wichmann H., MPI of Biophysical Chemistry,
            Dept of Molecular Genetics, P O Box 2841, D-3400 Goettingen, F R G
REFERENCE   2  (bases 1 to 4678)
  AUTHORS   Wichmann,H., Disela,C., Haubruck,H. and Gallwitz,D.
  TITLE     Nucleotide sequence of the mouse yptl gene encoding a ras-related
            GTP-binding protein
  JOURNAL   Nucleic Acids Res. 17 (16), 6737-6738 (1989)
  MEDLINE   89386011
COMMENT     Data kindly reviewed (27-OCT-1989) by Wichmann H.
FEATURES             Location/Qualifiers
     source          1. .4678
                     /organism="Mus musculus"
                     /strain="c(CH)/c(CH)"
                     /db_xref="taxon:10090"
     precursor_RNA   <1. .2682
                     /note="primary transcript"
     exon            239. .334
                     /note="Exon 4"
                     /usedin=X15744:gtpb_cds
     intron          335. .1521
                     /note="Intron IV"
     exon            1522. .1653
                     /note="Exon 5"
                     /usedin=X15744:gtpb_cds
     intron          1654. .1802
                     /note="Intron V"
     misc_feature    1803. .2000
                     /note="GTP-binding protein, exon 6 (AA 141-205)"
                     /usedin=X15744:gtpb_cds
     mRNA            1803. .2682
                     /note="Exon 6"
     misc_feature    2586. .2591
                     /note="pot. polyA signal (1.6 kb mRNA)"
     misc_feature    2654. .2659
                     /note="pot. alt. polyA signal (1.6 kb mRNA)"
     polyA_site      2682
                     /note="polyA site"
BASE COUNT     1338 a     913 c     905 g    1522 t
ORIGIN

  Query Match           10.3%;  Score 18;  DB 28;  Length 4678;
```



```
  Best Local Similarity 100.0%;  Pred. No. 1.66e+01;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    947 GCATATTTTTATGTGAAA 964
          ||||||||||||||||||
Cp    127 GCATATTTTTATGTGAAA 110


RESULT   31
LOCUS       STYSINRPHO   4880 bp    DNA             BCT       16-FEB-1993
DEFINITION  Salmonella typhimurium insert regulator nonspecific acid
            phosphatase (sinR) gene, complete cds.
ACCESSION   L04307
NID         g154370
KEYWORDS    nonspecific acid phophatase; nonspecific acid phosphatase; sinR
            gene; transcriptional regulation.
SOURCE      Salmonella typhimurium DNA.
  ORGANISM  Salmonella typhimurium
            Eubacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Salmonella.
REFERENCE   1  (bases 1 to 4880)
  AUTHORS   Groisman,E.A., Sturmoski,M.A., Solomon,F.R., Lin,R. and Ochman,H.
  TITLE     Molecular, functional and evolutionary analysis of sequences
            specific to Salmonella
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. (1993) In press
FEATURES             Location/Qualifiers
     source          1. .4880
                     /organism="Salmonella typhimurium"
                     /db_xref="taxon:602"
     RBS             2233. .2236
                     /gene="sinR"
     gene            2233. .3188
                     /gene="sinR"
     CDS             2241. .3188
                     /gene="sinR"
                     /codon_start=1
                     /transl_table=11
                     /product="nonspecific acid phosphatase"
                     /db_xref="PID:g154371"
                     /translation="MMELINNRGMRDWMIFIKVAEVGNLSRAARELDISISAVSKSLS
                     RLENSIEVTLLRRDSHHLELTGAGQTAYASMKRITSSPQSLLDELRNPDKIIRGSIKF
                     SAPAIVCEFLANKWIWEFTASYPDTKIYLDSRERSDFFSKSLEFDELVFKSGIIESED
                     LVYRKISPLKLVLCASPKYIRKYGRISHPGDLENHIIVGLHNHGLSGPLTLFRQDESY
                     TISGAVNVHLSSNNLLSVLNLVLEGKGINLMTPAWLATKYLKNNELEIILPEWRVPDL
                     PIYLVWRHRQYYSPLFQRFLSFIEDKWNNRPQIDFLNDD"
BASE COUNT     1288 a     984 c    1175 g    1433 t
ORIGIN

  Query Match           10.3%;  Score 18;  DB 15;  Length 4880;
  Best Local Similarity 100.0%;  Pred. No. 1.66e+01;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```



```
Db    566 TAAAAATAATTTATTTTT 583
          ||||||||||||||||||
Qy     39 TAAAAATAATTTATTTTT 56


RESULT   32
LOCUS       TOBPG13A     5837 bp    DNA             PLN       23-OCT-1991
DEFINITION  Nicotiana tabacum PG13 (ASF-1/G13) gene, complete cds.
ACCESSION   M62855
NID         g170284
KEYWORDS    leucine zipper DNA binding protein; tobacco nuclear factor;
            transcriptional activator.
SOURCE      Nicotiana tabacum (library: charon 35 lambda) DNA.
  ORGANISM  Nicotiana tabacum
            Eukaryotae; mitochondrial eukaryotes; Viridiplantae;
            Charophyta/Embryophyta group; Embryophyta; Magnoliophyta;
            Magnoliopsida; Solananae; Solanales; Solanaceae; Nicotiana.
REFERENCE   1  (sites)
  AUTHORS   Fromm,H., Katagiri,F. and Chua,N.-H.
  TITLE     The tobacco transcription activator TGA1a binds to a sequence in
            the 5' upstream region of a gene encoding a TGA1a-related protein
  JOURNAL   Mol. Gen. Genet. 229, 181-188 (1991)
  MEDLINE   92017650
FEATURES             Location/Qualifiers
     source          1. .5837
                     /organism="Nicotiana tabacum"
                     /db_xref="taxon:4097"
                     /tissue_lib="charon 35 lambda"
     mRNA            join(735. .972,1049. .1123,1549. .1608,1739. .1816,1946. .2009,
                     3334. .3605,4699. .4926,5285. .5437)
                     /gene="ASF-1/G13"
     gene            join(735. .972,1049. .1123,1549. .1608,1739. .1816,1946. .2009,
                     3334. .3605,4699. .4926,5285. .5437)
                     /gene="ASF-1/G13"
     exon            735. .972
                     /gene="ASF-1/G13"
                     /number=1
     CDS             join(784. .972,1049. .1123,1549. .1608,1739. .1816,1946. .2009,
                     3334. .3605,4699. .4926,5285. .5437)
                     /gene="ASF-1/G13"
                     /note="leucine-zipper DNA-binding protein"
                     /codon_start=1
                     /db_xref="PID:g170285"
                     /translation="MNSTYTQFVASKRMGICDPIHQIGMWGDFKNSSPPDTSATLILE
                     VENCLENEHPIMEKRPHNETEELSHGTVGTSMNYEPEARKPIDKVLRRLAQNREAARK
                     SRLRKKAYVQQLENSKLKLLQLEQELERTRQQGQYAGVGLDESQIGYTGTANSGIVAF
                     EMETGHWVEEQDRQTDELRSALNSQVGEIELQLLVEGCLNHYFDLFRMKAAAAHADVL
                     PLMTGTWKTSAERFPLWIAGFRPSELLKVLTPNVEPLTEQQLRDVCNLMQSCQQAEDA
                     LSQGMVKLHQILAEAVAAGTLGDGIILPKMAATIENLEALVRFVNQADHLRKETLLQM
                     SCILTPQQSAQGLLALGEYFKRLRALSSLWTSRTSEPA"
```

```
     intron          973. .1048
                     /gene="ASF-1/G13"
                     /number=1
     exon            1049. .1123
                     /gene="ASF-1/G13"
                     /number=2
     intron          1124. .1548
                     /gene="ASF-1/G13"
                     /number=2
     exon            1549. .1608
                     /gene="ASF-1/G13"
                     /number=3
     intron          1609. .1738
                     /gene="ASF-1/G13"
                     /number=3
     exon            1739. .1816
                     /gene="ASF-1/G13"
                     /number=4
     intron          1817. .1945
                     /gene="ASF-1/G13"
                     /number=4
     exon            1946. .2009
                     /gene="ASF-1/G13"
                     /number=5
     intron          2010. .3333
                     /gene="ASF-1/G13"
                     /number=5
     exon            3334. .3605
                     /gene="ASF-1/G13"
                     /number=6
     intron          3606. .4698
                     /gene="ASF-1/G13"
                     /number=6
     exon            4699. .4926
                     /gene="ASF-1/G13"
                     /number=7
     intron          4927. .5284
                     /gene="ASF-1/G13"
                     /number=7
     exon            5285. .5437
                     /gene="ASF-1/G13"
                     /number=8
BASE COUNT     1679 a   1087 c   1082 g   1989 t
ORIGIN

  Query Match           10.3%;  Score 18;  DB 23;  Length 5837;
  Best Local Similarity 100.0%;  Pred. No. 1.66e+01;
  Matches    18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   4360 AAAAATAAATTATTTTTA 4377
         ||||||||||||||||||
```



```
Cp     56 AAAAATAAATTATTTTTA 39


RESULT   33
LOCUS       RBFPS       10702 bp    DNA             VRL       19-DEC-1991
DEFINITION  Malignant rabbit fibroma virus DNA fragment.
ACCESSION   M32743
NID         g333609
KEYWORDS    .
SOURCE      Malignant rabbit fibroma virus DNA.
  ORGANISM  Rabbit fibroma virus
            Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
            Leporipoxvirus.
REFERENCE   1  (bases 1 to 10702)
  AUTHORS   Strayer,D.S., Jerng,H.H. and O'Connor,K.
  TITLE     Sequence and analysis of a portion of the genomes of Shope fibroma
            virus and malignant rabbit fibroma virus that is important for
            viral replication in lymphocytes
  JOURNAL   Virology 185, 585-595 (1991)
  MEDLINE   92074222
COMMENT     Draft entry and computer readable sequence for [Unpublished (1990)
            Univ Texas Health Center, 6431 Fannin Street,Houston, TX 7703]
            kindly submitted
            by G.Murray 09-MAR-1990.
FEATURES             Location/Qualifiers
     source          1. .10702
                     /organism="Rabbit fibroma virus"
                     /db_xref="taxon:10271"
BASE COUNT     3031 a   2225 c   2055 g   3391 t
ORIGIN

  Query Match           10.3%;  Score 18;  DB 33;  Length 10702;
  Best Local Similarity 100.0%;  Pred. No. 1.66e+01;
  Matches    18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2288 TTTATATGCCAGATAAAA 2305
         ||||||||||||||||||
Qy    130 TTTATATGCCAGATAAAA 147


RESULT   34
LOCUS       MUSBA       11763 bp    DNA             ROD       04-FEB-1997
DEFINITION  Mouse gene for basigin, complete cds (exon1-7).
ACCESSION   D82019
NID         g1127024
KEYWORDS    Bsg; basigin; immunoglobulin superfamily.
SOURCE      Mus musculus (strain:BALB/c) liver DNA.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae;
            Murinae; Mus.
```



```
REFERENCE   1  (sites)
  AUTHORS   Miyauchi,T., Kanekura,T., Yamaoka,A., Ozawa,M., Miyazawa,S. and
            Muramatsu,T.
  TITLE     Basigin, a new, broadly distributed member of the immunoglobulin
            superfamily, has strong homology with both the immunoglobulin V
            domain and the beta-chain of major histocompatibility complex class
            II antigen
  JOURNAL   J. Biochem. 107 (2), 316-323 (1990)
  MEDLINE   90299872
REFERENCE   2  (bases 1 to 11763)
  AUTHORS   Miyauchi,T., Jimma,F., Igakura,T., Yu,S., Ozawa,M. and Muramatsu,T.
  TITLE     Structure of the mouse basigin gene, a unique member of the
            immunoglobulin superfamily
  JOURNAL   J. Biochem. 118 (4), 717-724 (1995)
  MEDLINE   96157707
REFERENCE   3  (bases 1 to 11763)
  AUTHORS   Miyauchi,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-DEC-1995) to the DDBJ/EMBL/GenBank databases. Teruo
            Miyauchi, Japan Immunoresearch Laboratories Co., Ltd. (JIMRO);
            Nishiyokote-cho 351-1, Takasaki, Gunma 370, Japan
            (Tel:0273-53-1411, Fax:0273-53-1770)
FEATURES             Location/Qualifiers
     source          1. .11763
                     /organism="Mus musculus"
                     /strain="BALB/c"
                     /db_xref="taxon:10090"
                     /chromosome="10"
                     /tissue_type="liver"
     promoter        2344. .2579
                     /label=CpG_island
                     /phenotype="It is likely that CpG islands in the upstream
                     region of housekeeping genes serve as
                     transcription-regulation elements."
     prim_transcript 2602. .10070
                     /citation=[2]
     exon            2602. .2688
                     /label=exon1
                     /number=1
     5'UTR           2602. .2621
                     /citation=[2]
     sig_peptide     2622. .2684
                     /gene="Bsg"
                     /product="basigin signal peptide"
     gene            2622. .9659
                     /gene="Bsg"
     CDS             join(2622. .2688,7773. .7929,8160. .8242,8362. .8504,
                     8740. .9022,9358. .9382,9596. .9659)
                     /gene="Bsg"
                     /citation=[1]
                     /codon_start=1
```



```
                     /product="basigin precursor"
                     /db_xref="PID:d1012174"
                     /db_xref="PID:g1127025"
                     /translation="MAAALLLALAFTLLSGQGACAAAGTIQTSVQEVNSKTQLTCSLN
                     SSGVDIVGHRWMRGGEVLQEDTLPDLHTKYIVDADDRSGETSCIFLPEPVGRSEINVE
                     GPPRIKVGKKSEHSSEGELAKLVCKSDASYPPITDWFWFKTSDIGEEEAITNSTEANG
                     KTVVVSTPEKSQLTISNLDVNVDPGTYVCNATNAQGTTRETISLRVRSRMAALWPFLG
                     IVAEVLVLVTIIFIYEKRRKPDQTLDEDDPGAAPLKGSGTHMNDKDKNVRQRNAT"
     mat_peptide     join(2685. .2688,7773. .7929,8160. .8242,8362. .8504,
                     8740. .9022,9358. .9382,9596. .9656)
                     /gene="Bsg"
                     /product="basigin"
     intron          2689. .7772
                     /gene="Bsg"
                     /label=intron1
                     /number=1
     exon            7773. .7929
                     /gene="Bsg"
                     /label=exon2
                     /number=2
     intron          7930. .8159
                     /gene="Bsg"
                     /label=intron2
                     /number=2
     exon            8160. .8242
                     /gene="Bsg"
                     /label=exon3
                     /number=3
     intron          8243. .8361
                     /gene="Bsg"
                     /label=intron3
                     /number=3
     exon            8362. .8504
                     /gene="Bsg"
                     /label=exon4
                     /number=4
     intron          8505. .8739
                     /gene="Bsg"
                     /label=intron4
                     /number=4
     exon            8740. .9022
                     /gene="Bsg"
                     /label=exon5
                     /number=5
     intron          9023. .9357
                     /gene="Bsg"
                     /label=intron5
                     /number=5
     exon            9358. .9382
                     /gene="Bsg"
                     /label=exon6
```

```
                     /number=6
     intron          9383. .9595
                     /gene="Bsg"
                     /label=intron6
                     /number=6
     exon            9596. .10070
                     /label=exon7
                     /number=7
     3'UTR           9660. .10070
                     /citation=[2]
     polyA_signal    10052. .10057
                     /citation=[2]
     polyA_site      10070
                     /citation=[1]
BASE COUNT     2654 a   2934 c   3403 g   2772 t
ORIGIN

  Query Match           10.3%;  Score 18;  DB 28;  Length 11763;
  Best Local Similarity 100.0%;  Pred. No. 1.66e+01;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    123 TTAAAAATAATTTATTTT 140
          ||||||||||||||||||
Qy     38 TTAAAAATAATTTATTTT 55


RESULT   35
LOCUS       MIIICG      16550 bp    DNA   circular  BCT       05-SEP-1996
DEFINITION  Methanococcus jannaschii small extra-chromosomal element, complete
            sequence.
ACCESSION   L77119
NID         g1500688
KEYWORDS
SOURCE      Methanococcus jannaschii.
  ORGANISM  Methanococcus jannaschii
            Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
            Methanococcus.
REFERENCE   1  (bases 1 to 16550)
  AUTHORS   Bult,C.J., White,O., Olsen,G.J., Zhou,L., Fleischmann,R.D.,
            Sutton,G.G., Blake,J.A., FitzGerald,L.M., Clayton,R.A.,
            Gocayne,J.D., Kerlavage,A.R., Dougherty,B.A., Tomb,J.-F.,
            Adams,M.D., Reich,C.I., Overbeek,R., Kirkness,E.F., Weinstock,K.G.,
            Merrick,J.M., Glodek,A., Scott,J.L., Geoghagen,N.S.M.,
            Weidman,J.F., Fuhrmann,J.L., Presley,E.A., Nguyen,D.,
            Utterback,T.R., Kelley,J.M., Peterson,J.D., Sadow,P.W., Hanna,M.C.,
            Cotton,M.D., Hurst,M.A., Roberts,K.M., Kaine,B.P., Borodovsky,M.,
            Klenk,H.-P., Fraser,C.M., Smith,H.O., Woese,C.R. and Venter,J.C.
  TITLE     Complete genome sequence of the methanogenic archaeon,
            Methanococcus jannaschii
  JOURNAL   Science 273 (5278), 1058-1073 (1996)
  MEDLINE   96337999
```



```
REFERENCE   2  (bases 1 to 16550)
  AUTHORS   Bult,C.J., White,O., Olsen,G.J., Zhou,L., Fleischmann,R.D.,
            Sutton,G.G., Blake,J.A., FitzGerald,L.M., Clayton,R.A.,
            Gocayne,J.D., Kerlavage,A.R., Dougherty,B.A., Tomb,J.-F.,
            Adams,M.D., Reich,C.I., Overbeek,R., Kirkness,E.F., Weinstock,K.G.,
            Merrick,J.M., Glodek,A., Scott,J.L., Geoghagen,N.S.M.,
            Weidman,J.F., Fuhrmann,J.L., Presley,E.A., Nguyen,D.,
            Utterback,T.R., Kelley,J.M., Peterson,J.D., Sadow,P.W., Hanna,M.C.,
            Cotton,M.D., Hurst,M.A., Roberts,K.M., Kaine,B.P., Borodovsky,M.,
            Klenk,H.-P., Fraser,C.M., Smith,H.O., Woese,C.R. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-AUG-1996) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, Maryland 20850, USA
REFERENCE   3  (bases 1 to 16550)
  AUTHORS   White,O., Kerlavage,A.R., Adams,M.D. and Klenk,H.-P.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-SEP-1996) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, Maryland 20850, USA
  REMARK    The starts of translation for some of the genes in the
            Methanococcus jannaschii extra-chromosomal element have been
            extended.
COMMENT     Address all correspondence to:
            J. Craig Venter
            The Institute for Genomic Research
            9712 Medical Center Dr,
            Rockville, MD, 20850, USA

            Carol J. Bult is with the National Center for Geographic
            Information and Analysis, University of Maine, Orono, ME.

            Judith A. Blake is with the Jackson Laboratory, Bar Harbor,
            ME,04649-1500

            Hamilton O. Smith is with the Johns Hopkins University School of
            Medicine, Baltimore, MD, 21205.

            Gary J. Olson, Claudia I. Reich, Brian P. Klaine, and Carl R. Woese
            are with Microbiology Department, University of Illinois,
            Champaign-Urbana, IL, 61801.

            Joseph M. Merrick is with the State University of New York,
            Department of Microbiology, Buffalo, NY, 14214.

            All other authors are with The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD, 20850.

            Predicted open reading frames were determined using GeneMark
            software, kindly supplied by Mark Borodovsky, Georgia Institute of
            Technology, Atlanta, GA, 30332.
            e-mail: mark@amber.gatech.edu
```



```
            Submission and annotation: Owen White
            e-mail: owhite@tigr.org

            Biological role information, putative identifications, sequence
            alignments, on-line name and sequence search capability are
            available at TIGR's World Wide Web site.
            (URL: http://www.tigr.org/).
FEATURES             Location/Qualifiers
     source          1. .16550
                     /organism="Methanococcus jannaschii"
                     /db_xref="taxon:2190"
     gene            complement(429. .1259)
                     /gene="MJECS01"
     CDS             complement(429. .1259)
                     /gene="MJECS01"
                     /note="identified by GeneMark; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="M. jannaschii predicted coding region MJECS01"
                     /db_xref="PID:g1522635"
                     /translation="MKFIPEKYYRHPINKYNGSKITPISQKNIPVVIKVELDDNELNN
                     IHDEAKKLAEDVCEKNPSGDIRTEERKLLSSFCGLIAEKVVKQLITEKLMLIKDYIEI
                     KGANVSNENFNYESHNDIEIHLSNDEIITIEIRSSIATKYDLNDILNKDFDILGSYTT
                     SYKSKEHNKDYIFRIIPHNPGDYGYNPHLKKWEYNKRLENAPKQENKALDYFKNNVK
                     VTFVGGCTFDDLENYGEHDDLKQKNAETWIIRPITKGRDVLEILELKDILNKLGFL
                     "
     gene            complement(1269. .4814)
                     /gene="MJECS02"
     CDS             complement(1269. .4814)
                     /gene="MJECS02"
                     /note="identified by GeneMark; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="M. jannaschii predicted coding region MJECS02"
                     /db_xref="PID:g1522636"
                     /translation="MVVMAQTKLCSDIEIPKEVYNKWKDFIKLVGNIILSIKQIPELE
                     GKFKELLKKGRYNFKSDDPGGQLPEPPTRQKVIEPILEFLGYEPTSEISKKSPLGDRK
                     IPDYRVSVFNKEILIEAEPLGSDLNKKDSGIHQVKEWLIIKSYGVDTGIATNGLEWVL
                     LHYDDTIKEIRTLKELNLKSIPEYVLENKKDKDLENELKQVFSEPYYCPSKEYIEEYI
                     EVATKNIKHKKEEITNEFTKEFVKLVPGFEDVKDVKKKDKSSSEKDKGTKKCLYNCIE
                     APPNTSELDKKKFAVLLMNRLIFIKFLEDKGIVPRDLLRRTYEDYKKSNVLINYYDAY
                     LKPLPYEVLNTPEDERKENIRTNPYYKDIPYLNGGLFRSNNVPNELSPTIKDNEIIGE
                     VINFLERYKFTLSTSEGSEEVELNPDILGYVYEKLINILAEKGQKGLGAYYTPDEITS
                     YIAKNTIEPIVVERFKEIIKNWKINDINFSTLDEILNEDSKIAENKHILRAFLDELDK
                     IRILDPAVGSGHFLISALKELLQIKKRIYTLLREEMDIYKEKLGIILNNLYGVDIDDI
                     AVEIAKLRLWLALIENLDVEALKRGEVLLPNIEYNVRCGNSLVGWIDENLKQLSISYL
                     CDNVRIMCVLEGLLINAHNSEERKKLKKAKELLEERDGYVLDNYVEAYHLLYEVIRTS
                     HGLKANLLKELLDEIRDSIYESVTPAIYAEIYQNGNKKNNGKKSKKNRPRVEEFEKL
                     KPFHWKIDFGWIIKEEGFDVIIGNPPYGNLLSPTEKEIMKRRDTPEFDIFVTFIVHSS
                     KLLKNEGYLGFIIPSSFGTGVRYSNLRKELFTKMCLKKLIYLPFDVFSGAYVDNCIII
                     LHKKPPKSEDLVLIYAFPKKTKKISPEFKNDLFIEYSKILNDPKCRIFPKSPEIYIIL
```



```
                     DKIKQNCRESLTYLEDLTESTIGILASKYKFSDKKENEYYLPYLEGNVYRYETKLKLK
                     NYVDFSKHKNNEKLINLFMSPEKIFIRRIVNRQDRIMASYGNIEGVVKKDLFVPVLRP
                     DTPINYFFYLLGILNSELISYIYIGKSAIALKDDFRQTTLEELRELPIVIPKNKNIINA
                     LTQLSLKLRPELNDKLNENDRIPLSENIDSLVYGIYPQDLIPKEEELNEICNEINGIIKK
                     YDEKSIDCLKYCQNIIKLLNTINTNELVRKIRNKN"
     gene            complement(4848. .5174)
                     /gene="MJECS03"
     CDS             complement(4848. .5174)
                     /gene="MJECS03"
                     /note="identified by GeneMark; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="M. jannaschii predicted coding region MJECS03"
                     /db_xref="PID:g1522637"
                     /translation="MKIIFYASREEQGFYGEAEIEEVEFFENPMKILEKYKNNLFLTE
                     EEFKKYIEDSNKKWGYGKKKKKPWIVIILKNIRKYPKVVKPKRFIPVCGKYVKEDEYE
                     QILKKL"
     gene            complement(5456. .5863)
                     /gene="MJECS04"
     CDS             complement(5456. .5863)
                     /gene="MJECS04"
                     /note="identified by GeneMark; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="M. jannaschii predicted coding region MJECS04"
                     /db_xref="PID:g1500692"
                     /translation="MLKKAIIKLLGIDNYIEKIEELEGERKEIFENPEKLNKKLDNPE
                     NELNFIKKEIERIQLEFNNVINSNNNDYVSKKELNDIVNQLNTLSTLVNGLIINISHG
                     NNLKESSNTVDDVKNRLLELLQSEKDYCITEIN"
     gene            6383. .6817
                     /gene="MJECS05"
     CDS             6383. .6817
                     /gene="MJECS05"
                     /note="identified by GeneMark; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="M. jannaschii predicted coding region MJECS05"
                     /db_xref="PID:g1522638"
                     /translation="MESKEYRKLEYNYKAFLIFSKVAMLTFLTVGIGAIFTPQTYPIM
                     PTIGFIVVAGIVSLIGMTIGALIIHQQYETLPANEKLEFKQKLLPEAYYICIELFGYG
                     SLVLLYNTFTSNNPTLCVMSLLMAGLFILVVLVIWYFGYKSY"
     gene            complement(7006. .7428)
                     /gene="MJECS06"
     CDS             complement(7006. .7428)
                     /gene="MJECS06"
                     /note="identified by GeneMark; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="M. jannaschii predicted coding region MJECS06"
                     /db_xref="PID:g1500694"
                     /translation="MNIDDYIEKLEIQKEGFFYKCPYCNYTNADVKAIKKHIKSKHYD
```

```
                     IIAKEVENLNKQNKPQRKPMKKQPKKKDDDYKDYMLLFAHKKKCKIYLDNGMIVEGTV
                     KAKDRFNIMVLDAKVDDKEVERIIIQKGHIVALIPLEE"
     gene            complement(7425. .8750)
                     /gene="MJECS07"
     CDS             complement(7425. .8750)
                     /gene="MJECS07"
                     /note="identified by GeneMark; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="M. jannaschii predicted coding region MJECS07"
                     /db_xref="PID:g1522639"
                     /translation="METTIMVVAMDRLTYYKLTSRYLYYFCLDRLLSCHRLDLDRDRK
                     INWRHPKYWNYNNIPKGEKPYLRHIDLWDYRGNFKDWIIENLRSIYCSPYYFESPRYE
                     LEYPAIDYNYWDWVIDIDFKGNFKKSKEIAKEVREIFKDYGVKDIHIKFSGSKGFHIW
                     INARYLPNEILNKGYVEGSTILTNFLNFKLRKLLDFVGKEGDTGIDLSIYKENMTVTA
                     PFSLIYADIEGKTPISIPFPINKVEKFKPIYLQNVKPYFIRKLTKYYENFGEIRGDLT
                     DFVDDALRYYQATKKVNNDIKIDLEINTNKKAKPKVEHSKREYIKTTIGDKEVVLNKV
                     KLESYLNELINSNWEDGKMRGCTYLTSLCKLLGYGRDKTIKLLNRWASKYGNKYIKHI
                     DYEVRYLYDRNGKVCSIKWLKENIPEAKTQIEIYKKYYELDKNNYLVKK"
     gene            complement(8735. .11929)
                     /gene="MJECS08"
     CDS             complement(8735. .11929)
                     /gene="MJECS08"
...
Note: remainder of annotations omitted.

  Query Match          10.3%;  Score 18;  DB 16;  Length 16550;
  Best Local Similarity 100.0%;  Pred. No. 1.66e+01;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    5050 AAAAATAAATTATTTTTA 5067
           ||||||||||||||||||
Cp      56 AAAAATAAATTATTTTTA 39


RESULT   36
LOCUS       CET01C3     23210 bp    DNA             INV       21-SEP-1998
DEFINITION  Caenorhabditis elegans cosmid T01C3, complete sequence.
ACCESSION   Z78413
NID         g1495011
KEYWORDS    HTG; Fibrillarin; Glutamate receptor subunit epsilon; Lipase;
            Ribosomal protein S16; Zinc finger protein.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 23210)
  AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
```



```
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
  TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
  JOURNAL   Nature 368 (6466), 32-38 (1994)
  MEDLINE   94150718
REFERENCE   2  (bases 1 to 23210)
  AUTHORS   Wild,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-NOV-1996) Nematode Sequencing Project, Sanger Centre,
            Hinxton, Cambridge CB10 1RQ, England and Department of Genetics,
            Washington University, St. Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     Coding sequences below are predicted from computer analysis, using
            predictions from Genefinder (P. Green, U. Washington), and other
            available information.
            For a graphical representation of this sequence and its analysis
            see:-
            http://webace.sanger.ac.uk/cgi-
            bin/display?db=wormace&class=Sequence &object=T01C3
            Current sequence finishing criteria for the C. elegans genome
            sequencing consortium are that all bases are either sequenced
            unambiguously on both strands, or on a single strand with both a
            dye primer and dye terminator reaction, from distinct subclones.
            Exceptions are indicated by an explicit note.
            IMPORTANT:  This sequence is NOT necessarily the entire insert of
            the specified clone.  It may be shorter because we only sequence
            overlapping sections once, or longer because we arrange for a small
            overlap between neighbouring submissions.
            IMPORTANT: This sequence is not the entire insert of clone T01C3.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we arrange for a small overlap between
            neighbouring submissions.
            The true left end of clone T01C3 is at 1 in this sequence. The true
            right end of clone T01C3 is at 10399 in
            sequence Z81061.
            The true left end of clone F14H8 is at 23107 in this sequence. The
            true right end of clone F02D8 is at 12022 in this sequence. The
            start of this sequence (1. .104) overlaps with the end of sequence
            Z78411.
            The end of this sequence (23107. .23210) overlaps with the start of
            sequence Z81061.
FEATURES            Location/Qualifiers
     source         1. .23210
                    /organism="Caenorhabditis elegans"
                    /db_xref="taxon:6239"
```



```
                    /chromosome="V"
                    /clone="T01C3"
     gene           complement(1134. .4299)
                    /gene="T01C3.1"
     CDS            complement(join(1134. .1356,1408. .1710,2137. .2878,
                    3228. .3686,3740. .3934,3986. .4130,4183. .4299))
                    /gene="T01C3.1"
                    /note="Weak similarity to WD repeat protein (TR:G1163005);
                    cDNA EST yk296c6.3 comes from this gene; cDNA EST
                    yk296c6.5 comes from this gene; cDNA EST yk396a6.3 comes
                    from this gene; cDNA EST yk396a6.5 comes from this gene;
                    cDNA EST yk484e5.5 comes from this gene"
                    /codon_start=1
                    /db_xref="PID:e258716"
                    /db_xref="PID:g1495018"
                    /db_xref="SPTREMBL:Q22059"
                    /translation="MDFLHSTKARKRYLRYPSSIYEELESKHYSMPAGETDDHDTWVT
                    ARFSPHLNQEHILYMGDDPGNIGIFDVRKPQDRSVPLEERQLYFFPAHDGAIMDVVGV
                    PQKESQIVSISGDSTIRCWDLNQSTLDRKSQVFFGHEGSVRSICFAPDDPNVFVTGGR
                    DFQVKIWDMNVSTVKKMEEDCRMATITYKTAHPKPSKVLTSGTPKSKAKAKTIEGYKV
                    TSVLFLDEHHVASASENADSGIRVWDIRKPTRNGESQPARILKVPTSNKKSYGVTCLT
                    LDRFGNRLFASCTDSSIFEYSVPSESVSPKFLVSHQLRKSQTNKSITFLWHSMSCASTT
                    LPTHVLDQYTKNLELSVNSYTGATIHNFYTQVACSPVSDVIACGSEDSRAVVWDLQDQ
                    YNYMNDRKLPDDIDKRRTKLPRPSCDGHLKQVLNVGWSSRGTYFMSCDEGGVRIWSEP
                    RNRCTWKLNDEDDTSYPTTSQELGLSYEKIKKPELKESDEAMSCFDSISLSPRQRADS
                    SGLSGSPQKNRGSKRPIFESPLKSICTNSPKPLRLNRSPRAKMSKLSFSPPSPLQPTN
                    SNNQDLVGYRTPRQIRNKKKKNNPFYNEHPTEGLPNFVYDTFVKKLIGESSKSDIEDS
                    GKSLSKTGQKRIEDWWQTKGENVATVTRARLPSVSEFGESACSKVITEDERIALHSPR
                    KLVLKSSSTQSPCPSNSKPKPIPMTPRKPMDKRPTSRNLLHYFKK"
     gene           4575. .4960
                    /gene="T01C3.2"
     CDS            join(4575. .4658,4823. .4960)
                    /gene="T01C3.2"
                    /note="predicted using Genefinder"
                    /codon_start=1
                    /db_xref="PID:e258714"
                    /db_xref="PID:g1495015"
                    /db_xref="SPTREMBL:Q22056"
                    /translation="MNDDLIMRKKPEQIEMEFQDSGESALVKKARDDVSAKEVFDAVK
                    DMARREIPQEAKKKLYDQVLEFVTTTNKK"
     gene           complement(5272. .5975)
                    /gene="T01C3.3"
     CDS            complement(join(5272. .5429,5525. .5975))
                    /gene="T01C3.3"
                    /note="predicted using Genefinder; similar to Zinc finger,
                    C3HC4 type (RING finger); cDNA EST EMBL:D72550 comes from
                    this gene; cDNA EST EMBL:D75450 comes from this gene; cDNA
                    EST EMBL:D69007 comes from this gene; cDNA EST yk411d4.3
                    comes from this gene; cDNA EST yk411d4.5 comes from this
                    gene"
                    /codon_start=1
```



```
                    /db_xref="PID:e258713"
                    /db_xref="PID:g1495014"
                    /db_xref="SPTREMBL:Q22055"
                    /translation="MRMMEAEIKDQRNNLQALKNSQRLSVRGSIQSNMSSRTDGILQR
                    RLDETERKLAKTSAELKAKDEKLKKETASLEASREAHRLLQEESNKSKVSVMRLTFKL
                    NRITHESVKEQAVLKKKLLDCETRLATYSECLVCYQKFDENTRIPRVMDCGHTLCDFC
                    INQIVKMAGCYSATCPFDRVRIFGFGKSRRLEDRPCNRFIMK"
     gene           7318. .8440
                    /gene="T01C3.4"
     CDS            join(7318. .7424,7473. .8179,8265. .8440)
                    /gene="T01C3.4"
                    /note="similar to lipase"
                    /codon_start=1
                    /db_xref="PID:e258743"
                    /db_xref="PID:g1495019"
                    /db_xref="SPTREMBL:Q22060"
                    /translation="MLRRFVLLGLVCLSIVNAASDVAGPLTDDFQSWLIANGYSSYDF
                    VRADYGTQGSYGGKTANAPKAVNTPVVFIHGNSDAALKVSSAATGWSNSIEYFLEHGY
                    TVAELYATSWQDTNALHAASRTHNCKDLVRLRKFLEAVLAYTAAPKISVVTHSMGVTL
                    ARKIIKGGSVSASDGSCDLGLPLNKKVEVMIGISGANYGLCNCEGGSATLEKTCNRVN
                    GLWPGDSCGLNYLDCGLTPLPWPCSGVNYSSFLMQLNADKNKEGDYVPSMWSLSDDLI
                    EYGTQVWGRSTCFIPTSDGKVVYHVYTHMETKELTYVDQYQMVKYRTIPTPTA"
     gene           9978. .11763
                    /gene="T01C3.5"
     CDS            join(9978. .10152,10480. .10682,11105. .11163,11207. .11497,
                    11604. .11763)
                    /gene="T01C3.5"
                    /note="predicted using Genefinder"
                    /codon_start=1
                    /db_xref="PID:e258742"
                    /db_xref="PID:g1495016"
                    /db_xref="SPTREMBL:Q22057"
                    /translation="MQWNLAQLIKLLLVASFFVPSVNGDNENKCVPDDIATLPDDCLT
                    IEGKLLITGEDSEMDKVRNLKGVRRIESGLEVDGTDLETPDLLPQVEQILNPDGPA
                    IIFKNNGKLQRILMPKLLALNGKDEDVQFENDNFPSAAYTEVTSSRDLLILESKSRAS
                    HSRKECSAGFFQFKTEESSGMNAIVIGLIIACVIAFLAIIGGGYFMYKHWNDNDDSDT
                    SKRSNEASEKLPPTAGSKMEDPKMKTFSERKYFKSQFMYHLNLRNPFSLRTICPQIEK
                    YDTWFRHKNIYQVQNCVRL"
     gene           complement(13446. .13924)
                    /gene="T01C3.6"
     CDS            complement(join(13446. .13730,13775. .13924))
                    /gene="T01C3.6"
                    /note="predicted using Genefinder; Similarity to Human 40S
                    ribosomal protein S16 (SW:RS16_HUMAN); cDNA EST
                    EMBL:D72100 comes from this gene; cDNA EST EMBL:D72365
                    comes from this gene; cDNA EST EMBL:D72968 comes from this
                    gene; cDNA EST EMBL:D74899 comes from this gene; cDNA EST
                    EMBL:D75213 comes from this gene; cDNA EST EMBL:D75934
                    comes from this gene; cDNA EST EMBL:D76065 comes from this
                    gene; cDNA EST EMBL:D66030 comes from this gene; cDNA EST
                    EMBL:D69662 comes from this gene; cDNA EST EMBL:C07459
```

comes from this gene; cDNA EST EMBL:C08827 comes from this gene; cDNA EST yk488g3.3 comes from this gene; cDNA EST yk488g3.5 comes from this gene; cDNA EST yk484h2.3 comes from this gene; cDNA EST yk484h2.5 comes from this gene; cDNA EST yk484g2.3 comes from this gene; cDNA EST yk484g2.5 comes from this gene; cDNA EST yk483f12.3 comes from this gene; cDNA EST yk483f12.5 comes from this gene; cDNA EST yk410d6.3 comes from this gene; cDNA EST yk410d6.5 comes from this gene; cDNA EST yk385g4.3 comes from this gene; cDNA EST yk385g4.5 comes from this gene; cDNA EST yk460a3.3 comes from this gene; cDNA EST yk460a3.5 comes from this gene; cDNA EST yk459g5.3 comes from this gene; cDNA EST yk459g5.5 comes from this gene; cDNA EST yk471c12.3 comes from this gene; cDNA EST yk471c12.5 comes from this gene; cDNA EST yk470g6.3 comes from this gene; cDNA EST yk470g6.5 comes from this gene; cDNA EST yk468d6.3 comes from this gene; cDNA EST yk468d6.5 comes from this gene; cDNA EST yk458f12.3 comes from this gene; cDNA EST yk458f12.5 comes from this gene; cDNA EST yk456g5.3 comes from this gene; cDNA EST yk456g5.5 comes from this gene; cDNA EST yk454h6.3 comes from this gene; cDNA EST yk454h6.5 comes from this gene; cDNA EST yk454a7.3 comes from this gene; cDNA EST yk454a7.5 comes from this gene; cDNA EST yk427h12.3 comes from this gene; cDNA EST yk427h12.5 comes from this gene; cDNA EST yk399g4.3 comes from this gene; cDNA EST yk399g4.5 comes from this gene; cDNA EST yk392d7.3 comes from this gene; cDNA EST yk392d7.5 comes from this gene; cDNA EST yk337a12.3 comes from this gene; cDNA EST yk337a12.5 comes from this gene; cDNA EST yk336g10.3 comes from this gene; cDNA EST yk336g10.5 comes from this gene; cDNA EST yk292e11.3 comes from this gene; cDNA EST

....

Note: remainder of annotations omitted.

Query Match 10.3%; Score 18; DB 18; Length 23210;
Best Local Similarity 100.0%; Pred. No. 1.66e+01;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Db  11515 AAAATAAATTATTTTTAA 11532
          ||||||||||||||||||
Cp      55 AAAATAAATTATTTTTAA 38
```

RESULT 37
LOCUS CET07F10 26969 bp DNA INV 22-SEP-1998
DEFINITION Caenorhabditis elegans cosmid T07F10, complete sequence.
ACCESSION Z77669
NID g1469130
KETWORDS HTG; Aminopeptidase N; Human mRNA (KIAA0033) product like protein; Polyadenylate-binding protein.

SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 26969)
AUTHORS Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M., Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J., Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A., Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M., Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N., Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B., O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A., Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E., Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M., Vaughan,K., Waterston,R., Watson,A., Weinstock,L., Wilkinson-Sproat,J. and Wohldman,P.
TITLE 2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans
JOURNAL Nature 368 (6466), 32-38 (1994)
MEDLINE 94150718
REFERENCE 2 (bases 1 to 26969)
AUTHORS Smye,R.
TITLE Direct Submission
JOURNAL Submitted (06-NOV-1996) Nematode Sequencing Project, Sanger Centre, Hinxton, Cambridge CB10 1RQ, England and Department of Genetics, Washington University, St. Louis, MO 63110, USA. E-mail: jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT Coding sequences below are predicted from computer analysis, using predictions from Genefinder (P. Green, U. Washington), and other available information.
For a graphical representation of this sequence and its analysis see:-
http://webace.sanger.ac.uk/cgi-bin/display?db=wormace&class=Sequence &object=T07F10
Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.
IMPORTANT: This sequence is not the entire insert of clone T07F10. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.
The true left end of clone T07F10 is at 1 in this sequence. The true right end of clone T07F10 is at 10934 in sequence Z75548.
The true left end of clone R90 is at 17787 in this sequence. The

true right end of clone C44H9 is at 2700 in this sequence. The start of this sequence (1. .109) overlaps with the end of sequence Z75529.
The end of this sequence (26845. .26969) overlaps with the start of sequence Z75548.

FEATURES Location/Qualifiers
source 1. .26969
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="V"
/clone="T07F10"
gene 3970. .7962
/gene="T07F10.1"
CDS join(3970. .4115,4374. .4551,4748. .4960,5013. .5366, 5413. .5597,5766. .5973,6027. .6493,6553. .6754,6801. .7099, 7147. .7494,7540. .7686,7734. .7962)
/gene="T07F10.1"
/note="Similarity to Human aminopeptidase N (SW:AMPN_HUMAN); cDNA EST EMBL:D36412 comes from this gene; cDNA EST EMBL:D37688 comes from this gene; cDNA EST EMBL:D34550 comes from this gene; cDNA EST EMBL:D33568 comes from this gene; cDNA EST EMBL:D33736 comes from this gene; cDNA EST EMBL:D36630 comes from this gene; cDNA EST EMBL:C12942 comes from this gene; cDNA EST EMBL:C13604 comes from this gene; cDNA EST EMBL:C10949 comes from this gene; cDNA EST EMBL:C11447 comes from this gene; cDNA EST yk356h9.3 comes from this gene; cDNA EST yk356h9.5 comes from this gene; cDNA EST yk320e6.3 comes from this gene; cDNA EST yk320e6.5 comes from this gene; cDNA EST yk194b8.3 comes from this gene; cDNA EST yk194b8.5 comes from this gene; cDNA EST yk321f6.3 comes from this gene; cDNA EST yk321f6.5 comes from this gene; cDNA EST yk260d10.3 comes from this gene; cDNA EST yk260d10.5 comes from this gene"
/codon_start=1
/db_xref="PID:e255946"
/db_xref="PID:g1469132"
/db_xref="SPTREMBL:Q22317"
/translation="MGHPPPKETKKGCQVKSFVLPVFCTLLGVLITAFITWHITKSQY SRNDPHDEIESENVPAAKPEDNISASELRLPTSVSPISTQLTVKTYLPGYGYTADKNN LTPEGQVLIELNITKSIKKVSLNSKDLNYTEEFIKKSSILVNGKSIAFTLDDKQSTHE KIFFNLDETVEPTTSATLKVAFGAPLRTDMSGLYQTTYTNSKGESKMAAVTQMEPVYA RRMVPCFDEPAYKATWTVTVIHPNKTVAVSNGIEDKVEDGQPGFIISTFKPTPRMSSY LLAIFISEFEYNEATTKSGVRFRVWSRPEEKNSTHYAVEAGVKCLEYYEKYYNISFPL PKQDMVALPDFSAGAMENWGLITYRESALLFDPRIYSGSQKRRVAVVIAHELAHQWFG NLVTLKWWNDLWLNEGFATLVEYLGTDEISDGNMRMEWFTNDALWSALAADSVASTH PLTPKIDKAMEVLDSFDSVTYDKGGAVLAMVRKTIGEENFNTGINHYLTRHQFDNADA GNLLTALGEKIPDSVMGPKGVKLNISEFMDPWTKQLGYPLLNASRINNTHIIVEQSRP KLLATGKEEEEKYSNPVWGFKWDVPVWYQVVGSSELEMKWMKRNEPLIIKSDKNVIINA ESNGFYRAGYSSGLWKEISEMLKENHEQFSPQTRVRLIDDSFALARAGLLSYSIPLNL ITYLKNEKEYLPWSGAIAKIRELIDMYGSNPEKDIVNKPMIALAENAPARRSIDFVSK

NYLDEKKFYEVGAAQQIILNSCGFGDSVCQADMVKMFTEEVLAKCDATRILSECSQYV PNLDFQNTIQKLRIPAPFRAESYCEAVRNGNSDTFEKVFHWYKTERNQVEKLNLMTAL TCSRDILTLKKLLLDAMKPEGSSFRLQDCAALFAKISSNDATTDAMLNFLIDRWEDMQ KRLATDHSGPSRVLSSIVNTLKTRGGLDQLRKFRKKAPKASEPGLDKMEESAEVTVTW RETNLKFVTKIIEEISKSL"
gene 8085. .10530
/gene="T07F10.2"
CDS join(8085. .8239,8402. .8594,9913. .9997,10048. .10185, 10322. .10431,10483. .10530)
/gene="T07F10.2"
/note="cDNA EST EMBL:C12942 comes from this gene; cDNA EST yk386g12.3 comes from this gene"
/codon_start=1
/db_xref="PID:e255949"
/db_xref="PID:g1469135"
/db_xref="SPTREMBL:Q22320"
/translation="MYILSHTCFRNNSVSLSLNLCFLFLKNQHFNCFHSRIVVFIRLF LIQVKISRTTHKQPLQHLVHMLHSFVAEHNQSPGESTCMESVMAFVKNGATASLAQWQ SVCLVNRSTGKAPQDFWKNNFIKIIPIRGDPYEIFDISGKFLAQITFPTQRDFCLFPT TSNKSEKEPQKKLQTQMATTFRTNVARTLYPPSTLTRTIGMARKWLEERKERNAISTG WNDVDDDHSKIRGNLPTNRVPDII"
gene 10714. .10985
/gene="T07F10.5"
CDS join(10714. .10820,10913. .10985)
/gene="T07F10.5"
/codon_start=1
/db_xref="PID:e255948"
/db_xref="PID:g1469134"
/db_xref="SPTREMBL:Q22319"
/translation="MRNDVKLLELSPIDAHYLRYRPSDADQNHRINNSTKFICSFPMF NKHNFWQFFSFCINI"
gene complement(12732. .14591)
/gene="T07F10.3"
CDS complement(join(12732. .12901,13297. .13459,13543. .13870, 13918. .14054,14100. .14150,14203. .14472,14532. .14591))
/gene="T07F10.3"
/note="Similarity to S.pombe polyadenylate-binding protein (SW:PABP_SCHPO); cDNA EST EMBL:T00632 comes from this gene; cDNA EST EMBL:T00633 comes from this gene; cDNA EST yk500c11.3 comes from this gene"
/codon_start=1
/db_xref="PID:e255947"
/db_xref="PID:g1469133"
/db_xref="SPTREMBL:Q22318"
/translation="MAEEVRAFSLPSRRVVLQDEKMFQDEEHSSPTRSTSSNSDEPTE TIESGDQFDVNEPIPTIETLEPVIISAASVQQLPAISLSGNNAQRTLWIGDVPTDWTE ETLSQVFTECGHAPYKVKRVYVKDELKGYCFIEFITFDEARQTLYDLNGNRIPGFKDI RFNLCFANDSYNPNSEFNLHVSSVPDDMADAELYRIFDKYQSCRGAKMFRFVDGSSKG SGFVRFGNQTDQQMALVEMHRTKVGGCRIILKLAGSRGERLDRGESRQFKSQRNDRRR DEKSGQYRNGTAVPQQTQVVDAYGNPVVVPVDKNIYGTDHPWGPFPDDPDLDSIIGLD LLTTKYRPQIANKRLMIDTDQFMLDLESSRWSSIAFNTNIKKEDFEKRITDNPWTAVP

```
                     .
     gene            23144. .25921
                     /gene="T07F10.4"
     CDS             join(23144. .23237,23733. .23893,24244. .24310,24499. .24593,
                     24644. .24939,24987. .25074,25865. .25921)
                     /gene="T07F10.4"
                     /note="predicted using Genefinder; Similarity to Human
                     mRNA (KIAA0033) (TR:G436224)"
                     /codon_start=1
                     /db_xref="PID:e255945"
                     /db_xref="PID:g1469131"
                     /db_xref="SPTREMBL:Q22316"
                     /translation="MTRLFILPAIPGISSLSLWYMICSAPGWPESEGGVFEIPKQFDN
                     FTVLADKFRAYKEDHFGYITTLFICAYLYKQTFAIPGSFLLNVIAGVVYDLWSGFILC
                     CCLTTLGSTLCYMPSELFGREIVFYIYPGQKLTYLQQKIDDNSNRLLPFLLFARMPPIS
                     PSWLLNIVAPFLNIPLPIFVVSALFGIHFLHFSSFFLSLNFASFYKYSLGLGLAPYNF
                     ICVQAGYILRDLRSWDDVFSTTTMLKLFSFALIPLAYAIYIRPRANKHQLVSSSDEMD
                     EKVKKMEIV"
BASE COUNT     8363 a    4799 c    4566 g    9241 t
ORIGIN

  Query Match           10.3%;  Score 18;  DB 18;  Length 26969;
  Best Local Similarity 100.0%;  Pred. No. 1.66e+01;
  Matches    18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Db  13889 AAATAAATTATTTTTAAC 13906
          ||||||||||||||||||
Cp      54 AAATAAATTATTTTTAAC 37


RESULT   38
LOCUS       SPCC553      31975 bp    DNA                 PLN       27-MAY-1998
DEFINITION  S.pombe chromosome III cosmid c553.
ACCESSION   AL023704
NID         g3169057
KEYWORDS    cyclophillin related; DNA polymerase alpha/primase-associated
            subunit; DNA-damage-inducible protein; peptidyl prolyl cis/trans
            isomerase; peroxisome biosynthesis; possible pseudogene; tf2 type
            LTR; transcription factor IIa small subunit; translocation
            elongation factor; Tu family.
SOURCE      fission yeast.
  ORGANISM  Schizosaccharomyces pombe
            Eukaryota; Fungi; Ascomycota; Archiascomycetes;
            Schizosaccharomycetales; Schizosaccharomycetaceae;
            Schizosaccharomyces.
REFERENCE   1  (bases 1 to 31975)
  AUTHORS   Wood,V., Rajandream,M.A., Barrell,B.G., Murphy,L. and Harris,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAY-1998) European Schizosaccharomyces genome
            sequencing project, Sanger Centre, The Wellcome Trust Genome
            Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk
```



```
  COMMENT     Notes:
              Details of yeast sequencing at the Sanger Centre are available on
              the World Wide Web.
              (URL, http://www.sanger.ac.uk/Projects/S_pombe/)
              During 1995 to 1996 about 80% of S. pombe chromosome 1 was
              sequenced by the Sanger Centre.  The sequencing of the S. pombe
              genome is now being continued with funding from The European
              Commission. Fourteen European sequencing laboratories, including
              the Sanger Centre, are participating in the project.
              Protein coding regions (CDS) have been predicted with the help of
              computer analysis using the Genefinder program in PomBase (an ACEDB
              database) with additional predictions for the branch-acceptor sites
              supplied by the program Sp3splice. CAUTION: It is possible that for
              any individual CDS we may have underestimated or overestimated the
              number of introns/exons or we may not have chosen the correct
              splice donor/acceptor sites. CDS are numbered using the following
              system eg SPBC25H2.01c. SP (S. pombe), B (chromosome 2), c25H2
              (cosmid name), .01 (first CDS), c (complementary strand).
              The more significant matches with motifs in the PROSITE database
              are also included but some of these may be fortuitous. The length
              in codons is given for each CDS.
              IMPORTANT: This sequence MAY NOT be the entire insert of the
              sequenced clone.  It may be shorter because we only sequence
              overlapping sections once, or longer, because we arrange for a
              small overlap between neighbouring submissions. Cosmid c553 is
              overlapped at the 5' end by cosmid 794 and at the 3' end by cosmid
              736.
  FEATURES            Location/Qualifiers
       source         1. .31975
                      /organism="Schizosaccharomyces pombe"
                      /strain="972h-"
                      /db_xref="taxon:4896"
                      /chromosome="III"
                      /clone="cosmid c553"
                      /map="IIIL"
       misc_feature   complement(1. .181)
                      /gene="SPCC553.01c"
                      /note="nominal overlap with S. pombe chromosome 3 cosmid
                      c736"
       gene           complement(1. .1171)
                      /gene="SPCC553.01c"
       CDS            complement(<1. .1171)
                      /gene="SPCC553.01c"
                      /note="SPCC553.01c, partial; unknown, len:<390aa, low
                      similar eg. to YGR042W, YG1T_YEAST, P53227, hypothetical
                      protein, (271aa), fasta scores, opt:138, E():0.0027,
                      (26.5%identity in 204 aa overlap)"
                      /codon_start=1
                      /product="hypothetical protein"
                      /db_xref="PID:e1294515"
                      /db_xref="PID:g3169058"
```



```
                      /translation="MDTSSNVFHYQVLWTSDKAKKVKVWKDGTMRFHSFNNRGILYDS
                      DNRVVDDFFLNNKRIVLDSEIEFNRTIVYVENLQTTTEANIPPIPRKSKLSETGKRPA
                      YLLKPFQPPPFHSANSDVSNNSNGALSSNMLKTHAHHTNSTKSTIFRNANLQSYGVDDS
                      DIGSLRCSSKHQLSVTPEIQPKAKFRPNSYLASSVEESNRANNTSPTPPSNSSVTALP
                      KPSKIAYIPVLSKPKSPNGTLVEDPVEEFGEDDIASDFFSSPSADLEADDPTTLNNSN
                      AGHLLQNLSNDNVSVSVTALENGNEEVSNNHARSPHSSRACNTIPSNKNDGLNDDHQR
                      QDSSVAEPSPTSKRLRLTRLPLPLLRKPTGKRQNFINITVTDPSKNLYERSVKTAA"
       gene           3229. .5722
                      /gene="SPCC553.02"
       CDS            join(3229. .3257,3302. .3418,3619. .4628,4675. .5086,
                      5136. .5463,5516. .5722)
                      /gene="SPCC553.02"
                      /note="SPCC553.02, unknown, len:700aa, similar eg. to
                      YHR074W, YHM4_YEAST, P38795, hypothetical 80.7 kd protein;
                      weaksimilarity to B.subtilis spore outgrowth factor
                      B;(714aa),fasta scores, opt:2952, E():0, (60.4% identity
                      in705 aa overlap)"
                      /codon_start=1
                      /product="hypothetical protein"
                      /db_xref="PID:e1294516"
                      /db_xref="PID:g3169059"
                      /translation="MERYVTIASCQLNQWAMDFEGNRLRIIDSIKEAKRQNASLRVGP
                      ELEVTGYGCEDHFLESDTYYHSWEMLCSIIHDPDCQDILLDIGMPVMHKAMRHNCRIL
                      ALNGKILLIRPKIWLCDDGNPRESRWFTPWLRPRVVETHTYLPIFVARSLNQTTVPIGD
                      AILQCNETVVGVETCEELFTPNSPHIDMALDGVEIFINASGSHHELRKLTTRVNLIQN
                      ATEKCGGIYLYSNQRGCDGGRLYYDGSSMIFANGKMLAQGHQFSLKDVEVISATVDVD
                      TVRSYRFQPSHGIQGVTRPSYERIHVWFSLSSYQQDYDIYRKPTDPIEVTIPLPEEEI
                      TFGPACWLWDYLRRSHAAGFPLPLSGGLDSCSTAVLVYSMCRIVCKAMEEDDAQVLSD
                      VRRIVGDPSYSSTDPKKLLNHLFYTAFMGSEHSSKETRSRAKELSSLIGSYHTDVNID
                      TMTSAVVKLFALVTGKTPQFRSNGGTNAENLALQNIQARSRMLLGYLFAQLLPWVRGY
                      SGSLLVLGSSWVDECLRGYLTKYDCSSADINPIGGISKTDLKSFLRYAKEALDLPILQ
                      EFLDATPTAELEPTTESYVQSDEADMGMTYAELSVFGRLRKISKCGPYSMFTQLMHQW
                      GDRLSPSQVAEKVKRFFHYTGINRHRMTTLTPSYHAETYGVDDNRYDLRQFLYPSWTW
                      QNKKIDALASKFEQHQRK"
       misc_feature   3258. .3263
                      /gene="SPCC553.02"
                      /note="gtatgc, splice donor sequence"
       misc_feature   3289. .3301
                      /gene="SPCC553.02"
                      /note="ctaacttctgtag, splice branch and acceptor"
       misc_feature   3419. .3424
                      /gene="SPCC553.02"
                      /note="gtatgt, splice donor sequence"
       misc_feature   3603. .3618
                      /gene="SPCC553.02"
                      /note="ttaacaaacgcattag, splice branch and acceptor"
       misc_feature   4629. .4634
                      /gene="SPCC553.02"
                      /note="gtaagt, splice donor sequence"
       misc_feature   4654. .4674
                      /gene="SPCC553.02"
```



```
       misc_feature   /note="ctaattactgaccttcaatag, splice branch and acceptor"
                      5087. .5092
                      /gene="SPCC553.02"
       misc_feature   /note="gtatgc, splice donor sequence"
                      5120. .5135
                      /gene="SPCC553.02"
       misc_feature   /note="ctgacaaatttacaag, splice branch and acceptor"
                      5464. .5469
                      /gene="SPCC553.02"
       misc_feature   /note="gtaagt, splice donor sequence"
                      5503. .5515
                      /gene="SPCC553.02"
                      /note="ctaactttataag, splice branch and acceptor"
       gene           6500. .9358
                      /gene="SPCC553.03"
       CDS            join(6500. .6730,6776. .9358)
                      /gene="SPCC553.03"
                      /note="SPCC553.03, peroxisome biosynthesis protein,
                      len:937aa, similar eg. to P. pastoiris, PEX1_PICPA,
                      P46463, peroxisome biosynthesis protein, (1157aa), fasta
                      scores, opt:1005, E():0, (30.3% identity in 1161 aa
                      overlap), contains PS00017 ATP/GTP-binding site motif A
                      (P-loop) x2, also contains Pfam match to entry AAA
                      PF00004, ATPases associated with various cellular
                      activities (AAA) x2"
                      /codon_start=1
                      /product="peroxisome biosynthesis protein"
                      /db_xref="PID:e1294517"
                      /db_xref="PID:g3169060"
                      /translation="MRCIVSYKSLRSCLVNVPELLLESISEPVQNYAVQAVVCKNDIK
                      KTFYFGGIPSQFSFEIDSTYAHTLKLAENQEINLSIIDCTHEIEQLEIEPVTSNDW
                      EIAERNAAWLEENLLVQYRVATTPRFIILPSGTFIQFQPLKLIPSSLCGRLLRTTEV
                      LITPKPNTSAIEVKENRKVNLRCVVENRLLPDSVTADSPALCVFLPLNFPDPRPDVYM
                      DGGNLKSTIVCQCVSCPFQIPGHFFISKSLALSYSIKTGFKFQIWKAHNPPSSSKFIL
                      EQKGLPPESNLSSELVAAKLKNSYLMDGMTLKLVDIAVSYSVSGLSGVVKNPIQDIKI
                      TSDTNVPVNAGIRNNSPRLSMQPFPHEFAQVRNAVFLHQNIYINGPKGCGKSNLVHSL
                      FDYYSLNSIYFQMIVSCSEIDRSSFAKFQSFWNNVFIQAERYEPSIIYLDDVHCLISS
                      SNENGELGFVEEREIAFLQHQIINLKRKRKIIFIGFGEEFLTFSENLVLPLLFQIKIA
                      LPSLAVTRRKEILTTIFQENFSDITMDSIEPISVKTEGYLMTDLVLFVKRLLSEAFVE
                      KIQNGPTKLMNKGLIEKTLKDFVPLQLRKAKFVKSSIRWIDIAGMQEAKEAVRDIIES
                      PVKYSLIYKQCRLRLPTGILLFGYPGCGKTYLASAISSTFFVCPISIKGPELLDKYIG
                      KSEQGVRDLFSRAQMAKPCVLFFDEFDSVAPRRGQDSTGVTDRVVNQILTQMDGAESL
                      DGVYIVAATTRPDMIDPALLRPGRLDKLIFCDLPNEEERLEVLQKLANRPHIENAAML
                      KKLSTLTDGYTYADLSSLLYDAHLIAVHKLLKRVSINAVDPSQTTSSFTNLTTESKRN
                      ASMLALPPESRYNQNMQSMSDSKSVVIEDYMLMEALKKNSPSLNSEEPEHLSNLYRDF
                      RSKLFEPELNARNTDVGSKTRQI"
       misc_feature   6731. .6736
                      /gene="SPCC553.03"
                      /note="gtaaaa, splice donor sequence"
       misc_feature   6764. .6775
                      /gene="SPCC553.03"
```

```
                     /note="ctaatctcttag, splice branch and acceptor"
     misc_feature    7619. .8194
                     /gene="SPCC553.03"
                     /note="Pfam match to entry AAA PF00004, ATPases
                     associatedwith various cellular activities (AAA)"
     misc_feature    7679. .7702
                     /gene="SPCC553.03"
                     /note="PS00017 ATP/GTP-binding site motif A (P-loop)"
     misc_feature    8423. .8977
                     /gene="SPCC553.03"
                     /note="Pfam match to entry AAA PF00004, ATPases
                     associatedwith various cellular activities (AAA)"
     misc_feature    8483. .8506
                     /gene="SPCC553.03"
                     /note="PS00017 ATP/GTP-binding site motif A (P-loop)"
     gene            9483. .11370
                     /gene="SPCC553.04"
     CDS             join(9483. .9552,9608. .11370)
                     /gene="SPCC553.04"
                     /note="SPCC553.04, cyclophillin related; peptidyl prolyl
                     cis/trans isomerase len:611aa, similar eg. to H. sapiens,
                     Q15002, cyclophilin-related, (645aa), fasta scores,
                     opt:1478,E():0, (40.8% identity in 598 aa overlap),
                     contains PS00170 Cyclophilin-type peptidyl-prolyl
                     cis-trans isomerase signature, also contains Pfam match to
                     entry pro_isomerase PF00160, Peptidyl-prolyl cis-trans
                     isomerases"
...
Note: remainder of annotations omitted.

  Query Match           10.3%;  Score 18;  DB 24;  Length 31975;
  Best Local Similarity 100.0%;  Pred. No. 1.66e+01;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  31043 CACATTTCACATAAAAAT 31060
          ||||||||||||||||||
Qy    106 CACATTTCACATAAAAAT 123


RESULT   39
LOCUS       CEB0240     32346 bp    DNA             INV       21-SEP-1998
DEFINITION  Caenorhabditis elegans cosmid B0240, complete sequence.
ACCESSION   Z74026
NID         g1403144
KEYWORDS    NTG; 1-aminocyclopropane-1-carboxylate synthase; Erythrocyte plasma
            membrane glycoprotein like; Guanylate cyclase.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 32346)
```



```
  AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
  TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
  JOURNAL   Nature 368 (6466), 32-38 (1994)
  MEDLINE   94150718
REFERENCE   2  (bases 1 to 32346)
  AUTHORS   White,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-NOV-1996) Nematode Sequencing Project, Sanger Centre,
            Hinxton, Cambridge CB10 1RQ, England and Department of Genetics,
            Washington University, St. Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     Coding sequences below are predicted from computer analysis, using
            predictions from Genefinder (P. Green, U. Washington), and other
            available information.
            For a graphical representation of this sequence and its analysis
            see:-
            http://webace.sanger.ac.uk/cgi-
            bin/display?db=wormace&class=Sequence &object=B0240
            Current sequence finishing criteria for the C. elegans genome
            sequencing consortium are that all bases are either sequenced
            unambiguously on both strands, or on a single strand with both a
            dye primer and dye terminator reaction, from distinct subclones.
            Exceptions are indicated by an explicit note.
            IMPORTANT:  This sequence is NOT necessarily the entire insert of
            the specified clone.  It may be shorter because we only sequence
            overlapping sections once, or longer because we arrange for a small
            overlap between neighbouring submissions.
            IMPORTANT: This sequence is not the entire insert of clone B0240.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we arrange for a small overlap between
            neighbouring submissions.
            The true left end of clone B0240 is at 1 in this sequence. The true
            right end of clone B0240 is at 11447 in
            sequence Z72513.
            The true left end of clone T04F3 is at 32236 in this sequence. The
            true right end of clone W05E10 is at 1829 in this sequence. The
            start of this sequence (1. .106) overlaps with the end of sequence
            Z77670.
            The end of this sequence (32236. .32346) overlaps with the start of
            sequence Z72513.
```



```
FEATURES             Location/Qualifiers
     source          1. .32346
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /chromosome="V"
                     /clone="B0240"
     gene            complement(1862. .8842)
                     /gene="B0240.3"
     CDS             complement(join(1862. .2302,3161. .3424,3679. .3783,
                     3941. .4479,4529. .4708,5255. .5529,5584. .5681,5733. .5909,
                     6532. .6622,6668. .6742,6945. .7137,7410. .7692,8141. .8325,
                     8368. .8626,8674. .8768,8821. .8842))
                     /gene="B0240.3"
                     /note="similar to guanylate cyclase; cDNA EST EMBL:D34048
                     comes from this gene; cDNA EST EMBL:D37057 comes from this
                     gene; cDNA EST yk310b9.3 comes from this gene; cDNA EST
                     yk310b9.5 comes from this gene; cDNA EST yk323h9.5 comes
                     from this gene"
                     /codon_start=1
                     /db_xref="PID:e248941"
                     /db_xref="PID:g1403146"
                     /db_xref="SPTREMBL:Q17462"
                     /translation="MGPTCSSNFLRFANLTAEMQSLEINILKGYPVEHPTMIDMVTRS
                     PQNLAQNLVSLLHGFFNGQVGAVLCEECTEGDELASEIIFSTIEDIFENWNIALKETV
                     RIGKRENSVNISDAITIPEPSARVILLFLGNKLNDYTEFMTAMSMNNTTTEEYTPVIV
                     ISKNSLELTFPWKENDAIAELPDKAIIVTNNCTDKSKISSFLSSYSFSTIEETIISLQ
                     MYEGYHLLGYYLYTAITNTTLFNYVQPEKAISSMSIPGPFGEIPINSNGQRIAGIREL
                     IIWCRGYDVLVVDKSLNSNNFIMPLGTISTDKKCPDQACLNFVLNSTSSFEPLKDVPL
                     CGFHGEICDQTGVIIAIAVIMGVLLMFIIILTTIRKCCNGSKGRSISNPWVIPPQDVR
                     FIDLTNTEGSQHMSIQSLQRNMEEKQRLQSLARTKHIATVDQVYVLADKYVMRDKLRY
                     DKIDINLLYQMKSHLQHDNLNSFVGITIDKASHMYIIWNQCFRGSLHDHIFTKERQRG
                     TATRFEGLFLRDILKGLEYIHASAIDFHGNLTLHNCMLDSHWIVKLSGFGVNRLLVKW
                     KTSGQIPTEDHTPVIKSEELHYFDPAMKKIWKNYADRNERALITPQPGRKCDMYSPGV
                     ILLIKKKPNSQLDSPREEDDSVLIDDENDAIASRFPLPIIIPEGIEMHNDLIKH
                     LENCPGSVRPDIALARKIIDTVLKMSGSLVDLMIRNLTAYTQGLNETVKNRTAELEKE
                     QEKGDQLLMELLPKSVANDLKNGIAVDPEVYENATILIYSDIVGPTSLCSQSQPMEVVT
                     LLSGMYQRFDLIISQQGGTKVILNRIMETIGDAYCVAAGLPVVMEKDHVKSICMIALL
                     QRDCLHHFEIPHRPGTFLNCRWGFNSGPVFPAGVIGQKAPRYACFGEAVILASKMESSG
                     VEDRIQMTLASQQLLEENFPQFVCSNRGGRTIEGIGRILTTWLEGVNAGEQVKVVEPQ
                     NDLNDELSRIMKRDGELLAAATALKPKDKMTLAKEVIAERKNEEERLQRQQTLQEAL
                     EEHEEEIEMNEVLVDEDEGEGKPKEVDLTSIVSTQMEELEDEPAGRTIGHGRLDSQAS
                     TIPDN"
     gene            complement(11170. .13349)
                     /gene="B0240.1"
     CDS             complement(join(11170. .11232,11422. .11525,11576. .11709,
                     11909. .12042,12090. .12184,12232. .12441,12509. .12653,
                     12701. .12851,12898. .13028,13095. .13156,13205. .13349))
                     /gene="B0240.1"
                     /note="Similarity to Human erythrocyte plasma membrane
                     50KD glycoprotein (SW:RH50_HUMAN)"
                     /codon_start=1
```



```
                     /db_xref="PID:e257637"
                     /db_xref="PID:g1483241"
                     /db_xref="SPTREMBL:Q17463"
                     /translation="MWSVLHRRQFAIIAGLMQTVPIVLPAKYVKYIDPLDDSRRVYSG
                     TDYPLPQDVHLMIFVGPGFLMAFLKRYGFSAVSVNLLLSAFVIQFAMLLRGFMTVAPQ
                     ETGLFSIGIPEMISAESSCAAVLITMGVLLGRLTPVQFLLLAFFETGINVLVEHTVFN
                     YLHVNDSGRSLSVHTFGAIYFGLAAACVGHKKNVMEMDEHGGIHHSDLFSMIGTLLLWV
                     FFPSFNAAIQEPEDARHRAIMNTYLAMASGTVTTFMISSCVDTLGRFMMIHIQSSTLA
                     GGVAIGSSANAVLHPYHAVIVGVIAALLSVIGHAWISPRLERTFHLFDTCGVHNLHGM
                     PGILAGLLSIGFAYFYEPESYGKTLYHIYPYWIGGELHGDRENVSQAQYQALGLLTIL
                     VTAVIGGLLTGCILKIKVWNQVDDPDFPHGEMNYYAQSDVNFLSKYKHAQEQERLRER
                     EQMQEIY"
     gene            complement(14506. .20158)
                     /gene="B0240.2"
     CDS             complement(join(14506. .14572,14619. .14719,14823. .15497,
                     15556. .15813,15862. .16085,16130. .16234,16282. .16662,
                     17055. .17206,17643. .17764,17812. .18072,18139. .18176,
                     18294. .18450,18507. .18914,18959. .19148,19221. .19438,
                     19629. .19692,19740. .19844,19896. .20027,20073. .20158))
                     /gene="B0240.2"
                     /note="cDNA EST yk452h4.3 comes from this gene; cDNA EST
                     yk452h4.5 comes from this gene"
                     /codon_start=1
                     /db_xref="PID:e249061"
                     /db_xref="PID:g1403145"
                     /db_xref="SPTREMBL:Q17461"
                     /translation="MGLIGLWNVKRRANRAVRKFQFVVNNIRREKSGQNRVLRYSERL
                     KKWLSVLYPPWKKEKDDGIYDNLPEDGTRANSIVLLFRNILMIELPGIVMYFLLSFKY
                     SQYPEPIGTSISTIIHVFLIFTLIFPKGPAMNVIGNIHQVASGVACVQIDVMSSRNDV
                     EGNVMDKGAMLVSRFRALLQNVSGPMNKVKKMLLVLDEKMTKFVDIMRRHYRVIANLS
                     NQCRNMLKTPYTKCLDIFDDAYFYCKGKARFPGSHGDACEMIQKTAKICHQAKGFSTE
                     VCSLPSIIGKGVKGAAVPFYAAYFQAIEFAVKKIFTVHIEAAKWALKKVNPVIEEIKK
                     ANNIRAEYQSSDDQKPNAADDTTANLRASIKSLNGIVNVYIRIINIISFILRYCIMP
                     GLLIWPFITTLRFMYNYNYNDDPRNRFLTKEFTKIDQDCAFRGARKVMPLMANESKQV
                     GFSVVSLLSLCVCLIMTLSTVHTLMNRTNIDTPAHYELKVAGNSSMSDVMNEFLDVFS
                     PPTKSIRDRENRWRRCFNEPNPPNYTENMVMPLMPIAALFLCRLKAYPGRQTLVLADY
                     FFPNRVRIRALSLYNKILQSRRNLLTEINVRVCVHCSAFYCINCFSVTLPCVKCEEEM
                     QCINGIELYYEESNAIIDDSEMMGDSHSSSPTTTTDEHLTNQFSPGRRYQLSKDSYNEN
                     NYRKNDFPAASSSSSSSPNLRRSPNRTVSSPRVQQKPITIFDQIVDWFQAEISVRKRLA
                     GAACGYLSTIPFIVTVSILKLTIWAPFSSVQDSLAWWIYPNAWASIIFVGIASVAMSL
                     FSIIKFCKVDQLPRLAATDTFALAGVALEFVTRLTFVTFAPCVADFSFSREFAFVAIS
                     LAIAISSALVVFRSDYQLNFSHIQVNSVKTLIDPGTSLPYANISEICGIDAAISTTAA
                     VALILVVGPMVSGFSAWWLLLNIPFIVVLPGLCFTQQFTSKISWKIVNQIVMKPISFP
                     FPPPPYTVHSPTPSQTSTLPWVIETDDSLLKFPALHDLRTIAWNDERRVDVFSLSQPG
                     KHPRNWKAVSLPCVRMLDELCSRMTVSAARLVGYSWDDDHDIENEDVPRDALLMPRKMR
                     EMAYRGTGQSRQQKSMAPIRSHNTQTVGLLSKISNFLGPGVTEKLVISRFDAHMNAYA
                     AEALYMLVVDSMGEDRPGVVQKDLKDLITLLCKLIAAIDTYERAKASVADKSDVTFLR
                     IVDASLKSSLQRVVTTPGSHLSSLNLPEEHSRTIRMICLTDEL"
     CDS             join(25995. .27008,27060. .27296,27728. .32346,
                     Z72513:112. .1315,Z72513:1414. .1537,Z72513:1597. .1984,
                     Z72513:2055. .2221,Z72513:2267. .2613,Z72513:2662. .2791,
```

Z72513:3142. .3542,Z72513:3789. .4242,Z72513:4324. .4463,
Z72513:4762. .4980)
/gene="T04F3.1"
/note="similar to 1-aminocyclopropane-1-carboxylate
synthase; cDNA EST EMBL:D34239 comes from this gene; cDNA
EST EMBL:D35575 comes from this gene; cDNA EST EMBL:D64242
comes from this gene; cDNA EST EMBL:D67126 comes from this
gene; cDNA EST EMBL:D67854 comes from this gene; cDNA EST
EMBL:D68321 comes from this gene; cDNA EST EMBL:C12830
comes from this gene; cDNA EST yk212e9.5 comes from this
gene; cDNA EST yk240g7.5 comes from this gene; cDNA EST
yk244f9.5 comes from this gene; cDNA EST yk260b1.5 comes
from this gene; cDNA EST yk270f5.5 comes from this gene;
cDNA EST yk323c2.5 comes from this gene; cDNA EST
yk324d10.5 comes from this gene; cDNA EST yk353c9.5 comes
from this gene; cDNA EST yk420d11.5 comes from this gene"
/codon_start=1
/db_xref="PID:e1321666"
/db_xref="PID:g3641878"
/translation="MSPIAPPRKRLGREESLHYYRDDDEVIQEQSSTEKSRRVTVKSP
EPGNTSVSMNDNVEGRNLTAGITSISMNDSCTDTSYESRFSTIPRARSSPPGNHRVNH
IQVNVIGPIRVSEPESRHPVVGSVGEDVTMDIYIERRRSGSNPRRDSHVSPETLEAPP
KPPRRSKYDEHIRIGDPIATSTPMTTIERKQTREERVSTESGQSSRKSTRSSRPPDIN
IIECETFLKETRSPARTPATPIVVINQTPSSRRRSQDLLGEPGQRRRKSTVTDIDWYA
AFNMKVSWLNRLYRGLLTQFIRNSKHTVIILCHFRPSKFYIYKHVFDISKIVKIFKRR
KKKKDPSSPTKIARPSFASHNEERKASRDLRRRASDSSMIGIPPADIDLNQIPLCTVP
PEIKKDEKCQCNACRLIRMSEAERAETRSSTLPSRHRTAPVDIELEDIPNPKPFSHPS
ANSKPPTPPNRRRHPPSASVAHSQSSFVDESPSQHFVTTTIDRNQVTPVTTTTTNMRE
SGPLLKENSHNQLLSSSDNWVAQMMNDASEWENNVSSRKSSILSTSSTSARKASVARR
ISVDELTKPEKRKSRQAVPLPPSDVSIDDIFTALTPRKEEVQSERKTPIVTRKQRAQD
VDAIDPEKDRETPIAPPRSGKRKSDATENQPEITLELDVEKHEIDSSKVSTSTINLND
ESMETRNTNDSEDSFDDEVNPRNQRLVVEIPPSEPRVTSTATIQLESSDVAGENSENK
RPVISMRSKSEIAKKEKDAQRSGFVIIPHSKEHILDESNISMDDVFNTTPHDQCRVPD
IDADSSKHTSDSREVSTVTINLDNVFPTEEPKLVAKDNCEIAEEERIGKRIKQFER
TTGEQEISKNSEPTEDEMPDEKDHRTSAVSIDLDKVFVQGTAKKPENDEPDEKIKRGI
AEFERSKQEKEVQRSGVAETSHSGKHIPDESNISMDDVFNTSQKYKSDEKLSTPERTV
...

Note: remainder of annotations omitted.

Query Match 10.3%; Score 18; DB 18; Length 32346;
Best Local Similarity 100.0%; Pred. No. 1.66e+01;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 9701 AAAAATAAAACAAGTGAA 9718
||||||||||||||||||
Qy 144 AAAAATAAAACAAGTGAA 161

RESULT 40
LOCUS CELZC21 36087 bp DNA INV 22-OCT-1993
DEFINITION C. elegans cosmid ZC21.

ACCESSION L16685 L18807
NID g289729
KEYWORDS breakpoint cluster region protein; dual bar protein; transient receptor potential protein.
SOURCE Caenorhabditis elegans (strain Bristol N2) hermaphrodite mixed whole animal DNA.
ORGANISM Caenorhabditis elegans
Eukaryotae; mitochondrial eukaryotes; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea; Rhabditidae; Caenorhabditis.
REFERENCE 1 (bases 1 to 36087)
AUTHORS Sulston,J., Du,Z., Thomas,K., Wilson,R., Hillier,L., Staden,R., Halloran,N., Green,P., Thierry-Mieg,J., Qiu,L., Dear,S., Coulson,A., Craxton,M., Durbin,R., Berks,M., Metzstein,M., Hawkins,T., Ainscough,R. and Waterston,R.
TITLE The C.elegans genome sequencing project: A beginning
JOURNAL Nature 356, 37-41 (1992)
MEDLINE 92168156
REFERENCE 2 (bases 1 to 36087)
AUTHORS Du,Z. and Waterston,R.
TITLE Sequence of the C. elegans cosmid ZC21
JOURNAL Unpublished (1993)
COMMENT Submitted by:
Nematode Sequencing Project
Department of Genetics, Washington University,
St. Louis, MO 63110, USA, and
MRC Laboratory of Molecular Biology,
Cambridge CB2 2QH, England
e-mail: rw@nematode.wustl.edu and jes@cele.mrc-lmba.cam.ac.uk NOTE:
Coding sequences below are predicted from computer analysis, using the program Genefinder (P. Green and L. Hillier, in preparation)
Neighboring cosmid information:
This sequence is the entire sequence of C. elegans cosmid ZC21.
Bases 1 to 200 of this cosmid correspond to bases 10234 to 10433 of the C. elegans cosmid C04D8 entry (GenBank: L16687). Bases 1 to 200 of the GenBank entry for cosmid C02D5 (GenBank: CELC02D5) correspond to bases 35888 to 36087 of this entry.
The end of the actual sequence of the C04D8 cosmid is at 7298 of this entry.
The beginning of the actual C02D5 cosmid sequence is at 22573 of this entry.
The beginning of the actual K10C7 cosmid sequence is at 30071 of this entry.
FEATURES Location/Qualifiers
source 1. .36087
/organism="Caenorhabditis elegans"
/strain="Bristol N2"
/db_xref="taxon:6239"
/dev_stage="mixed"
/sex="hermaphrodite"
/tissue_type="whole animal"

exon 1540. .1713
/note="putative"
/number=1
CDS join(1540. .1713,1754. .1843)
/partial
/standard_name="ZC21.5 protein"
/note="putative"
/codon_start=1
/db_xref="PID:g289736"
/translation="IASLASSSCSTATTSEFLRSQHTLQQQQQQQTNQKHQQTVNELS
AGIVSRMFLFYPFLFSILRKNSLNFGRFLLFPEINYVPFLIKK"
exon 2360. .2456
/partial
/note="ZC21.8: orphan; putative"
/number=2
misc_feature 3399
/note="possible trans-spliced leader site for ZC21.4"
CDS join(3490. .4296,4696. .5541,5657. .5802,5854. .6064,
6551. .6742)
/standard_name="ZC21.4 protein"
/note="homology with breakpoint cluster region protein; putative"
/codon_start=1
/db_xref="PID:g289735"
/translation="MATTSSSFHHHSSQAGPSRDIENGEAPTATATTPKSGRKWKKSK
AAKQGSGGGSSGSSSGSQQQGAAGAPQPVLGVRIADCPTGSCEDHVPMIVQACVCVIE
TYGMDTVGIYRIPGNTAAVNALKESLSNRGFDSVDLSKVESLDPRWRDVNVVSSLLKM
FLRKLPEPLLTDKLYPFFIDANRISTHHNRLHKLRNLLRKLPRPHYDTLRFLIVHLSE
ITKHSDVNKMSCRNLALMFGPSIVRPSDDNMATMVTHMSDQCKIIETLIHYNLWMFDE
SSTTEDAVPEQHPADQGMPLEPGGYGVGVPTGVSAASFNDMHNLIRKANEDQAAAMMN
EGKGQKIKNMLRRNSRRDKSKSKLKIESTAPAAVNPRGWTQPTPSNTSAASVESAFCG
NYQERDIDABIESRQTVSPQMTSGSADGASSTRLDQSPSLESSLGSLPDTSRTEPILG
SSGYDDDEAKEAARRKRQEEMYSARRIFIAGAAGAAAAATTTADAEKAAIDALANHSQ
HLHLASSPAFEVLSEETREKIRRMQKKQSWHDTKELRSGELLKTYSPTKDLTDALSCT
SDYSTTSSAPLSTNPPLAVACADQPNSSSDYASSDPSPCARNPSTSPASRPSNLAISP
AQLHATSSSGQSHQPMSRSQKIRLRTKLGSRDPARRHTLSDVDTLKEGRLDKLARWFG
IRKSSPDVSRDEVSDDEKNHQEAPPLPAAAPPVIVRTSPNELTPVSGDELL"
exon 4696. .5541
/note="putative"
/number=3
exon complement(5169. .5380)
/partial
/note="ZC21.9: orphan; putative"
/number=6
exon 5657. .5802
/note="putative"
/number=4
exon 5854. .6064
/note="putative"
/number=5
exon 7581. .7714

/note="complement; putative"
/number=7
CDS complement(join(7581. .7714,7759. .7855))
/partial
/standard_name="ZC21.6 protein"
/note="putative"
/codon_start=1
/db_xref="PID:g289734"
/translation="TKAKSGKGTNTAKSSGGNNGTNLNAKRSNTTQDDVKFNLNIKNT
EQKRDEFDDDEENPLAKIPTKARLPKKVSGLL"
exon 14870. .15023
/note="putative"
/number=8
CDS join(14870. .15023,15131. .15373,16179. .16710,16862. .16973)
/partial
/standard_name="ZC21.3 protein"
/note="homology with dual bar protein; putative"
/codon_start=1
/db_xref="PID:g289733"
/translation="IKNLKARFLKDYIKNGPESHSKTFEIFLLREKFIPGKVRFKNFL
ILEIASRRTSSGSSSYNTNGVQRDSTDGSSNNQPGYLPDYQGDRYTDPQRIAETQALN
KAFFDSIGGAPSPPPLTPAQQGMNGSSSSTGAPSTTYSPTSTYGYSSQNYNPSSASSS
NVYRNPSTSYGNTQQSNQNTVAYDTTNQNSPYYYQRVPSTAQYNSVLGVNAGFPQSSN
QNQVLRDQSNNQGYMNNNQGMYNTQTTQGYGNNQGSYQNQMMTNNNPQNTQTTTFLY
ASSIQPQQQQQQQQYNNQYNNQQMNSNGYNNQPITLHQSISMDESNNFSNFLYIFSLA
TFALSDCIDYSE"
exon 15131. .15373
/note="putative"
/number=9
exon 16179. .16710
/note="putative"
/number=10
misc_feature 24684
/note="possible trans-spliced leader site for ZC21.2"
CDS join(24685. .24824,25082. .25214,25262. .25418,25468. .25681,
25738. .26125,26232. .26381,26426. .26563,26646. .26842,
26887. .27067,27321. .27890,28061. .28264)
/standard_name="ZC21.2 protein"
/note="homology with transient receptor potential protein; putative"
/codon_start=1
/db_xref="PID:g289732"
/translation="MLSAAGRRFSRCAPSPRPRRNYALANIPPPAAVHIRTNDMISPE
ERRKSRVKSLKHAQLLHGYIILSNNLFRFLEAAELGNKPTLQECLDYDGDRRLNVNCL
DSMGRTALEIAVDNENMEVVELLLQQPDIRIGNALLCAIRECGVYRLVEVLVNHPNITR
EMLGDGWSQALDPSEAASAEYSSDISPVILAAQLNQFEILQMLIRKDASIEKPHRHSC
ICETCDRERLNDSLQYSLKRINTFRALASPAWMSLTSPDPILSAFKLSWDLQRLAFEE
HEFRETTLQLSEQCKQYSCDLLSQCRSSEEVIAILNKDGNVNDNIDVWASKLSLSLSRL
KLAIKYEQKAFVSHPHCQQLLTSIWYEGIPYRQRSGTWANFPLYAFLLFLWPIPCLMY
ILMPKSRLGRLVRSPFMKFFYYSVSFATFLGLLTWATFEDYRIYEKGERGGMTRASDRG
PPATWVESLVFTWVIGMLWSEIKQLWEEGFKRYMRQWWNWLDFLMICLYLCTISIRLS

```
                     AYYIFTYREDPYRYTVRTYWTSEEPMLVAEALFAVGNVFSFARIIYLFQTNPYLGPLQ
                     ISLGCMLVDVAKFCFIFVLIISSFSIGLAQLYWYYDPNTDVCLPGATCKHSSNVFSSI
                     ADSYLTLLWSLFSITKPEDTDVVENHKITQWVGQGMFIMYHCTSIIVLLNMLIAMMSH
                     SFQIIINDHADLEWKFHRTKLWWAHFDEGSSLPPPFNIIVTPKSLIYVMNCLFNTVRWL
                     LGKYTYQKNRMATIRRPGYSRKRNEMEKSGGHDDDSLKPLTYADIITRLVARFIHQT
                     KKDMKKDGGVNEDDLHEIKQDISSLR*
     exon            25082. .25214
                     /note="putative"
                     /number=11
     exon            25262. .25418
                     /note="putative"
                     /number=12
     exon            25468. .25681
                     /note="putative"
                     /number=13
     exon            25738. .26125
                     /note="putative"
                     /number=14
     exon            26232. .26381
                     /note="putative"
                     /number=15
     exon            26426. .26563
                     /note="putative"
                     /number=16
     exon            26646. .26842
                     /note="putative"
                     /number=17
     exon            26887. .27067
                     /note="putative"
                     /number=18
     exon            27321. .27890
                     /note="putative"
                     /number=19
     CDS             join(29237. .29352,30055. .30276,30403. .30568,30655. .30785,
                     30832. .30988)
                     /standard_name="ZC21.1 protein"
                     /note="putative"
                     /codon_start=1
                     /db_xref="PID:g289731"
                     /translation="MKFRRIRNTLPRDIIFINFQAKTCFFLQIDRPTRYSPQLYELRD
                     DRRREIVRSSSHIDAVKRDIMRTMSTTSRRPYGGSMRLPKTRPSVAEESEEDDKSDET
                     SSTDEEADETRSRKSSVLYIPPVNSTLPGIISEEAPVKKSTKRRASEADSKLPDRPLS
                     DTYTSSPTPKLLEVPAQPPHSALRKTDTSMSFPINGTALDTSIVKPPRGVLRASQENL
                     PAVELIETLKKEKNEKLDRLISGISENVKSPSPASHSHVGFNVEK*
     exon            30055. .30276
...
Note: remainder of annotations omitted.

  Query Match          10.3%;  Score 18;  DB 18;  Length 36087;
  Best Local Similarity 100.0%;  Pred. No. 1.66e+01;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```



```
Db  10658 CAAAAATAAATTATTTTT 10675
          |||||||||||||||||
Cp     57 CAAAAATAAATTATTTTT 40

RESULT   41
LOCUS       CELB0238    41754 bp    DNA             INV       06-AUG-1997
DEFINITION  Caenorhabditis elegans cosmid B0238.
ACCESSION   AF016450
NID         g2315480
KEYWORDS    .
SOURCE      Caenorhabditis elegans strain=Bristol N2.
  ORGANISM  Caenorhabditis elegans
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Nematoda;
            Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea;
            Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 41754)
  AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
  TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
  JOURNAL   Nature 368 (6466), 32-38 (1994)
  MEDLINE   94150718
REFERENCE   2  (bases 1 to 41754)
  AUTHORS   Blanchard,M. and Bradshaw,H.
  TITLE     The sequence of C. elegans cosmid B0238
  JOURNAL   Unpublished (1997)
REFERENCE   3  (bases 1 to 41754)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JUL-1997)
COMMENT     Submitted by:
                  Genome Sequencing Center
                  Department of Genetics, Washington University,
                  St. Louis, MO  63110, USA, and
                  Sanger Centre, Hinxton Hall
                  Cambridge CB10 1RQ, England
                  e-mail: rw@nematode.wustl.edu and jes@sanger.ac.uk


            NOTICE:  This sequence may not be the entire insert of this clone.
```


It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:

The 5' end of this cosmid lies in a gap;3' cosmid is C52A10, 200 bp overlap. Actual start of this cosmid is at base position 1 of CELB0238; actual end is at 41754 of CELB0238. This cosmid lies in an unanchored cluster, the orientation of which is unknown.

NOTES:

Coding sequences below are predicted from computer analysis, using the program Genefinder(P. Green and L. Hillier, ms in preparation).

```
FEATURES             Location/Qualifiers
     source          1. .41754
                     /organism="Caenorhabditis elegans"
                     /strain="Bristol N2"
                     /db_xref="taxon:6239"
                     /chromosome="V"
                     /clone="B0238"
     gene            3596. .5788
                     /gene="B0238.9"
     CDS             join(3596. .3730,3801. .3966,4018. .4298,4349. .4588,
                     4940. .5633,5712. .5788)
                     /gene="B0238.9"
                     /note="similar to drosophila HCD protein (NID:g902624);
                     coded for by C. elegans cDNA yk56f3.5; coded for by C.
                     elegans cDNA yk159g11.5; coded for by C. elegans cDNA
                     yk159g11.3; coded for by C. elegans cDNA yk56f3.3"
                     /codon_start=1
                     /db_xref="PID:g2315488"
                     /translation="MRIVRTHRDEFLRTLCLNLFCCLLINSIEKSKQIQSSAYFFRNS
                     HSFAIEKFKRKQQKMPRGLRRADLVKRHRHSTTGDKDGGVPEVIGCPVLDPIICQCPK
                     DEIELGEGVKMTCTWESCPYSSRPLHHICTQLLEDNLVKRLASLGSARGWTVPQRRNN
                     LWERKGQSLIGKFCRCRCDRGQNTRDKQALIEKEKAVEKEKKKAKKAAKQLPQLQFNS
                     KPLAAIEEKRGDADVFHSPSIASSTRHHTFSTTTRSRLHTDRSASSILTHTIGRTWS
                     ESSFAGETNGQYDNNQEPHPSNCECVFHHDYDADDQIDTDFECESNHSDVIVPAPLPP
                     LQARSYAATIMRNGTPKVTNYSPDSGLDQQTPRFSLSSSSGGDVDNQHGDFHVETRIS
                     EHLNALGLSIMSPVENANENVNYEESPPYPELTSTPIVSKKQREPLRAKKSTSVSKLP
                     LAPSSQLFNEESRCGFRFNVPVREMMDIWQESGALSPAIRETQAENTEKRAENASGVL
                     QYGWTPFPGNGFNLGERLYYFP*
     misc_feature    complement(12417. .12418)
                     /note="SL1 trans-splice site; see EST yk78a7.5"
     gene            complement(13177. .13999)
                     /gene="B0238.12"
     CDS             complement(join(13177. .13497,13699. .13820,13871. .13925,
                     13997. .13999))
                     /gene="B0238.12"
                     /note="similar to Lytechinus variegatus ECM 18
                     (GB:U40064)"
                     /codon_start=1
                     /evidence=not_experimental
                     /db_xref="PID:g2315490"
                     /translation="MKVFIVLAALIGCYVYAQDAQILPVGGQVGGGQRLPCRGRNEET
                     KTCGTACEPSCTNPNPMCTKQCINNVCQCRSGYVRNEITRQCVRQAQCSRPGTGFGSS
                     TPFPSQSPQRCGRNETFRTCGSSCEPSCTTPRPQACTMQCIVNVCQCSEGFVRGPSGC
                     VRQRDC"
     gene            14524. .17987
                     /gene="B0238.8"
```

```
     CDS             join(14524. .14823,16402. .16557,17129. .17464,17518. .17744,
                     17793. .17873,17927. .17987)
                     /gene="B0238.8"
                     /codon_start=1
                     /evidence=not_experimental
                     /db_xref="PID:g2315487"
                     /translation="MRTILPIEIEILKLPKKCAKSQFAGNFNSEICRFADLPEISIST
                     ICRFADLPKILIPATCRFADLPEILSPGSVSICRFAENPQTRQFVDEPATINFFFPAV
                     LNKDKQSANVTYKLMNLINFLQLVQASGHFLTAPILIFPNLLSRFDFVIRTIGCIMNS
                     FWVADLPVMTLLAVCRILIFSNIINAKKFPPFPIKIVLPSLMSWIFPVIVSGSITQMMK
                     FSPPWWGYDYEMPYAEVFDSIEIYLSFPCLAISYLAYLIIIFLITYCKNIKSTAKSRKI
                     EISILLQYTFVTTYITFMVIIWHPTLFSLLSFIDMSNKRNQAILNGLWIIHCYVNPTM
                     LLIFNRSIREDVGNVLDNQMKCGNKVKAGIVSSALNKNDLNFLCRCLKCLKN"
     gene            18803. .21570
                     /gene="B0238.7"
     CDS             join(18803. .18952,18999. .19157,19208. .19294,19971. .20304,
                     20369. .20491,20535. .20744,20797. .20972,21023. .21297,
                     21348. .21570)
                     /gene="B0238.7"
                     /note="strong similarity to C. elegans esterase CM06B1
                     (SP:Q07085)"
                     /codon_start=1
                     /evidence=not_experimental
                     /db_xref="PID:g2315493"
                     /translation="MSLSFYWGIGCLLLSIVEGGFLVNLTPEKNSQILNAPCGPIRGN
                     IYKHKNGTKIVDGYLGIPFAKAPVGALKFKKPVAAEKWTEPRDCYEYGPGCPQTGRFA
                     ALTAKSTAAFAEENCLTLNVFAPRWKLAEPPKGLPVLVYIYGGGYEIGFTSYMDDYSL
                     TGTLPLKDVIVVTMNYRLGPLGFLTIADGIANGNYGLWDQTLALQWVQDNIASFGGNP
                     ESVTLSGTSAGATSTDFLSLSPHSNQLFQRSISMSGTAFCNFAIRPKDVEVGVCKEPA
                     QFQGYTGTDSQSLFDWYMAQNVSKFKEVAGPQSTAAGAVPFIPNFDGDFFPKPFDILR
                     REAPKKDALITVVTMESVGMIIFNPLFANAYNNFEGFINETYGPDVTDYYADVKKKLT
                     GFYLKDVNASDKATVNRNRVVDYVSDSVFNSGTLDAVKSYVKYGNKVYLGSFNYFNDAT
                     KDMVELMQPFKEASHSSDFKYIPSGVHANFTPTDEEMEVMDMSGTLMANFVKYGDPN
                     GKSGTKPWKPYTILKPNMYYKIDYPTCEMSDNFQNGRLRIYDDINKNSKKYQQIVYGK
                     VSKRVVINARAG"
     gene            22354. .23866
                     /gene="B0238.6"
     CDS             join(22354. .22458,22500. .22655,23056. .23391,23444. .23673,
                     23722. .23866)
                     /gene="B0238.6"
                     /codon_start=1
                     /evidence=not_experimental
                     /db_xref="PID:g2315486"
                     /translation="MSLFQLNNDVAMVKYGVAYFCIALPLFPVFFVLMKTICIKDREN
                     SNLTYKLMNLINFCQVGQTISHFISGPMLIFPNLLMKLDVIIRMMGCIMNSLWIADFP
                     VMTLLAVCRILIFSNIINSKKFPNSIKCALFLIITWTMPLIFVGCATLNMQLVTPGWD
                     YDFSVSSAEILANLEVLLSFACLASSYVAYLFMAYLIYAKKNLVSCVHSRKNEIAILV
                     QSFVTTYITAMIFVWHQSLFSKISFIDMESKRNQAILNCCLIIHCYVNPIMTLVCNK
                     SIRDECLRFLKIKKRPSRPSRSDILSKLSSIHPINPSANNEHNIESQ"
     gene            24479. .25760
                     /gene="B0238.5"
```



```
     CDS             join(24479. .24601,24651. .24806,24965. .25300,25347. .25576,
                     25670. .25760)
                     /gene="B0238.5"
...
Note: remainder of annotations omitted.

  Query Match           10.3%;  Score 18;  DB 18;  Length 41754;
  Best Local Similarity 100.0%;  Pred. No. 1.66e+01;
  Matches    18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  39316 TTAAAAATAATTTATTTT 39333
          ||||||||||||||||||
Qy     38 TTAAAAATAATTTATTTT 55


RESULT   42
LOCUS       CELF55A8    45131 bp    DNA             INV       09-JUN-1998
DEFINITION  Caenorhabditis elegans cosmid F55A8.
ACCESSION   AF067612
NID         g3193146
KEYWORDS    .
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 45131)
  AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
  TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
  JOURNAL   Nature 368 (6466), 32-38 (1994)
  MEDLINE   94150718
REFERENCE   2  (bases 1 to 45131)
  AUTHORS   Langston,Y., Wohldmann,P. and Duckels,G.
  TITLE     The sequence of C. elegans cosmid F55A8
  JOURNAL   Unpublished (1998)
REFERENCE   3  (bases 1 to 45131)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAY-1998) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     Submitted by:
```


Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 1RQ, England
e-mail: rw@nematode.wustl.edu and jes@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:

The 5' cosmid is W03A6, 200 bp overlap;3' cosmid is Y48A5B, 200 bp overlap. Actual start of this cosmid is at base position 1 of CELF55A8; actual end is at 45131 of CELF55A8

NOTES:

Coding sequences below are predicted from computer analysis, using the program Genefinder(P. Green and L. Hillier, ms in preparation).

```
FEATURES             Location/Qualifiers
     source          1. .45131
                     /organism="Caenorhabditis elegans"
                     /strain="Bristol N2"
                     /db_xref="taxon:6239"
                     /clone="F55A8"
                     /chromosome="IV"
     gene            <1399. .9573
                     /gene="F55A8.2"
     CDS             join(<1399. .1521,3219. .3326,5285. .5889,6821. .7454,
                     9478. .9573)
                     /gene="F55A8.2"
                     /note="contains similarity to eukaryotic protein kinase
                     domains (Pfam; pkinase, score; 197.98), protein kinase C
                     terminal domains (Pfam; pkinase_C, score; 21.56) and
                     cyclic nucleotide binding domains (Pfam; cNMP_binding,
                     scores; 19.92 and 121.99); partial CDS; coded for by C.
                     elegans cDNA yk2f8.5; coded for by C. elegans cDNA
                     yk317d2.5; coded for by C. elegans cDNA yk395h7.5; coded
                     for by C. elegans cDNA yk349d8.5; coded for by C. elegans
```



```
                     cDNA yk443d2.5; coded for by C. elegans cDNA yk410a4.5;
                     coded for by C. elegans cDNA yk81h11.5; coded for by C.
                     elegans cDNA yk385d7.5; coded for by C. elegans cDNA
                     yk55f2.5; coded for by C. elegans cDNA yk55a6.5; coded for
                     by C. elegans cDNA yk486b10.5; coded for by C. elegans
                     cDNA yk469c5.5; coded for by C. elegans cDNA yk52d1.5;
                     coded for by C. elegans cDNA yk30f8.5; coded for by C.
                     elegans cDNA yk41a5.3; coded for by C. elegans cDNA
                     yk17d12.3; coded for by C. elegans cDNA yk52d1.3; coded
                     for by C. elegans cDNA yk4g10.3; coded for by C. elegans
                     cDNA yk87g1.3; coded for by C. elegans cDNA yk81h11.3;
                     coded for by C. elegans cDNA yk87g12.3; coded for by C.
                     elegans cDNA yk55f2.3; coded for by C. elegans cDNA
                     yk55a6.3"
                     /codon_start=1
                     /db_xref="PID:g3193148"
                     /translation="TASVQALTDVQLWVLDRSVFQNITQRLGMERHSQLMNFLTKVSI
                     FQNLSEDRISKMADVMDQDYYDGGHYIIRQGEKGDAPFVINSGQVKVTQQIEGETEPR
                     EIRVLNQGDFFGERALLGEEVRTANIIAQAPGVEVLTLDRESFGKLIGDLESLKKDYG
                     DKERLAQVVREPPSPVKIVDDFREEFAQVTLKNVKRLATLGVGGFGRVELVCVNGDKA
                     KTFALKALKKKHIVDTRQQEHIFAERNIMMETSTDWIVKLYKTFRYSLQQQNTSKNIW
                     FLLDQKFVYMLLEVCLGGELWTTLRDRGHFDDYTARFYVACVLEGLEYLHRKNIVYRD
                     LKPENCLLANTGYLKLVDFGFAKKLASGRKTWTFCGTPEYVSPEIILNKGHDQAADYW
                     ALGIYICELMLGRPPFQASDPMKTYTLILKGVDALEIPNRRIGKTATALVKKLCRDNP
                     GERLGSGSGGVNDIRKHRWFMGFDWEGLRSRTLKPPILPKVSNPADVTNFDNYPPDND
                     VPPDEFSGWDEGF"
     repeat_region   complement(17051. .17356)
                     /note="C. elegans telomere-like hexamer repeat"
     repeat_region   complement(19875. .20021)
                     /note="C. elegans telomere-like hexamer repeat"
     repeat_region   complement(32568. .32960)
                     /note="C. elegans telomere-like hexamer repeat"
     misc_feature    43087. .43088
                     /note="SL1 trans-splice site; see ESTs yk474a4.5 and
                     yk484g12.5"
     gene            43121. .>44568
                     /gene="F55A8.1"
     CDS             join(43121. .43231,43277. .43807,44328. .>44568)
                     /gene="F55A8.1"
                     /note="contains similarity to a GATA-type zinc finger
                     (PROSITE; ZN_FING); partial CDS; coded for by C. elegans
                     cDNA yk330b8.5; coded for by C. elegans cDNA yk524e12.5;
                     coded for by C. elegans cDNA yk484g12.5; coded for by C.
                     elegans cDNA yk115f7.5; coded for by C. elegans cDNA
                     yk390a2.5; coded for by C. elegans cDNA yk334d12.5; coded
                     for by C. elegans cDNA yk96g8.5; coded for by C. elegans
                     cDNA yk404g2.5; coded for by C. elegans cDNA yk345d6.5;
                     coded for by C. elegans cDNA yk330f7.5; coded for by C.
                     elegans cDNA yk474a4.5; coded for by C. elegans cDNA
                     yk446e9.5; coded for by C. elegans cDNA yk341e9.5"
                     /codon_start=1
```

```
                     /db_xref="PID:g3193147"
                     /translation="MTETRPESAEQQHHEVLQRPSDEPCSGCKQLQKDVAKTISMVME
                     RMDKLQYRLDELLKENNELKSSSVSSGKASPSPAESRSSPKLVETVVAPVSGARKRKP
                     KERSPPAAASPLPDPSNLMNGFMPDPLMMSNPMGMMQLLSMVQQQQQQQQHHQHIENQ
                     QSVSPPQSKSVKIEDPMDQDVKQESESERSDIPTATEAQNLLDALTAQPSSNGQATSTT
                     SPPSSSSQVQAVIEAVATPSSQSQDSSMPEKTETSGDPNAARCSNCRTDKTTAWRRDA
                     EGKLVCNPCGLYYRLHKV"
BASE COUNT    13643 a   8985 c   8063 g  14440 t
ORIGIN

  Query Match           10.3%;  Score 18;  DB 18;  Length 45131;
  Best Local Similarity 100.0%;  Pred. No. 1.66e+01;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  39333 TTATTTTTAACCTCCAAA 39350
          ||||||||||||||||||
Cp     47 TTATTTTTAACCTCCAAA 30


RESULT   43
LOCUS       AB017065    61290 bp    DNA              PLN       02-SEP-1998
DEFINITION  Arabidopsis thaliana genomic DNA, chromosome 5, P1 clone: MFC16,
            complete sequence.
ACCESSION   AB017065
NID         g3510341
KEYWORDS    HTG.
SOURCE      Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui P1
            clone:MFC16.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Charophyta/Embryophyta group;
            Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta;
            Magnoliophyta; eudicotyledona; Rosidae; Capparales; Brassicaceae;
            Arabidopsis.
REFERENCE   1  (sites)
  AUTHORS   Nakamura,Y.
  TITLE     Structural Analysis of Arabidopsis thaliana Chromosome 5. IX
  JOURNAL   Unpublished (1998)
REFERENCE   2  (bases 1 to 61290)
  AUTHORS   Nakamura,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-AUG-1998) to the DDBJ/EMBL/GenBank databases.
            Yasukazu Nakamura, Kazusa DNA Research Institute, Laboratory of
            Gene Structure 2; 1532-3, Yana, Kisarazu, Chiba 292, Japan
            (E-mail:ynakamu@kazusa.or.jp, Tel:+81-438-52-3935,
            Fax:+81-438-52-3934)
FEATURES             Location/Qualifiers
     source          1. .61290
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="5"
```



```
                     /clone="MFC16"
                     /clone_lib="Mitsui P1"
BASE COUNT    19498 a  11162 c  10939 g  19691 t
ORIGIN

  Query Match           10.3%;  Score 18;  DB 24;  Length 61290;
  Best Local Similarity 100.0%;  Pred. No. 1.66e+01;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  17437 TTAAAAATAATTTATTTT 17454
          ||||||||||||||||||
Qy     38 TTAAAAATAATTTATTTT 55


RESULT   44
LOCUS       AC004140    74918 bp    DNA              PRI       13-FEB-1998
DEFINITION  Homo sapiens PAC clone DJ0841B21 from 7q21.1-q31.1, complete
            sequence.
ACCESSION   AC004140
NID         g2880082
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 74918)
  AUTHORS   Kalicki,J. and Angell,S.
  TITLE     The sequence of Homo sapiens PAC clone DJ0841B21
  JOURNAL   Unpublished (1998)
REFERENCE   2  (bases 1 to 74918)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-FEB-1998) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     SUBMITTED BY:  WUGSC
            Genome Sequencing Center
            Department of Genetics
            Washington University
            St. Louis MO 63108, USA
            http://genome.wustl.edu/gsc
            mailto:sapiens@watson.wustl.edu

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded or sequenced with an alternate
            chemistry; an attempt was made to resolve all sequencing problems,
            such as compressions and repeats; all regions were covered by
```



```
            sequence from more than one subclone; and the assembly was
            confirmed by restriction digest.

            MAPPING INFORMATION:
            The sequence of this clone was established as part of a mapping and
            sequencing collaboration between the NHGRI Chromosome 7 Mapping
            Project and the Washington University Genome Sequencing Center.
            For additional information about the map position of this sequence,
            see http://www.nhgri.nih.gov/DIR/GTB/CHR7 or
            mailto:egreen@nhgri.nih.gov

            SOURCE INFORMATION:
            This clone was derived from human PAC library RPCI-5, prepared by
            Pieter de Jong and coworkers at Roswell Park Cancer Institute,
            using the method described by Ioannou et al., Nature Genetics
            6:84-9 (1994).  The library is from one male donor.  For further
            details, see http://bacpac.med.buffalo.edu/
            The clone is available from Genome Systems, Inc.
            (http://www.genomesystems.com).
            VECTOR:  pCYPAC2
            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is DJ1090P18, 200 bp overlap. The
            actual start of this clone is at base position 1 of DJ0841B21;
            actual end is at 74918 of DJ0841B21. The orientation of this clone
            unknown.

            This clone contains STS sWSS30 (NID:g1254687).
FEATURES             Location/Qualifiers
     source          1. .74918
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /clone="DJ0841B21"
                     /clone_lib="RPCI-5"
                     /map="7q21.1-q31.1"
     repeat_region   271. .558
                     /rpt_family="MaLR"
     repeat_region   2235. .2333
                     /rpt_family="Alu"
     repeat_region   2806. .2853
                     /rpt_family="Alu"
     repeat_region   4672. .4851
                     /rpt_family="Alu"
     repeat_region   4852. .5147
                     /rpt_family="Alu"
     repeat_region   5680. .5898
                     /rpt_family="L1"
     repeat_region   6525. .6825
                     /rpt_family="Alu"
     repeat_region   6964. .7022
                     /rpt_family="MaLR"
```



```
     repeat_region   7024. .7077
                     /rpt_family="L1"
     repeat_region   7099. .13212
                     /rpt_family="L1"
     repeat_region   13227. .13528
                     /rpt_family="Alu"
     repeat_region   13531. .13818
                     /rpt_family="L1"
     repeat_region   13831. .13904
                     /rpt_family="MaLR"
     repeat_region   13956. .14134
                     /rpt_family="MaLR"
     repeat_region   14147. .14210
                     /rpt_family="MaLR"
     repeat_region   14370. .14433
                     /rpt_family="(CA)n"
     repeat_region   14492. .14600
                     /rpt_family="MER1_type"
     repeat_region   14634. .14684
                     /rpt_family="MER1_type"
     repeat_region   14685. .14833
                     /rpt_family="MER1_type"
     repeat_region   15257. .15393
                     /rpt_family="MER21_g"
     repeat_region   15408. .15852
                     /rpt_family="MER21_g"
     repeat_region   16139. .16216
                     /rpt_family="L2"
     repeat_region   18452. .18549
                     /rpt_family="L2"
     repeat_region   20729. .21029
                     /rpt_family="Alu"
     repeat_region   22635. .23998
                     /rpt_family="L1"
     repeat_region   24246. .24346
                     /rpt_family="L1"
     repeat_region   24884. .27496
                     /rpt_family="L1"
     repeat_region   27671. .27933
                     /rpt_family="L1"
     repeat_region   27984. .28122
                     /rpt_family="Alu"
     repeat_region   28123. .28472
                     /rpt_family="L1"
     repeat_region   28486. .28788
                     /rpt_family="Alu"
     repeat_region   29567. .29873
                     /rpt_family="Alu"
     repeat_region   31594. .31845
                     /rpt_family="MIR"
     repeat_region   34169. .34368
```

```
                     /rpt_family="MER1_type"
     gene            complement(37224. .41031)
                     /gene="WUGSC:H_DJ0841B21.1"
     CDS             complement(join(37224. .37345,39352. .39396,39825. .39921,
                     40909. .41031))
                     /gene="WUGSC:H_DJ0841B21.1"
                     /note="tachykinin neurotransmitter; possible alternatively
                     spliced products; similar to P49110 (PID:g1351255);
                     H_DJ0841B21.1"
                     /codon_start=1
                     /evidence=not_experimental
                     /db_xref="PID:g2880083"
                     /translation="MKILVALAVFFLVSTQLFAEEIGANDDLNYWSDWYDSDQIKEEL
                     PEPFEHLLQRIARRPKPQQFFGLMGKRDADSSIEKQVALLKALYGHKTDSFVGLMGKR
                     ALNSGMYEIMTENRQYLKSISIFSKT"
     misc_feature    complement(37298. .37347)
                     /gene="WUGSC:H_DJ0841B21.1"
                     /note="match to EST AA447465 (NID:g2161135) zw89h02.r1"
     repeat_region   39127. .39223
                     /rpt_family="(GA)n"
     misc_feature    complement(39351. .39399)
                     /gene="WUGSC:H_DJ0841B21.1"
                     /note="match to EST AA447465 (NID:g2161135) zw89h02.r1"
     misc_feature    complement(39823. .39925)
                     /gene="WUGSC:H_DJ0841B21.1"
                     /note="match to EST AA447465 (NID:g2161135) zw89h02.r1"
     misc_feature    39938. .41823
                     /note="CpG_island (%GC=60.4, o/e=0.80, #CpGs=137)"
     misc_feature    complement(40906. .41049)
                     /note="match to EST AA447465 (NID:g2161135) zw89h02.r1"
     misc_feature    complement(41449. .41589)
                     /note="match to EST AA447465 (NID:g2161135) zw89h02.r1"
     repeat_region   43188. .43541
                     /rpt_family="MaLR"
     repeat_region   44153. .44486
                     /rpt_family="MER2_type"
     repeat_region   44537. .44815
                     /rpt_family="L1"
     repeat_region   45262. .45557
                     /rpt_family="Alu"
     repeat_region   46357. .46649
                     /rpt_family="Alu"
     repeat_region   46974. .47419
                     /rpt_family="L1"
     repeat_region   48615. .48822
                     /rpt_family="MIR"
     repeat_region   48826. .49128
                     /rpt_family="Alu"
     repeat_region   49238. .49655
                     /rpt_family="Retroviral"
     repeat_region   50731. .51627
```



```
                     /rpt_family="L1"
     repeat_region   51632. .51930
                     /rpt_family="Alu"
     repeat_region   51931. .52358
                     /rpt_family="L1"
     repeat_region   52359. .52691
                     /rpt_family="MaLR"
     repeat_region   52701. .52785
                     /rpt_family="L1"
     repeat_region   52974. .53909
                     /rpt_family="L1"
...
Note: remainder of annotations omitted.

  Query Match           10.3%;  Score 18;  DB 26;  Length 74918;
  Best Local Similarity 100.0%;  Pred. No. 1.66e+01;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  31013 AAAATAATTTATTTTTGC 31030
          ||||||||||||||||||
Qy     41 AAAATAATTTATTTTTGC 58


RESULT   45
LOCUS       AC004422    77137 bp    DNA              INV       02-APR-1998
DEFINITION  Drosophila melanogaster (P1 DS08493 (D141)) DNA sequence, complete
            sequence.
ACCESSION   AC004422 AC002729 AC003180 AC003194 AC003744 AC003185 AC003191
            AC003187 AC003189 AC003182 AC002728 AC003190 AC002731 AC002733
            AC002730 AC002735 AC002723 AC002721 AC002725 AC002724 AC002726
            AC002734 AC002722 AC003179 AC003188 AC003186 AC003181 AC003183
            AC003184 AC003192 AC003193 AC002727 AC002732
NID         g3006215
KEYWORDS    HTG.
SOURCE      Drosophila melanogaster (Subclones in Sac from P1 clone DS08493
            (D141)) DNA.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea;
            Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 77137)
  AUTHORS   Celniker,S.E., Aghavani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A.,
            Hummasti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B.,
            Lomotan,M.A., Mazda,P., Mok,M.S., Moshrefi,A.R., Moshrefi,M.,
            Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,D., Snir,E.,
            Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R., Zieran,L.L.
            and Kimmel,B.E.
  TITLE     Sequencing of Drosophila chromosome 2L, region 23D2-23D4
  JOURNAL   Unpublished (1997)
REFERENCE   2  (bases 1 to 77137)
```



```
  AUTHORS   Celniker,S.E., George,R.A., Galle,R., Svirskas,R.R., Hoskins,R.A.,
            Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Chavez,C.,
            Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A.,
            Hummasti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B.,
            Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S., Moshrefi,A.R.,
            Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,E.,
            Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R.,
            Zieran,L.L. and Kimmel,B.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAR-1998) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Apr 2, 1998 this sequence version replaced gi:2995563.
            Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site
            (http://fruitfly.berkeley.edu/sequence-archive.html) or send email
            to drosophila@genome.lbl.gov.
            Library location: 45_89.
            This P1 was assembled from the following subclones: 2_b9
            (AC002729), 1_c7 (AC003180), 1_c1, 2_h2 (AC003194), 2_h7
            (AC003744), 1_g9 (AC003185), 2_e10 (AC003191), 2_b10 (AC003187),
            2_d12 (AC003189), 1_e10 (AC003182), 2_b7 (AC002728), 2_d2
            (AC003190), 2_e12 (AC002731), 2_g12 (AC002733), 2_c3 (AC002730),
            2_h9 (AC002735), 1_d3 (AC002723), 1_c2 (AC002721), 1_f2 (AC002725),
            1_d7 (AC002724), 1_e6, 1_g8 (AC002726), 2_g2 (AC002734), 1_c8
            (AC002722), 1_b7 (AC003179), 2_b4 (AC003188), 2_d5, 2_e8
            (AC003186), 1_d8 (AC003181), 1_e8 (AC003183), 1_f4 (AC003184), 2_e6
            (AC003192), 2_g3 (AC003193), 2_b6 (AC002727).
FEATURES                Location/Qualifiers
     source             1. .77137
                        /organism="Drosophila melanogaster"
                        /db_xref="taxon:7227"
                        /chromosome="2L"
                        /map="23D2-23D4"
                        /clone="P1 DS08493 (D141)"
BASE COUNT     22877 a  15210 c  15774 g  23276 t
ORIGIN

  Query Match           10.3%;  Score 18;  DB 18;  Length 77137;
  Best Local Similarity 100.0%;  Pred. No. 1.66e+01;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  24356 AAAAATAATTTATTTTTG 24373
          ||||||||||||||||||
Qy     40 AAAAATAATTTATTTTTG 57
```



```
RESULT   46
LOCUS       ATAC003105  95167 bp    DNA              PLN       09-JAN-1998
DEFINITION  Arabidopsis thaliana chromosome II BAC F18A8 genomic sequence,
            complete sequence.
ACCESSION   AC003105
NID         g2760829
KEYWORDS    HTG.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryotae; Viridiplantae; Charophyta/Embryophyta group;
            Embryophyta; Tracheophyta; seed plants; Magnoliophyta;
            eudicotyledons; Rosidae; Capparales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 95167)
  AUTHORS   Rounsley,S.D., Lin,X., Ketchum,K.A., Crosby,M.L., Brandon,R.C.,
            Sykes,S.M., Kaul,S., Mason,T.M., Kerlavage,A.R., Adams,M.D.,
            Somerville,C.R. and Venter,J.C.
  TITLE     Arabidopsis thaliana chromosome II BAC F18A8 genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 95167)
  AUTHORS   Rounsley,S.D. and Lin,X.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-NOV-1997) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, rounsley@tigr.org
REFERENCE   3  (bases 1 to 95167)
  AUTHORS   Rounsley,S.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JAN-1998) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
COMMENT     On Jan 9, 1998 this sequence version replaced gi:2708756.
            Address all correspondence to:
            Steve Rounsley
            The Institute for Genomic Research
            9712 Medical Center Dr,
            Rockville, MD 20850,
            USA
            e-mail: rounsley@tigr.org
            BAC clone F18A8 is from Arabidopsis chromosome II and close to the
            molecular marker B68.
            The orientation of the sequence is from SP6 to T7 end of the BAC
            clone.
            Genes were identified by a combination of three methods: Gene
            prediction programs including GRAIL (available by anonymous ftp
            from arthur.epm.ornl.gov), Genefinder (Phil Green, University of
            Washington), Genscan (Chris Burge,
            http://gnomic.stanford.edu/~chris/GENSCANW.html), and NetPlantGene
            (http://www.cbs.dtu.dk/netpgene/cbsnetpgene.html), searches of the
            complete sequence against a peptide database and the Arabidopsis
            EST database at TIGR (http://www.tigr.org/tdb/at/at.html).
            Annotated genes are named to indicate the level of evidence for
            their annotation. Genes with similarity to other proteins are named
            after the database hits. Genes without significant peptide
```

```
            similarity but with EST similarity are named as 'unknown' proteins.
            Genes without protein or EST similarity, that are predicted by more
            than two gene prediction programs over most of their length are
            annotated as 'hypothetical' proteins. Genes encoding tRNAs are
            predicted by tRNAscan-SE (Sean Eddy,
            http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are
            identified by repeatmasker (Arian Smit,
            http://ftp.genome.washington.edu/RM/RepeatMasker.html). Regions of
            genomic sequence that are not annotated as genes but have predicted
            exons by GRAIL are annotated as misc features.
FEATURES              Location/Qualifiers
     source           1. .95167
                      /organism="Arabidopsis thaliana"
                      /cultivar="Columbia"
                      /db_xref="taxon:3702"
                      /chromosome="II"
                      /map="B68"
                      /clone="F18A8"
     misc_feature     1. .178
                      /note="overlap with BAC clone T9J22
                      (AC002505:115674. .115851)."
     mRNA             <86. .1745
                      /gene="F18A8.1"
     gene             <86. .1745
                      /gene="F18A8.1"
     CDS              86. .1615
                      /gene="F18A8.1"
                      /codon_start=1
                      /product="putative beta-ketoacyl-CoA synthase"
                      /db_xref="PID:g2760830"
                      /translation="MDVEQKKPLIESSDRNLPDFKKSVKLKYVKLGYHYLITHGMYLF
                      LSPLVLVIAAQISTFSVTDLRSLWEHLQYNLISVVVCSMLLVFLMTIYFMTRPRPVYL
                      VNFSCFKPDESRKCTKKIFMDRSKLTGSFTEENLEFQRKILQRSGLGESTYLPEAVLN
                      VPPHPCMKEARKEAETVMFGAIDELLAKTNVNPKDIGILIVNCSLFNPTPSLSAMVVN
                      HYKLRGNILSYNLGGMGCSAGLISIDLAKHLLHSIPNTTAMVISMENITLNWYFGNDR
                      SKLVSNCLFRMGGAAILLSNKRWDRRRSKYELVDTVRTHRGADDKCPGCITQEEDSAS
                      KIGVTLSKELMAVAGDALKTNITTLGPLVLPTSEQLLFPATLVGRKLFKMKIKPYIPD
                      FKLAFEHPCIHAGGRAVLDELEKNLKLTEWHMEPSRMTLYRFGNTSSSSLWYELAYSE
                      AKGRIKKGDRIWQIAFGSGFKCNSSVWRAVRSVNPKKEKNPWMDEIHEPPVEVPKVST
                      I"
     misc_feature     complement(1788. .1833)
                      /note="exon predicted by xgrail, quality
                      marginal_shadowexon"
     mRNA             complement(join(1934. .2789,2883. .2958,3053. .3247,
                      3350. .3535,3659. .3691,3774. .3872,3962. .4207,4289. .4606,
                      4715. .4772,4842. .5198,6264. .6493))
                      /gene="F18A8.2"
     gene             complement(1934. .6493)
                      /note="identical to GB:U06745"
                      /gene="F18A8.2"
     CDS              complement(join(1957. .2789,2883. .2958,3053. .3247,
```



```
                      3350. .3535,3659. .3691,3774. .3872,3962. .4207,4289. .4606,
                      4715. .4772,4842. .5198,6264. .6436))
                      /gene="F18A8.2"
                      /codon_start=1
                      /product="potassium transporter, AKT1"
                      /db_xref="PID:g2760831"
                      /translation="MRGGALLCGQVQDEIEQLSRESSHFSLSTGILPSLGARSNRRVK
                      LRRFVVSPYDHKYRIWEAFLVVLVVYTAWVSPFEFGFLRKPRPPLSITDNIVNAFFAI
                      DIIMTFFVGYLDKSTYLIVDDRKQIAFKYLRSWFLLDLVSTIPSEAAMRISSQSTGLF
                      NMLRLWRLRRVGALFARLEKDRNFNYFWVRCAKLVCVTLFAVHCAACFYYLIAARNSN
                      PAKTWIGANVANFLEESLWMRYVTSMYWSITTLTTVGYGDLHPVNTKEMIFDIFYMLF
                      NLGLTAYLIGNMTNLVVHGTSRTRNFRDTIQAASNFAHRNHLPPRLQDQMLAHLCLKY
                      RTDSEGLQQQETLDALPKAIRSSISHFLFYSLMDKVYLFRGVSNDLLFQLVSEMKAEY
                      FPPKEDVILQNEAPTDFYILVNGTADLVDVDTGTESIVREVKAGDIIGEIGVLCYRPQ
                      LPTVRTKRLCQLLRMNRTTFLNIIQANVGDGTIIMNNNLLQHLKEMNDPVMTNVLLEIE
                      NMLARGKMDLPLNLCFAAIREDDLLLWQLLKRGLDPNESDNNGRTPLHIAASKGTLNC
                      VLLLLEYHADPNCRDAEGSVPLWEAMVEGHEKVVKVLLEHGSTIDAGDVGHFACTAAE
                      QGNLKLLKEIVLHGGDVTRPRATGTSALHTAVCEENIEMVKYLLEQGADVNKQDMHGW
                      TPRDLAEQQGHEDIKALFREKLHERRVHIETSSSVPILKTGIRPLGRFTSEPNIRPAS
                      REVSFRIRETRARRKTNNPDNSLPGILANQSVPKNGLATVDEGRTGNPVRVTISCAEK
                      DDIAGKLVLLPGSFKELLELGSNKFGIVATKVMNKDNNAEIDDVDVIRDGDHLIFATD
                      S"
     repeat_region    5808. .5855
                      /rpt_family="AT_rich"
     repeat_region    6042. .6099
                      /rpt_family="AT_rich"
     misc_feature     6997. .7046
                      /note="exon predicted by xgrail, quality
                      marginal_shadowexon"
     misc_feature     7593. .7817
                      /note="exon predicted by xgrail, quality excellent"
     repeat_region    8035. .8072
                      /rpt_family="(TA)n"
     repeat_region    8278. .8318
                      /rpt_family="AT_rich"
     misc_feature     complement(8454. .8618)
                      /note="exon predicted by xgrail, quality
                      excellent_shadowexon"
     repeat_region    8825. .8858
                      /rpt_family="AT_rich"
     mRNA             join(<8924. .9214,9644. .9770,10250. .>10695)
                      /gene="F18A8.3"
     gene             <8924. .>10695
                      /gene="F18A8.3"
     CDS              join(8924. .9214,9644. .9770,10250. .10695)
                      /gene="F18A8.3"
                      /codon_start=1
                      /product="similar to barley ids-4 gene product"
                      /db_xref="PID:g2760832"
                      /translation="MKFGKSLSNQIEETLPEWRDKFLSYKELKKKLKLMEPRSVENRP
                      NKRSRSDSNSVDTDPTVGMTKEELDFISLLEDELEKFNSFFVEQEEETIIRLKELKDQ
```



```
                      VAKAKNSNEEMINIKKEIVDFRGEMVLLMNYSALNYTGLAKILKKYDKRTGALIRLPF
                      IQKVLQEPFFTTDLLNTFVKECEAMLDRLFPSNKSRNLDEEGEPTTSGMVKTGTDDSE
                      LLRVPKELSEIEYMESLYMKSTVSALKVLKEIRSGSSTVSVFSLPPLPASGLEDDSWK
                      KEVGVLEQVAK"
     repeat_region    11325. .11451
                      /rpt_family="AT_rich"
     misc_feature     complement(11553. .11663)
                      /note="exon predicted by xgrail, quality
                      excellent_shadowexon"
     mRNA             join(11752. .12231,12887. .13110,13186. .13357,13399. .>13472)
                      /gene="F18A8.4"
     gene             11752. .>13472
                      /gene="F18A8.4"
     CDS              join(11808. .12231,12887. .13110,13186. .13357,13399. .13472)
                      /gene="F18A8.4"
                      /note="unknown protein"
                      /codon_start=1
                      /db_xref="PID:g2760833"
                      /translation="MAYLAPISSSLSIFKNPQLSRPQFSSSSSPNPLFLRPRIQILSMT
                      MNKSPSLVVVAATTAAEKQKKRYPGESKGFVEEMRPVAMRLHTKDQAKEGEKETKSIE
                      ERPVAKWEPTVEGYLRFLVDSKLVYDTLELIIQDSNFPTYAEFKNTGLERAEKLSTDL
                      EWFKEQGYEIPEPTAPGKTYSQYLKELAEKDPQAFICHFYNIYFAHSAGGRMIGRKVA
                      ERILDNKELEFYKWDGELSQLLQNVREKLNKVAEEWTREEKNHCLEETEKSFKYSVSL
                      DISIRLVATKMMHIVFIPVLY"
     misc_feature     13562. .13626
                      /note="exon predicted by xgrail, quality
                      marginal_shadowexon"
     misc_feature     13682. .13784
                      /note="exon predicted by xgrail, quality good_shadowexon"
     mRNA             complement(join(<13995. .14354,14681. .>15280))
                      /gene="F18A8.5"
     gene             complement(<13995. .>15280)
                      /gene="F18A8.5"
     CDS              complement(join(13995. .14354,14681. .15280))
                      /gene="F18A8.5"
                      /note="unknown protein"
                      /codon_start=1
                      /db_xref="PID:g2760835"
                      /translation="MASAWKKDKNQKLLSPKTLISLLVLSIIFLSSLFFFFSNSSLQY
                      SNPNLSINSPYPIPPFDCFKCPQSKPIIANVVENLKYPFVYSLADLGNLPEKPHKNIV
                      RLLKGKPFRKPDISATIQEVLDSMRASGKNGIVVDVGANVGMASFAAAVMGFKVLAPE
                      PVFENLQRICDGIWFNRVASLVTVFEAAASDRTGDITFHKLVGRLDNSAVSEVGARLA
                      FKSNKEIAVQVKSIPLDKLIPPSQPVLLIKIDVQGWEYHVLKGAKKLLSGKPAEAPYL
                      IYEEDERLLTASNSSSSKEIRDFLKSVGYSKCSQHGTDAHCTKE"
     repeat_region    15169. .15217
                      /rpt_family="(GAA)n"
     misc_feature     15346. .15477
                      /note="exon predicted by xgrail, quality
...
Note: remainder of annotations omitted.
```



```
  Query Match           10.3%;  Score 18;  DB 23;  Length 95167;
  Best Local Similarity 100.0%;  Pred. No. 1.66e+01;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  36873 TGCAAAAATAAATTATTT 36890
          ||||||||||||||||||
Cp     59 TGCAAAAATAAATTATTT 42


RESULT   47
LOCUS       AC004548   105127 bp    DNA              PRI       10-APR-1998
DEFINITION  Homo sapiens PAC clone DJ1182N03 from 7q11.23-q21.1, complete
            sequence.
ACCESSION   AC004548
NID         g3046308
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 105127)
  AUTHORS   Leonard,S., Hinds,K. and Harper,M.
  TITLE     The sequence of Homo sapiens PAC clone DJ1182N03
  JOURNAL   Unpublished (1998)
REFERENCE   2  (bases 1 to 105127)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-APR-1998) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     SUBMITTED BY:  WUGSC
            Genome Sequencing Center
            Department of Genetics
            Washington University
            St. Louis MO 63108, USA
            http://genome.wustl.edu/gsc
            mailto:sapiens@watson.wustl.edu

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded or sequenced with an alternate
            chemistry; an attempt was made to resolve all sequencing problems,
            such as compressions and repeats; all regions were covered by
            sequence from more than one subclone; and the assembly was
            confirmed by restriction digest.

            MAPPING INFORMATION:
            The sequence of this clone was established as part of a mapping and
```

sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see http://www.nhgri.nih.gov/DIR/GTB/CHR7 , send mailto:egreen@nhgri.nih.gov , or see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
This clone was derived from human PAC library RPCI-5, prepared by Pieter de Jong and coworkers at Roswell Park Cancer Institute, using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor. For further details, see http://bacpac.med.buffalo.edu/
The clone is available from Genome Systems, Inc. (http://www.genomesystems.com).
VECTOR: pCYPAC2
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the right is DJ0608H12, 200 bp overlap. Actual start of this clone is at base position 1 of DJ1182N03; actual end is at 17513 of DJ0608H12.

FEATURES Location/Qualifiers
source 1. .105127
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q11.23-q21.1"
/clone_lib="RPCI-5"
/clone="DJ1182N03"
repeat_region 3309. .3606
/rpt_family="Alu"
repeat_region 5464. .6171
/rpt_family="L1"
repeat_region 6183. .8511
/rpt_family="L1"
repeat_region 8513. .9673
/rpt_family="L1"
repeat_region 9699. .9933
/rpt_family="L1"
repeat_region 9943. .10778
/rpt_family="L1"
repeat_region 11105. .11198
/rpt_family="L1"
repeat_region 11900. .12008
/rpt_family="MIR"
repeat_region 12056. .12358
/rpt_family="Alu"
repeat_region 12485. .13316
/rpt_family="L1"
repeat_region 13318. .13619
/rpt_family="Alu"

repeat_region 14206. .14322
/rpt_family="L1"
repeat_region 15998. .16426
/rpt_family="L1"
repeat_region 16761. .17244
/rpt_family="L1"
repeat_region 17681. .17963
/rpt_family="LOR1b"
repeat_region 17974. .20708
/rpt_family="L1"
repeat_region 20775. .21162
/rpt_family="L1"
repeat_region 21585. .21659
/rpt_family="MIR"
repeat_region 25781. .25974
/rpt_family="L2"
repeat_region 25976. .26271
/rpt_family="Alu"
repeat_region 26405. .26538
/rpt_family="L2"
misc_feature 28121. .28353
/note="match to EST (NID:g966746)"
misc_feature complement(28137. .28414)
/note="match to EST AA831575 (NID:g2904674) oc84f02.s1"
repeat_region 28707. .28794
/rpt_family="Alu"
repeat_region 28924. .28998
/rpt_family="Alu"
repeat_region 29256. .29441
/rpt_family="Retroviral"
repeat_region 29525. .29998
/rpt_family="Retroviral"
repeat_region 30197. .30549
/rpt_family="MaLR"
repeat_region 30560. .30629
/rpt_family="L2"
repeat_region 31041. .31172
/rpt_family="MER1_type"
repeat_region 33710. .34039
/rpt_family="Alu"
repeat_region 35980. .36317
/rpt_family="MaLR"
repeat_region 36830. .37284
/rpt_family="MaLR"
repeat_region 38331. .39035
/rpt_family="L1"
repeat_region 39360. .39705
/rpt_family="MER2_type"
repeat_region 40488. .40537
/rpt_family="MIR"
repeat_region 40866. .41014

/rpt_family="MER73-group"
repeat_region 41038. .41336
/rpt_family="L1"
repeat_region 41334. .41964
/rpt_family="L1"
repeat_region 42431. .42685
/rpt_family="MaLR"
repeat_region 43125. .43275
/rpt_family="MaLR"
repeat_region 44230. .44668
/rpt_family="L1"
repeat_region 44865. .44950
/rpt_family="Alu"
repeat_region 45427. .45588
/rpt_family="MER4-group"
repeat_region 46675. .47423
/rpt_family="Retroviral"
repeat_region 47651. .47955
/rpt_family="Alu"
repeat_region 47983. .48025
/rpt_family="L1"
repeat_region 49475. .49613
/rpt_family="L1"
repeat_region 49566. .49932
/rpt_family="L1"
repeat_region 50390. .50679
/rpt_family="MaLR"
repeat_region 51432. .51716
/rpt_family="Alu"
repeat_region 51896. .52190
/rpt_family="L1"
repeat_region 52196. .53015
/rpt_family="L2"
repeat_region 53450. .53608
/rpt_family="MER1_type"
repeat_region 53643. .53943
/rpt_family="Alu"
repeat_region 54108. .54189
/rpt_family="MIR"
repeat_region 55634. .55747
/rpt_family="MER1_type"
repeat_region 57577. .57652
/rpt_family="MIR"
repeat_region 57663. .58105
/rpt_family="L1"
repeat_region 58108. .58824
/rpt_family="L1"
repeat_region 58832...58900
/rpt_family="MER2_type"
repeat_region 58901. .59110
/rpt_family="Alu"

repeat_region 59111. .59206
/rpt_family="MER2_type"
repeat_region 59256. .60025
/rpt_family="MER2_type"
repeat_region 60255. .60312
/rpt_family="L2"
...
Note: remainder of annotations omitted.

Query Match 10.3%; Score 18; DB 26; Length 105127;
Best Local Similarity 100.0%; Pred. No. 1.66e+01;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 38510 AATAAATTATTTTTAACC 38527
||||||||||||||||||
Cp 53 AATAAATTATTTTTAACC 36

RESULT 48
LOCUS AB001523 122638 bp DNA PRI 03-JUL-1997
DEFINITION Homo sapiens gene for TMEM1 and PWP2,complete and partial cds.
ACCESSION AB001523
NID g2244605
KEYWORDS TMEM1; PWP2.
SOURCE Homo sapiens pre-pro-B cell cell_line:FLEB14-14 DNA, clone_lib:human BAC library clone:KB86A5.
ORGANISM Homo sapiens
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 122638)
AUTHORS Shimizu,N.
TITLE Direct Submission
JOURNAL Submitted (28-FEB-1997) to the DDBJ/EMBL/GenBank databases. Nobuyoshi Shimizu, Keio University School of Medicine, Department of Molecular Biology; 35 Shinanomachi, Shinjuku-ku, Tokyo 160, Japan (E-mail:shimizu@dmb.med.keio.ac.jp, Tel:03-3351-2370, Fax:03-3351-2370)
REFERENCE 2 (sites)
AUTHORS Nagamine,K., Kudoh ,J., Minoshima,S., Kawasaki,K., Asakawa,S., Ito,F. and Shimizu,N.
TITLE Nucleotide Sequence and Organization of the TMEM1 Gene
JOURNAL Unpublished (1997)
REFERENCE 3 (sites)
AUTHORS Nagamine,K., Kudoh,J., Kawasaki,K., Minoshima,S., Asakawa,S., Ito,F. and Shimizu,N.
TITLE Genomic organization and complete nucleotide sequence of the TMEM1 gene on human chromosome 21q22.3
JOURNAL Biochem. Biophys. Res. Commun. 235 (1), 185-190 (1997)
MEDLINE 97339461
FEATURES Location/Qualifiers

```
     source          1. .122638
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="PLEB14-14"
                     /cell_type="pre-pro-B cell"
                     /chromosome="21"
                     /clone="KB86A5"
                     /clone_lib="human BAC library"
                     /map="21q22.3"
     5'UTR           25441. .25577
     exon            25441. .25644
                     /number=1
                     /product="TMEM1"
     CDS             join(25578. .25644,45175. .45256,50876. .51011,65364. .65560,
                     68782. .68977,72187. .72298,76622. .76869,87408. .87554,
                     88123. .88237,89636. .89712,90777. .90868,92648. .92788,
                     93099. .93211,95872. .96386,97206. .97347,99886. .100044,
                     100783. .101012,102917. .103017,105007. .105133,
                     107147. .107317,111441. .111618,115865. .116059,
                     116380. .116618)
                     /codon_start=1
                     /product="TMEM1"
                     /db_xref="PID:d1021944"
                     /db_xref="PID:g2244606"
                     /translation="MDASEEPLPPVIYTMENKPIVTCAGDQNLPTSVYPTLSQQLPRE
                     PHEWRRSYGRAPKMIHLESNFVQPKEELLPKEGNKALLTPPPLHIYWTECCDTEVYKA
                     TVKDDLTKWQNVLKAHSSVDWLIVIVENDAKKNKTNILPRTSIVDKIRNDPCNKQSD
                     RCVVLSDPLKDSSRTQESWNAPLTKLRTLLLMSFTKNLGKFEDDMRTLREKRTEPGWS
                     FCEYFMVQEELAFVFEMLQQFEDALVQYDELDALFSQYVVNFGAGDGANWLTPFCQPV
                     KSWNGLILRKPIDMEKRESIQRREATLLDLRSYLFSRQCTLLLFLQRPWEVAQRALEL
                     LHNCVQELKLLEVSVPPGALDCWVPLSCLEVLQRIEGCCDRAQIDSNIAHTVGLWSYA
                     TEKLKSLGYLCGLVSEKGPNSEDLNRTVDLLAGLGAERPETANTAQSPYKKLKEALSS
                     VEAFEKHYLDLSHATIEMTTSIGRIRSAKFVGKDLAEFTMRKKAPQRAEIYLQGALKN
                     YLAEGWALPITHTRKQLAECQKHLGQIENYLQTSSLLASDHHLTEEERKHFCQEILDF
                     ASQPSDSPGHKIVLPKHSPAQLRDLHFDPSNAVVHVGGVLCVEITMTSQMPVPVHVEQ
                     IVVRVHFSIEKHSYRKTAEWLTKHKTSNGIINPPPETAPPPVSQNSLPALELVEMPER
                     SPSDNSLNTTGIICRNVKMLLRRQESSSSLEMPSGVALEEGAHVLRCSHVTLEPGANQ
                     ITFRTQAKEPGTYTLRQLCASVGSVWFVLPHIYPIVQYDVYSQEPQLHVEPLADSLLA
                     GIPQRVKFTVTTGHYTIKNGDSLQLSNAEAMLILCQAESRAVVYSNTREQSSEAALRI
                     QSSDKVTSISLPVAPAYHVIEPELEVLSLPSAPALGGESDMLGMAEPHRKHKDKQRTG
                     RCMVTTDHKVSIDCPWSIYSTVIALTFSVPFRTTHSLLSSGTRKYVQVCVQNLSELDF
                     QLSDSYLVDTGDSTDLQLVPLNTQSQQPIYSKQSVFFVWELKWTEEPPPSLHCRFSVG
                     FSPASEEQLSISLKPYTYEFKVENFFTLYNVKAEIFPPSGMEYCRTGSLCSLEVLITR
                     LSDLLEVDKDEALTESDEHFSTKLMYEVVDNSSNWAVCGKSCGVISMPVAARATHRVH
                     MEVMPLFAGYLPLPDVRLFKYLPHHSAHSSQLDADSWIENDSLSVDKHGDDQPDSSSL
                     KSRGSVHSACSSEHKGLPMPRLQALPAGQVFNSSSGTQVLVIPSQDDHVLEVSVT"
     exon            45175. .45256
                     /number=2
                     /product="TMEM1"
     exon            50876. .51011
                     /number=3
```



```
     exon            65364. .65560
                     /number=4
     exon            68782. .68977
                     /number=5
     exon            72187. .72298
                     /number=6
     exon            76622. .76869
                     /number=7
     exon            87408. .87554
                     /number=8
     exon            88123. .88237
                     /number=9
     exon            89636. .89712
                     /number=10
     exon            90777. .90868
                     /number=11
     exon            92648. .92788
                     /number=12
     exon            93099. .93211
                     /number=13
     exon            95872. .96386
                     /number=14
     exon            97206. .97347
                     /number=15
     exon            99886. .100044
                     /number=16
     exon            100783. .101012
                     /number=17
     exon            102917. .103017
                     /number=18
     exon            105007. .105133
                     /number=19
     exon            107147. .107317
                     /number=20
     exon            111441. .111618
                     /number=21
     exon            115865. .116059
                     /number=22
     exon            116380. .119637
                     /number=23
     3'UTR           116619. .119637
     5'UTR           120476. .120499
     exon            120476. .120517
                     /number=1
                     /product="PWP2"
     CDS             join(120500. .120517,122067. .>122179)
                     /codon_start=1
                     /product="PWP2"
                     /db_xref="PID:d1021945"
                     /db_xref="PID:g2244607"
                     /translation="MKFAYRFSNLLGTVYRRGNLNFTCNGNSVISPVGNRVTVFDLK"
```



```
     exon            122067. .122179
                     /number=2
                     /product="PWP2"
BASE COUNT    29257 a  27930 c  30176 g  35275 t
ORIGIN

  Query Match           10.3%;  Score 18;  DB 26;  Length 122638;
  Best Local Similarity 100.0%;  Pred. No. 1.66e+01;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  37763 TTCACTTGTTTTATTTTT 37780
          ||||||||||||||||||
Cp    161 TTCACTTGTTTTATTTTT 144


RESULT   49
LOCUS       HS19408    130019 bp    DNA             HTG       08-OCT-1998
DEFINITION  Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone 19408;
            HTGS phase 1.
ACCESSION   AL031769
NID         g3724217
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 130019)
  AUTHORS   Sulston,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-OCT-1998) Wellcome Trust Genome Campus, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquires:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 9, 1998 this sequence version replaced gi:3717969.
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and the release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc. Order of segments is not known; 800 n's separate
            segments. Unfinished: dJ19408  Contig_ID: 01485  acc=AL031769
            Length: 22324 bp Unfinished: dJ19408  Contig_ID: 00001
            acc=AL031769  Length: 58717 bp Unfinished: dJ19408  Contig_ID:
            01245  acc=AL031769  Length: 45574 bp Unfinished: dJ19408
            Contig_ID: 00558  acc=AL031769  Length: 1004 bp.
            ***                                                    ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                    ***
            * This sequence is unfinished. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
FEATURES              Location/Qualifiers
```



```
     source          1. .130019
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="19408"
BASE COUNT    40336 a  23496 c  23296 g  40491 t   2400 others
ORIGIN

  Query Match           10.3%;  Score 18;  DB 17;  Length 130019;
  Best Local Similarity 100.0%;  Pred. No. 1.66e+01;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  73574 TCTTTTGAGCTTTGCCTG 73591
          ||||||||||||||||||
Cp     26 TCTTTTGAGCTTTGCCTG 9


RESULT   50
LOCUS       AC002421   131047 bp    DNA             HTG       12-AUG-1997
DEFINITION  *** SEQUENCING IN PROGRESS *** Human Chromosome X; HTGS phase 1, 4
            unordered pieces.
ACCESSION   AC002421
NID         g2323248
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 131047)
  AUTHORS   Chen,E., Brownstein,B.H., States,D.J., Schlessinger,D. and
            Mazzarella,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished (1997)
REFERENCE   2  (bases 1 to 131047)
  AUTHORS   Brownstein,B.H., States,D.J. and Mazzarella,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-AUG-1997) Center for Genetics in Medicine, Box 8232,
            Washington University School of Medicine, 4566 Scott Avenue, St.
            Louis, MO 63110, USA
COMMENT     Current status of this project is available at:
            'http://genome.wustl.edu/cgm/seq_projects.html'
            Submitted by:
            Ellson Chen,
            Advanced Center for Genetic Technology,
            Applied Biosystems Division of Perlin Elmer Corp.,
            850 Lincoln Center Drive,
            Foster City, CA 94404 USA
            e-mail: ellson@genseq.apldbio.com

            and
```

Buddy Brownstein,
Center for Genetics in Medicine,
Washington University School of Medicine, Box 8232
4566 Scott Avenue,
St. Louis, MO 63110, USA
e-mail: buddy@genetics.wustl.edu

and

David J. States,
Institute for Biomedical Computing
Washington University in St. Louis
700 South Euclid Ave.
St. Louis, MO 63108 USA
e-mail: states@ibc.wustl.edu.

```
                ***                                                  ***
                *** WARNING: Phase 1 High Throughput Genome Sequence ***
                ***                                                  ***
                * This sequence is unfinished. It consists of 4 contigs for
                * which the order is not known; their order in this record is
                * arbitrary. In some cases, the exact lengths of the gaps
                * between the contigs are also unknown; these gaps are  presented
                * as runs of N as a convenience only. When sequencing is complete,
                * the sequence data presented in this record will be replaced
                *by a single finished sequence with the same accession number.
                *        1    22595: contig of 22595 bp in length
                *    22596    66883: contig of 44288 bp in length
                *    66884   122647: contig of 55764 bp in length
                *   122648   131047: contig of 8400 bp in length.
FEATURES             Location/Qualifiers
     source          1..131047
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="pWXD1"
                     /chromosome="X"
BASE COUNT    40590 a  23040 c  23923 g  43494 t
ORIGIN

  Query Match           10.3%;  Score 18;  DB 17;  Length 131047;
  Best Local Similarity 100.0%;  Pred. No. 1.66e+01;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  50176 AAAAATAATTTATTTTTG 50193
          ||||||||||||||||||
Qy     40 AAAAATAATTTATTTTTG 57


Search completed: Tue Dec 22 20:40:45 1998
Job time : 552 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Tue Dec 22 20:58:54 1998; MasPar time 213.22 Seconds
1075.039 Million cell updates/sec

Tabular output not generated.

Title: >US-08-530-112A-1320
Description: (1-128) from US08530112A.seq
Perfect Score: 128
N.A. Sequence: 1 GATCAGAATTTTTAAAGCAG.....AGAATATCTTGCTGGGCAAA 128
Comp: CTAGTCTTAAAAATTTCGTC.....TCTTATAGAACGACCCGTTT

Scoring table: TABLE jmetric
Gap 60

Nmatch STD : Dbase 0; Query 0

Searched: 2275026 seqs, 895388244 bases x 2

Post-processing: Minimum Match 0%
Listing first 1000 summaries

Database: embl-est55
1:em_est1 2:em_gss1 3:em_gss2 4:em_gss3
Database: genbank-est107
5:gb_est1 6:gb_est10 7:gb_est11 8:gb_est12 9:gb_est13
10:gb_est14 11:gb_est15 12:gb_est16 13:gb_est17
14:gb_est18 15:gb_est19 16:gb_est2 17:gb_est20
18:gb_est21 19:gb_est3 20:gb_est4 21:gb_est5 22:gb_est6
23:gb_est7 24:gb_est8 25:gb_est9 26:gb_gss1 27:gb_gss2
28:gb_gss3 29:gb_gss4

Statistics: Mean 7.890; Variance 1.408; scale 5.605

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
	1	128	100.0	128	7	HUMGS01470	Human HL60 3'directed	2.46e-235
	2	83	64.8	401	19	H78523	yu13c02.r1 Homo sapien	6.54e-134
	3	79	61.7	380	21	AA053439	z170b05.r1 Stratagene	4.24e-125
c	4	20	15.6	356	7	AA461465	zx68e06.s1 Soares tota	4.21e-07
c	5	19	14.8	363	26	AQ017960	CIT-HSP-2314N16.TF CIT	1.29e-05
c	6	19	14.8	528	21	W68051	zd42c09.s1 Soares feta	1.29e-05
	7	18	14.1	358	26	B76117	RPCI11-13K22.TP RPCI11	3.50e-04
	8	18	14.1	374	18	AI154735	ud36h11.r1 Soares mous	3.50e-04
	9	18	14.1	377	16	R92292	yq03b03.r1 Homo sapien	3.50e-04
	10	18	14.1	408	28	AQ181367	HS_3216_A2_E03_MR CIT	3.50e-04
c	11	18	14.1	411	5	R37679	yf61g04.s1 Homo sapien	3.50e-04
	12	18	14.1	439	27	B83406	RPCI11-13K22.TPB RPCI1	3.50e-04
c	13	18	14.1	510	16	R93755	yq35b06.r1 Homo sapien	3.50e-04
c	14	18	14.1	536	29	AQ245173	HS_2008_A1_E10_T7 CIT	3.50e-04
	15	17	13.3	111	17	AI023497	ov70h11.s1 Soares_test	8.40e-03
c	16	17	13.3	152	13	AA496694	ae30e05.s1 Gessler Wil	8.40e-03
	17	17	13.3	169	8	AA513115	nh78d10.s1 NCI_CGAP_Br	8.40e-03
	18	17	13.3	172	22	AA025396	ze90b03.s1 Soares feta	8.40e-03
	19	17	13.3	190	18	AI185506	qe54e09.x1 Soares_feta	8.40e-03
	20	17	13.3	202	12	AA775160	ac78e09.s1 Stratagene	8.40e-03
	21	17	13.3	241	11	AA702752	zi82g01.s1 Soares feta	8.40e-03
	22	17	13.3	250	12	AA779558	af24g10.s1 Soares tota	8.40e-03
	23	17	13.3	306	9	AA552090	ng48d01.s1 NCI_CGAP_Co	8.40e-03
	24	17	13.3	309	9	AA579359	nf29c03.s1 NCI_CGAP_Pr	8.40e-03
	25	17	13.3	315	6	AA007422	zh98g03.s1 Soares feta	8.40e-03
	26	17	13.3	336	13	AA804853	of44g06.s1 NCI_CGAP_CN	8.40e-03
	27	17	13.3	341	6	AA114829	zk88e05.s1 Soares preg	8.40e-03
	28	17	13.3	341	17	AI004353	ot95b09.x1 Soares_tota	8.40e-03
	29	17	13.3	343	5	R22948	yh31a06.s1 Homo sapien	8.40e-03
	30	17	13.3	343	8	AA485269	ab07g06.s1 Stratagene	8.40e-03
c	31	17	13.3	353	23	AA200434	mu35g12.r1 Soares 2NbM	8.40e-03
	32	17	13.3	360	28	AQ098400	HS_3046_A2_E12_MF CIT	8.40e-03
	33	17	13.3	366	11	AA693353	ah19e01.s1 Gessler Wil	8.40e-03
	34	17	13.3	368	14	AA953621	oq65a08.s1 NCI_CGAP_Ki	8.40e-03
	35	17	13.3	376	15	AI056065	ox46d08.s1 Soares_tota	8.40e-03
	36	17	13.3	377	14	AA864358	oh54f06.s1 NCI_CGAP_GC	8.40e-03
	37	17	13.3	378	10	AA648400	ns22a01.s1 NCI_CGAP_GC	8.40e-03
	38	17	13.3	381	12	AA766793	oa35f06.s1 NCI_CGAP_GC	8.40e-03
	39	17	13.3	383	17	AI034146	ow15c06.x1 Soares_para	8.40e-03
c	40	17	13.3	386	15	AI006803	ua74b07.r1 Soares 2NbM	8.40e-03
c	41	17	13.3	388	17	AA910013	ol40a12.s1 Soares_NFL_	8.40e-03
	42	17	13.3	401	22	W02793	zc64c10.s1 Soares feta	8.40e-03

	43	17	13.3	410	8	AA486288	ab35h08.s1 Stratagene	8.40e-03
c	44	17	13.3	413	17	AI025564	ov87c05.s1 Soares_test	8.40e-03
	45	17	13.3	416	16	H10401	ym08e07.s1 Homo sapien	8.40e-03
	46	17	13.3	419	22	N57697	yv55b06.s1 Soares feta	8.40e-03
	47	17	13.3	420	8	AA490320	aa51c09.s1 NCI_CGAP_GC	8.40e-03
	48	17	13.3	422	9	AA535623	nf87f03.s1 NCI_CGAP_Co	8.40e-03
	49	17	13.3	423	10	AA424324	zv82a05.s1 Soares tota	8.40e-03
	50	17	13.3	424	21	W04711	zb94d02.s1 Soares para	8.40e-03
c	51	17	13.3	425	12	AA761490	nz27a03.s1 NCI_CGAP_GC	8.40e-03
	52	17	13.3	430	11	AA678487	ah06e09.s1 Gessler Wil	8.40e-03
	53	17	13.3	431	18	AI149572	qc60b04.x1 Soares_plac	8.40e-03
	54	17	13.3	432	19	H82910	yq46f12.r1 Homo sapien	8.40e-03
	55	17	13.3	434	13	AA205308	zq79a10.s1 Stratagene	8.40e-03
	56	17	13.3	436	18	AI167824	ox90e09.x1 Soares_para	8.40e-03
	57	17	13.3	438	10	AA603729	np24h11.s1 NCI_CGAP_GC	8.40e-03
	58	17	13.3	441	22	AA148861	zo01d12.s1 Stratagene	8.40e-03
	59	17	13.3	441	9	AA610141	np54h03.s1 NCI_CGAP_Br	8.40e-03
c	60	17	13.3	442	16	R98191	yq66e11.r1 Homo sapien	8.40e-03
	61	17	13.3	443	17	AI129996	qc50c12.x1 Soares_preg	8.40e-03
	62	17	13.3	446	18	AI149727	qf32b09.x1 Soares_test	8.40e-03
c	63	17	13.3	448	20	W94384	zd77b01.r1 Soares feta	8.40e-03
	64	17	13.3	450	18	AI095615	qb24d01.x1 Soares_preg	8.40e-03
	65	17	13.3	451	17	AI126023	qc50g09.x1 Soares_preg	8.40e-03
	66	17	13.3	456	19	N30763	yw74f04.s1 Homo sapien	8.40e-03
	67	17	13.3	457	19	N20849	yx51f09.s1 Homo sapien	8.40e-03
c	68	17	13.3	457	29	AQ236055	HS_2052_A2_D11_MR CIT	8.40e-03
	69	17	13.3	458	11	AA682773	ah52h04.s1 Soares test	8.40e-03
	70	17	13.3	458	18	AI078179	os12e11.x1 Soares_feta	8.40e-03
	71	17	13.3	458	18	AI086786	ox77c05.x1 Soares_sene	8.40e-03
	72	17	13.3	458	17	AI027706	ov89g10.x1 Soares_test	8.40e-03
	73	17	13.3	461	21	W15581	zb92c11.s1 Soares para	8.40e-03
	74	17	13.3	467	17	AI126199	qd86f02.x1 Soares_test	8.40e-03
	75	17	13.3	468	6	AA001737	zh86f08.s1 Soares feta	8.40e-03
	76	17	13.3	468	26	B71890	RPCI11-11K4.TP RPCI11	8.40e-03
	77	17	13.3	468	5	T78293	yc99b05.r1 Homo sapien	8.40e-03
	78	17	13.3	472	6	W90372	zh78e11.s1 Soares feta	8.40e-03
	79	17	13.3	472	19	N67842	za05g12.s1 Homo sapien	8.40e-03
	80	17	13.3	476	6	AA433927	zw52e08.s1 Soares tota	8.40e-03
c	81	17	13.3	477	26	B42936	HS-1056-B2-D05-MF.abi	8.40e-03
	82	17	13.3	480	20	W67753	zd38g01.s1 Soares feta	8.40e-03
	83	17	13.3	492	19	N73273	za09a06.s1 Homo sapien	8.40e-03
	84	17	13.3	493	18	AI105210	EST214499 Normalized r	8.40e-03
	85	17	13.3	495	23	W92013	zh47g10.s1 Soares feta	8.40e-03
	86	17	13.3	498	18	AI038928	ox96h02.x1 Soares_sene	8.40e-03
	87	17	13.3	501	13	AA811726	ob84c07.s1 NCI_CGAP_GC	8.40e-03
	88	17	13.3	508	17	AI093170	qa97h09.x1 Soares_feta	8.40e-03
c	89	17	13.3	511	8	AA130788	zo13f04.r1 Stratagene	8.40e-03
	90	17	13.3	514	13	AA854206	aj64f03.s1 Soares para	8.40e-03
	91	17	13.3	515	13	AA643184	nr61b08.s1 NCI_CGAP_Ly	8.40e-03
c	92	17	13.3	528	18	AI094716	qa08e01.x1 NCI_CGAP_Br	8.40e-03
c	93	17	13.3	529	10	AA620152	vo63b06.r1 Soares mous	8.40e-03

	94	17	13.3	530	19	H97860	yw09f06.s1 Homo sapien	8.40e-03
	95	17	13.3	548	21	W45570	zc26f08.s1 Soares sene	8.40e-03
	96	17	13.3	552	10	AA630321	ac08g03.s1 Stratagene	8.40e-03
	97	17	13.3	553	18	AI147025	qa82b07.x1 Soares_feta	8.40e-03
	98	17	13.3	554	12	AA780210	af52f01.s1 Soares tota	8.40e-03
	99	17	13.3	560	19	H97575	yx03d01.s1 Homo sapien	8.40e-03
	100	17	13.3	564	13	AA216596	zg95e06.s1 Stratagene	8.40e-03
	101	17	13.3	567	20	AA004205	zh97b07.s1 Soares feta	8.40e-03
	102	17	13.3	571	19	N25768	yx84b03.s1 Homo sapien	8.40e-03
	103	17	13.3	582	18	AI056248	oy90g09.x1 Soares_feta	8.40e-03
	104	17	13.3	600	8	AA527268	ng39e01.s1 NCI_CGAP_Co	8.40e-03
	105	17	13.3	636	20	W74607	zd77b01.s1 Soares feta	8.40e-03
	106	16	12.5	174	24	AA266720	mz98e08.r1 Soares mous	1.75e-01
	107	16	12.5	179	16	H48810	yq80d06.s1 Homo sapien	1.75e-01
	108	16	12.5	185	15	AI058564	UI-R-C1-kv-a-11-0-UI.s	1.75e-01
c	109	16	12.5	203	8	AA491953	ng16g06.s1 NCI_CGAP_Li	1.75e-01
	110	16	12.5	210	19	AI785979	A. thaliana transcribe	1.75e-01
c	111	16	12.5	224	9	AA570109	ne57d03.s1 NCI_CGAP_Co	1.75e-01
c	112	16	12.5	243	9	AA533707	nj92b04.s1 NCI_CGAP_Pr	1.75e-01
c	113	16	12.5	291	25	AA358180	EST67129 Fetal lung II	1.75e-01
c	114	16	12.5	297	12	AA721114	nz67c02.s1 NCI_CGAP_GC	1.75e-01
c	115	16	12.5	300	25	AA350881	EST58391 Infant brain	1.75e-01
c	116	16	12.5	300	6	C08025	C.elegans cDNA clone y	1.75e-01
c	117	16	12.5	321	11	AA723492	zg69f12.s1 Soares feta	1.75e-01
c	118	16	12.5	322	27	AQ067492	HS_2237_B2_C06_MR CIT	1.75e-01
c	119	16	12.5	323	7	C24351	Dictyostelium discoide	1.75e-01
	120	16	12.5	335	13	AA860358	aj59b07.s1 Soares test	1.75e-01
c	121	16	12.5	336	13	AA883066	am18g01.s1 Soares NFL	1.75e-01
c	122	16	12.5	338	5	R23149	yh27g12.r1 Homo sapien	1.75e-01
c	123	16	12.5	338	5	R23136	yh27e12.r1 Homo sapien	1.75e-01
c	124	16	12.5	343	14	AA894885	nz47g01.s1 NCI_CGAP_Pr	1.75e-01
	125	16	12.5	344	19	H67731	yr72g05.s1 Homo sapien	1.75e-01
	126	16	12.5	346	5	R34243	yh80f12.s1 Homo sapien	1.75e-01
c	127	16	12.5	360	9	C66499	C.elegans cDNA clone y	1.75e-01
c	128	16	12.5	360	9	C39594	C.elegans cDNA clone y	1.75e-01
	129	16	12.5	367	14	H34522	EST111531 Rat PC-12 ce	1.75e-01
	130	16	12.5	367	10	AA613906	no93f02.s1 NCI_CGAP_Pr	1.75e-01
c	131	16	12.5	367	28	AQ105917	HS_3057_A2_G06_MF CIT	1.75e-01
c	132	16	12.5	370	22	AA136428	zn89h01.r1 Stratagene	1.75e-01
c	133	16	12.5	375	19	H74008	yu05g03.s1 Homo sapien	1.75e-01
	134	16	12.5	377	11	AA112250	zm64g12.s1 Stratagene	1.75e-01
c	135	16	12.5	386	19	H78891	yu11a08.r1 Homo sapien	1.75e-01
c	136	16	12.5	393	13	AA841732	MB3D6V6E08T3 Brugia ma	1.75e-01
	137	16	12.5	403	28	AQ164988	HS_3016_A1_D09_MR CIT	1.75e-01
c	138	16	12.5	406	23	AA067181	mm38a09.r1 Stratagene	1.75e-01
c	139	16	12.5	410	5	R23901	yh28c12.r1 Homo sapien	1.75e-01
	140	16	12.5	419	28	AQ180859	HS_3225_A2_A05_MR CIT	1.75e-01
	141	16	12.5	425	12	AA766609	oa33h03.s1 NCI_CGAP_GC	1.75e-01
	142	16	12.5	426	6	AA436284	zv22b12.r1 Soares NhHM	1.75e-01
	143	16	12.5	429	22	AA165561	zo81d03.r1 Stratagene	1.75e-01
	144	16	12.5	430	19	H90861	yu85e06.s1 Homo sapien	1.75e-01

c	145	16	12.5	430	6	AA411297	zv24f05.r1 Soares NhHM	1.75e-01
c	146	16	12.5	431	27	AQ053132	CIT-HSP-2342M15.TF CIT	1.75e-01
c	147	16	12.5	433	18	AI156297	ue09a10.r1 Soares mous	1.75e-01
	148	16	12.5	443	17	AU012855	Schizosaccharomyces po	1.75e-01
	149	16	12.5	444	7	AA442254	zv61g11.s1 Soares test	1.75e-01
c	150	16	12.5	449	27	AQ035243	CIT-HSP-2333H6.TF CIT-	1.75e-01
c	151	16	12.5	451	6	AA436158	zv22b12.s1 Soares NhHM	1.75e-01
	152	16	12.5	453	13	AA804755	ob98e09.s1 NCI_CGAP_GC	1.75e-01
	153	16	12.5	454	20	W88478	zh72a08.r1 Soares feta	1.75e-01
	154	16	12.5	457	8	AA459476	aa27b04.r1 NCI_CGAP_GC	1.75e-01
	155	16	12.5	457	12	AA754456	97SN1778 Rice Immature	1.75e-01
c	156	16	12.5	458	9	AA607928	vm39d07.r1 Knowles Sol	1.75e-01
c	157	16	12.5	466	29	AQ137252	HS_3066_B1_H01_MF CIT	1.75e-01
	158	16	12.5	468	27	AQ082673	RPCI11-54A6.TJ RPCI11	1.75e-01
	159	16	12.5	471	28	AQ129568	HS_2248_B1_C09_MR CIT	1.75e-01
c	160	16	12.5	474	15	AA947748	oq32g04.s1 NCI_CGAP_GC	1.75e-01
	161	16	12.5	478	10	SSC16F10	Sus scrofa mRNA; expre	1.75e-01
c	162	16	12.5	484	18	AI170857	EST216798 Normalized r	1.75e-01
c	163	16	12.5	494	28	AQ183672	HS_3184_B2_E02_T7 CIT	1.75e-01
	164	16	12.5	496	22	AA190036	mt91g05.r1 Soares mous	1.75e-01
	165	16	12.5	506	28	AQ192001	HS_3229_B2_E06_T7 CIT	1.75e-01
c	166	16	12.5	506	10	AA620016	v176g04.r1 Knowles Sol	1.75e-01
	167	16	12.5	511	17	AA993218	ou02b12.s1 Soares_tota	1.75e-01
c	168	16	12.5	511	23	AA217496	mu87h09.r1 Soares mous	1.75e-01
c	169	16	12.5	514	11	AA677682	zj72f10.s1 Soares feta	1.75e-01
	170	16	12.5	518	7	AA522396	vi45g02.r1 Beddington	1.75e-01
	171	16	12.5	519	28	AQ136742	HS_2192_A2_D02_MR CIT	1.75e-01
c	172	16	12.5	524	23	AA230668	mw06e10.r1 Soares mous	1.75e-01
	173	16	12.5	527	27	AQ003844	CIT-HSP-2289F18.TF CIT	1.75e-01
c	174	16	12.5	528	18	AI176804	EST220400 Normalized r	1.75e-01
c	175	16	12.5	546	27	AQ040572	CIT-HSP-2325P16.TV CIT	1.75e-01
c	176	16	12.5	558	14	AA990167	ua60e07.r1 Soares 2NbM	1.75e-01
	177	16	12.5	568	23	AA088048	mo01e07.r1 Stratagene	1.75e-01
c	178	16	12.5	579	13	C87773	Mus musculus fertilize	1.75e-01
	179	16	12.5	584	13	C85074	Mus musculus fertilize	1.75e-01
	180	16	12.5	610	14	AA889813	al51d08.s1 Soares_NFL_	1.75e-01
	181	16	12.5	639	27	B90337	CIT-HSP-2173E6.TR CIT-	1.75e-01
	182	16	12.5	661	27	AQ051248	nbxb0001cE10r CUGI Ric	1.75e-01
c	183	16	12.5	712	28	AQ116188	RPCI11-42G14.TK RPCI11	1.75e-01
c	184	16	12.5	719	27	AQ026950	CIT-HSP-2323K6.TR CIT-	1.75e-01
	185	16	12.5	835	26	B13008	T27M5-Sp6 TAMU Arabido	1.75e-01
c	186	16	12.5	945	10	W12977	ma74g10.r1 Soares mous	1.75e-01
	187	16	12.5	986	26	B06528	cSRL-81g12-u cSRL flow	1.75e-01
c	188	16	12.5	1000	8	C23374	Jpanese flounder liver	1.75e-01
c	189	15	11.7	113	21	W25448	zb90f06.r1 Soares sene	3.09e+00
	190	15	11.7	137	10	AA424457	zv82g11.r1 Soares tota	3.09e+00
c	191	15	11.7	148	26	FR0004880	F.rubripes GSS sequenc	3.09e+00
	192	15	11.7	163	17	AI040696	ox18d03.s1 Soares_feta	3.09e+00
c	193	15	11.7	163	12	AA776958	zf23b05.s1 Soares feta	3.09e+00
c	194	15	11.7	169	15	AA923507	ol66e09.s1 NCI_CGAP_Ki	3.09e+00
c	195	15	11.7	172	6	AA055787	zl75h11.r1 Stratagene	3.09e+00

	196	15	11.7	174	11	C83866	Dictyostelium discoide	3.09e+00
	197	15	11.7	178	13	AA832424	oc99d11.s1 NCI_CGAP_GC	3.09e+00
c	198	15	11.7	179	8	AA282018	zs89c01.s1 NCI_CGAP_GC	3.09e+00
	199	15	11.7	180	24	AA299646	EST12207 Uterus tumor	3.09e+00
	200	15	11.7	185	25	AA330053	EST33921 Embryo, 12 we	3.09e+00
	201	15	11.7	186	7	AA537606	vj98f05.r1 Knowles Sol	3.09e+00
c	202	15	11.7	186	16	R82900	e17-15.2.3- Rattus nor	3.09e+00
	203	15	11.7	196	22	AA020557	mh62f01.r1 Soares mous	3.09e+00
	204	15	11.7	197	16	H44062	yn58a06.s1 Homo sapien	3.09e+00
	205	15	11.7	199	24	AA299191	EST11678 Uterus Homo s	3.09e+00
	206	15	11.7	199	12	AA745087	nv43g10.r1 NCI_CGAP_Ew	3.09e+00
	207	15	11.7	201	27	B64302	CIT-HSP-2022G6.TF CIT-	3.09e+00
	208	15	11.7	204	27	B80696	CIT-HSP-2050M13.TR CIT	3.09e+00
c	209	15	11.7	218	16	H19504	yn59e03.s1 Homo sapien	3.09e+00
c	210	15	11.7	224	14	AA895142	vx47b03.r1 Stratagene	3.09e+00
c	211	15	11.7	233	20	HSC0WB112	H. sapiens partial cDN	3.09e+00
	212	15	11.7	239	24	AA301376	EST14332 Testis tumor	3.09e+00
c	213	15	11.7	239	20	W24003	zb47h05.r1 Soares feta	3.09e+00
	214	15	11.7	240	9	C66012	C.elegans cDNA clone y	3.09e+00
	215	15	11.7	243	24	AA322373	EST25001 Cerebellum II	3.09e+00
c	216	15	11.7	243	12	AA751441	ISYS0111 Rice Immature	3.09e+00
c	217	15	11.7	244	25	AA364572	EST75179 Pineal gland	3.09e+00
	218	15	11.7	248	19	N71780	yz94e05.s1 Homo sapien	3.09e+00
	219	15	11.7	255	6	AA432843	vd75b01.r1 Beddington	3.09e+00
c	220	15	11.7	259	25	AA357995	EST66821 Fetal lung II	3.09e+00
c	221	15	11.7	260	24	AA322367	EST25080 Cerebellum II	3.09e+00
	222	15	11.7	261	14	AA849924	EST192691 Normalized r	3.09e+00
c	223	15	11.7	261	16	T36029	EST95952 Homo sapiens	3.09e+00
c	224	15	11.7	264	5	T73238	yc71a02.s1 Homo sapien	3.09e+00
	225	15	11.7	269	19	N31170	yx64d04.r1 Homo sapien	3.09e+00
c	226	15	11.7	270	13	AA609153	af11h06.s1 Soares test	3.09e+00
c	227	15	11.7	272	17	AA995206	ou90d11.s1 NCI_CGAP_Ki	3.09e+00
c	228	15	11.7	272	8	AA235988	zs41g10.s1 Soares NhHM	3.09e+00
c	229	15	11.7	272	25	AA341454	EST46845 Fetal kidney	3.09e+00
c	230	15	11.7	272	17	AA993830	ot93g01.r1 Soares_tota	3.09e+00
	231	15	11.7	273	17	AA995388	or74e01.s1 NCI_CGAP_Lu	3.09e+00
c	232	15	11.7	273	6	C20659	HUMGS0004664, Human Ge	3.09e+00
c	233	15	11.7	275	11	AA580944	nc83e05.s1 NCI_CGAP_GC	3.09e+00
c	234	15	11.7	276	25	AA360430	EST69824 T-cell lympho	3.09e+00
c	235	15	11.7	279	14	AA911956	oh88f10.s1 NCI_CGAP_Co	3.09e+00
c	236	15	11.7	279	16	H41653	yn94h06.s1 Homo sapien	3.09e+00
c	237	15	11.7	279	25	AA331239	EST35069 Embryo, 6 wee	3.09e+00
c	238	15	11.7	283	25	AA336106	EST40621 Epididymus Ho	3.09e+00
c	239	15	11.7	283	25	AA353357	EST61737 Activated T-c	3.09e+00
c	240	15	11.7	284	25	AA348966	EST55505 Infant adrena	3.09e+00
c	241	15	11.7	286	25	AA360591	EST69832 T-cell lympho	3.09e+00
c	242	15	11.7	286	25	AA338033	EST42847 Endometrial t	3.09e+00
	243	15	11.7	292	17	AI120272	ub68d07.r1 Soares mous	3.09e+00
c	244	15	11.7	292	5	T06017	EST03906 Homo sapiens	3.09e+00
c	245	15	11.7	294	25	AA341548	EST47083 Fetal kidney	3.09e+00
c	246	15	11.7	294	21	N95636	zb64g10.s1 Soares feta	3.09e+00

c	247	15	11.7	294	24	AA328152	EST31593 Embryo, 12 we	3.09e+00
	248	15	11.7	295	27	AQ087770	HS_2234_B2_D08_MF CIT	3.09e+00
	249	15	11.7	296	7	AA469641	vg53b07.r1 Beddington	3.09e+00
c	250	15	11.7	297	27	AQ037884	CIT-HSP-2316L13.TR CIT	3.09e+00
c	251	15	11.7	298	28	AQ094278	HS_3022_B1_H08_MR CIT	3.09e+00
c	252	15	11.7	298	7	RICS13976A	Rice cDNA, partial seq	3.09e+00
c	253	15	11.7	298	21	W33146	zc08h09.r1 Soares para	3.09e+00
	254	15	11.7	300	19	H88336	yw19d01.s1 Homo sapien	3.09e+00
c	255	15	11.7	300	6	C07270	C.elegans cDNA clone y	3.09e+00
c	256	15	11.7	300	9	C31759	C.elegans cDNA clone y	3.09e+00
c	257	15	11.7	300	9	C30486	C.elegans cDNA clone y	3.09e+00
c	258	15	11.7	300	9	C33211	C.elegans cDNA clone y	3.09e+00
c	259	15	11.7	300	9	C57946	C.elegans cDNA clone y	3.09e+00
c	260	15	11.7	300	6	C11022	C.elegans cDNA clone y	3.09e+00
c	261	15	11.7	300	9	C58070	C.elegans cDNA clone y	3.09e+00
c	262	15	11.7	300	9	C38229	C.elegans cDNA clone y	3.09e+00
c	263	15	11.7	301	7	RICS13718A	Rice cDNA, partial seq	3.09e+00
c	264	15	11.7	301	7	HUMGS01787	Human HL60 3'directed	3.09e+00
c	265	15	11.7	304	24	AA297123	EST11270 Uterus Homo s	3.09e+00
c	266	15	11.7	304	28	AQ220236	HS_2262_B1_G09_T7 CIT	3.09e+00
c	267	15	11.7	305	24	AA318632	EST20812 Spleen I Homo	3.09e+00
c	268	15	11.7	306	11	AA701899	zi64e03.s1 Soares feta	3.09e+00
	269	15	11.7	307	13	AA842656	MBAPC29A01T3 Brugia ma	3.09e+00
c	270	15	11.7	309	17	AI000648	os64b02.s1 NCI_CGAP_Br	3.09e+00
	271	15	11.7	310	5	R09800	yf28e01.r1 Homo sapien	3.09e+00
	272	15	11.7	310	27	B74157	CIT-HSP-2028L7.TR CIT-	3.09e+00
	273	15	11.7	310	13	AA840671	CLF40 Leaf cDNA librar	3.09e+00
	274	15	11.7	310	27	AQ005419	CIT-HSP-2296A13.TF CIT	3.09e+00
c	275	15	11.7	310	11	AA687964	nv13h08.s1 NCI_CGAP_Pr	3.09e+00
c	276	15	11.7	310	21	W24728	zb58f10.r1 Soares feta	3.09e+00
c	277	15	11.7	311	24	AA300683	EST13531 Testis tumor	3.09e+00
c	278	15	11.7	311	13	AA182716	sp47g08.r1 Stratagene	3.09e+00
	279	15	11.7	312	17	AI120357	ub69f09.r1 Soares mous	3.09e+00
c	280	15	11.7	313	25	AA378493	EST91728 Synovial sarc	3.09e+00
	281	15	11.7	315	5	T11995	A894R Homo sapiens cDN	3.09e+00
c	282	15	11.7	315	11	AA659197	nu12d04.s1 NCI_CGAP_Pr	3.09e+00
c	283	15	11.7	316	14	H31613	EST105826 Rat PC-12 ce	3.09e+00
c	284	15	11.7	316	28	AQ102110	HS_3004_A2_G11_MR CIT	3.09e+00
c	285	15	11.7	316	8	AA547693	EST188713 B. glabrata	3.09e+00
	286	15	11.7	317	16	H04367	yj20f10.r1 Homo sapien	3.09e+00
c	287	15	11.7	317	27	AQ051834	RPCI11-53L1.TJ RPCI11	3.09e+00
c	288	15	11.7	318	25	AA370382	EST81997 Prostate glan	3.09e+00
c	289	15	11.7	318	25	AA377726	EST90363 Synovial sarc	3.09e+00
c	290	15	11.7	318	25	AA361312	EST70596 T-cell lympho	3.09e+00
c	291	15	11.7	318	16	H21456	yl32d01.r1 Homo sapien	3.09e+00
c	292	15	11.7	319	11	AA681235	vr74a06.s1 Knowles Sol	3.09e+00
	293	15	11.7	321	17	AI070348	UI-R-Y0-lv-f-02-0-UI.s	3.09e+00
	294	15	11.7	321	6	AA404958	zu19b07.s1 Soares NhHM	3.09e+00
c	295	15	11.7	321	26	AQ020617	CIT-HSP-2304B23.TF CIT	3.09e+00
c	296	15	11.7	321	27	AQ077932	CIT-HSP-2354H17.TF CIT	3.09e+00
c	297	15	11.7	322	8	AA256112	zs29f03.r1 NCI_CGAP_GC	3.09e+00

c	298	15	11.7	322	5	T90035	ye13c05.s1 Homo sapien	3.09e+00
	299	15	11.7	323	24	AA317795	EST19803 Retina II Hom	3.09e+00
c	300	15	11.7	323	25	AA363585	EST73769 Bone VII Homo	3.09e+00
c	301	15	11.7	323	10	HUMNK778	Human keratinocyte cDN	3.09e+00
	302	15	11.7	326	11	AA081451	zn06g08.r1 Stratagene	3.09e+00
c	303	15	11.7	328	8	AA542321	vk53e04.r1 Stratagene	3.09e+00
	304	15	11.7	330	23	AA053361	zf57e04.r1 Soares reti	3.09e+00
c	305	15	11.7	332	20	HSB81G102	H. sapiens partial cDN	3.09e+00
c	306	15	11.7	332	5	T18590	hbc2153 Homo sapiens c	3.09e+00
	307	15	11.7	333	16	H48804	yq80c07.s1 Homo sapien	3.09e+00
	308	15	11.7	333	26	B74965	RPCI11-12D9.TP RPCI11	3.09e+00
	309	15	11.7	333	15	AA471152	PMY2271 KG1-a Lambda Z	3.09e+00
c	310	15	11.7	333	8	AA082404	zn38h11.r1 Stratagene	3.09e+00
	311	15	11.7	334	17	AI112473	UI-R-Y0-mm-g-13-0-UI.s	3.09e+00
	312	15	11.7	334	5	R10815	yf36h05.r1 Homo sapien	3.09e+00
c	313	15	11.7	335	22	N21790	TNSSFH0014 TBN95TM-SSF	3.09e+00
c	314	15	11.7	335	5	T73763	yc53g10.r1 Homo sapien	3.09e+00
c	315	15	11.7	336	23	AA247445	csg2867.seq.F Human fe	3.09e+00
	316	15	11.7	337	14	C93399	Dictyostelium discoide	3.09e+00
c	317	15	11.7	339	24	AA304319	EST17087 Aorta endothe	3.09e+00
	318	15	11.7	340	15	AI009535	EST203986 Normalized r	3.09e+00
c	319	15	11.7	340	25	AA336680	EST41309 Endometrial t	3.09e+00
	320	15	11.7	341	26	AQ011957	HS_2190_A1_MR_D06 CIT	3.09e+00
c	321	15	11.7	341	6	CELK009A8R	C.elegans cDNA clone y	3.09e+00
	322	15	11.7	342	9	HSE599426	Homo sapiens mRNA; exp	3.09e+00
	323	15	11.7	343	21	R94165	yt74e10.r1 Soares feta	3.09e+00
c	324	15	11.7	344	6	AA434415	zw31d03.s1 Soares ovar	3.09e+00
	325	15	11.7	345	14	C90007	Dictyostelium discoide	3.09e+00
	326	15	11.7	345	15	AA917441	ol51d01.s1 Soares_NFL_	3.09e+00
	327	15	11.7	346	9	AA571862	vl95h01.r1 Knowles Sol	3.09e+00
c	328	15	11.7	346	8	AA195695	zr33a07.s1 Soares NhHM	3.09e+00
	329	15	11.7	347	14	AA848140	oe05c02.s1 NCI_CGAP_Ov	3.09e+00
c	330	15	11.7	347	28	AQ193432	CIT-HSP-2380P3.TF CIT-	3.09e+00
	331	15	11.7	349	5	R32520	yh73h03.s1 Homo sapien	3.09e+00
	332	15	11.7	350	14	C92834	Dictyostelium discoide	3.09e+00
	333	15	11.7	353	27	AQ030082	RPCI11-37P24.TP RPCI11	3.09e+00
	334	15	11.7	353	22	AA092609	l15885.seq.F Fetal hea	3.09e+00
	335	15	11.7	354	13	AA199846	zq52b08.s1 Stratagene	3.09e+00
	336	15	11.7	354	22	AA211473	zp75h08.s1 Stratagene	3.09e+00
c	337	15	11.7	354	9	AA230164	nc38e03.s1 NCI_CGAP_Pr	3.09e+00
	338	15	11.7	355	18	AI070453	oz05h05.x1 Soares_feta	3.09e+00
	339	15	11.7	355	27	B59844	CIT-HSP-345G15.TPC CIT	3.09e+00
c	340	15	11.7	355	25	AA372112	EST83975 Parathyroid g	3.09e+00
c	341	15	11.7	355	17	AI071929	UI-R-C1-lo-c-08-0-UI.s	3.09e+00
	342	15	11.7	357	8	AA278323	zs80e08.s1 NCI_CGAP_GC	3.09e+00
c	343	15	11.7	357	25	AA355227	EST63713 Jurkat T-cell	3.09e+00
c	344	15	11.7	357	14	AA865595	oh51f06.s1 NCI_CGAP_GC	3.09e+00
c	345	15	11.7	358	8	AA243763	zr67d12.s1 Soares NhHM	3.09e+00
c	346	15	11.7	359	25	AA334543	EST38775 Embryo, 9 wee	3.09e+00
	347	15	11.7	360	5	T95040	ye39a06.s1 Homo sapien	3.09e+00
c	348	15	11.7	360	14	AA865510	oh30b02.s1 NCI_CGAP_Ki	3.09e+00

c	349	15	11.7	360	21	W40158	zc82b11.r1 Pancreatic 3.09e+00
	350	15	11.7	361	21	W08336	mb41b05.r1 Soares mous 3.09e+00
c	351	15	11.7	361	26	AQ018190	CIT-HSP-2312K16.TF CIT 3.09e+00
	352	15	11.7	362	28	AQ098270	HS_3044_A1_A05_MF CIT 3.09e+00
c	353	15	11.7	363	16	R45423	yg42f04.s1 Homo sapien 3.09e+00
c	354	15	11.7	363	24	AA299133	EST111602 Uterus Homo s 3.09e+00
	355	15	11.7	364	12	AA788909	ae93g09.s1 Stratagene 3.09e+00
	356	15	11.7	364	23	AA140382	mq03h05.r1 Soares 2NbM 3.09e+00
	357	15	11.7	364	26	AQ012587	CIT-HSP-2298H11.TF CIT 3.09e+00
c	358	15	11.7	364	5	T59278	yc02b11.r1 Homo sapien 3.09e+00
c	359	15	11.7	364	25	AA341350	EST46708 Fetal kidney 3.09e+00
	360	15	11.7	366	24	AA268734	vb01h08.r1 Soares mous 3.09e+00
c	361	15	11.7	368	8	AA491218	aa47b07.s1 NCI_CGAP_GC 3.09e+00
c	362	15	11.7	368	13	AA167279	ep14d10.r1 Stratagene 3.09e+00
	363	15	11.7	369	14	AA807237	oc33a04.s1 NCI_CGAP_GC 3.09e+00
c	364	15	11.7	369	27	AQ084367	HS_2250_B1_C01_MF CIT 3.09e+00
c	365	15	11.7	369	20	W40111	zc81g08.r1 Pancreatic 3.09e+00
	366	15	11.7	370	28	AQ100453	HS_3055_A1_H07_MF CIT 3.09e+00
c	367	15	11.7	370	17	AI096409	qb91h10.x1 Soares_feta 3.09e+00
c	368	15	11.7	370	19	H94469	yv19b06.s1 Homo sapien 3.09e+00
c	369	15	11.7	370	26	AQ022612	HS_2183_A1_B09_MR CIT 3.09e+00
	370	15	11.7	372	27	AQ031378	HS_2222_A2_B09_MF CIT 3.09e+00
c	371	15	11.7	372	28	AQ128222	HS_3092_B1_H10_MR CIT 3.09e+00
	372	15	11.7	373	9	C65183	C.elegans cDNA clone y 3.09e+00
c	373	15	11.7	373	13	AA815017	oa88a06.s1 NCI_CGAP_GC 3.09e+00
c	374	15	11.7	374	29	AQ235104	HS_2052_B1_F07_MR CIT 3.09e+00
	375	15	11.7	375	11	AA683306	ae90b05.s1 Stratagene 3.09e+00
	376	15	11.7	375	9	C69114	C.elegans cDNA clone y 3.09e+00
c	377	15	11.7	375	27	AQ059327	CIT-HSP-2348F16.TF CIT 3.09e+00
	378	15	11.7	376	10	AA618346	nq18a05.s1 NCI_CGAP_Th 3.09e+00
c	379	15	11.7	377	17	AI082669	ox59b07.s1 Soares_NhHM 3.09e+00
	380	15	11.7	378	5	T98100	ye56g03.r1 Homo sapien 3.09e+00
c	381	15	11.7	378	27	B69095	CIT-HSP-2053B19.TR CIT 3.09e+00
	382	15	11.7	379	13	AA880210	vx39b06.r1 Stratagene 3.09e+00
c	383	15	11.7	380	20	W17201	zb13d08.r1 Soares feta 3.09e+00
c	384	15	11.7	380	6	AA400716	zv09c11.s1 Soares_NhHM 3.09e+00
c	385	15	11.7	382	8	AA122319	zn36d10.s1 Stratagene 3.09e+00
	386	15	11.7	383	23	T03370	IB1429 Infant brain, B 3.09e+00
c	387	15	11.7	383	28	AQ206149	HS_3240_A2_G12_MR CIT 3.09e+00
c	388	15	11.7	383	20	W01841	za72h07.r1 Soares feta 3.09e+00
c	389	15	11.7	383	28	AQ178132	HS_2217_B2_H07_MF CIT 3.09e+00
c	390	15	11.7	383	25	AA378213	EST90881 Synovial sarc 3.09e+00
c	391	15	11.7	383	28	AQ096779	HS_3039_A1_B05_MF CIT 3.09e+00
c	392	15	11.7	383	27	AQ008311	CIT-HSP-2280C18.TFB CI 3.09e+00
c	393	15	11.7	386	14	AA962357	oq79g06.s1 NCI_CGAP_Ki 3.09e+00
	394	15	11.7	387	20	ATTS1268	A. thaliana transcribe 3.09e+00
c	395	15	11.7	387	11	AA076260	zm90c11.r1 Stratagene 3.09e+00
c	396	15	11.7	388	20	W19460	zb31f09.r1 Soares para 3.09e+00
	397	15	11.7	389	28	AQ170558	HS_2165_A2_A02_MF CIT 3.09e+00
c	398	15	11.7	389	16	H52089	yq32a04.s1 Homo sapien 3.09e+00
c	399	15	11.7	390	26	B96605	T30L14TR TAMU Arabidop 3.09e+00

c	400	15	11.7	390	20	ATTS1274	A. thaliana transcribe 3.09e+00
	401	15	11.7	391	26	AQ022786	HS_2180_A2_D09_MR CIT 3.09e+00
c	402	15	11.7	391	8	AA454939	aa30c11.s1 NCI_CGAP_GC 3.09e+00
c	403	15	11.7	392	18	AI073663	ox98c11.s1 Soares_aene 3.09e+00
c	404	15	11.7	393	27	B82152	RPCI11-25O7.TV RPCI11 3.09e+00
	405	15	11.7	394	9	AA522852	ni72d11.s1 NCI_CGAP_Pr 3.09e+00
c	406	15	11.7	394	11	AA081726	sm62h03.r1 Stratagene 3.09e+00
	407	15	11.7	395	17	AI126680	qc52d09.x1 Soares_preg 3.09e+00
c	408	15	11.7	395	15	AA827877	od56g08.s1 NCI_CGAP_GC 3.09e+00
	409	15	11.7	396	28	AQ129791	HS_2253_A1_G10_MF CIT 3.09e+00
c	410	15	11.7	396	23	W86516	zh63c03.s1 Soares feta 3.09e+00
c	411	15	11.7	396	20	HHEA35%	H. sapiens partial cDN 3.09e+00
	412	15	11.7	398	5	R15052	yf86g02.r1 Homo sapien 3.09e+00
	413	15	11.7	398	5	R06501	yf09e07.r1 Homo sapien 3.09e+00
c	414	15	11.7	398	17	AI088582	qb12f04.x1 Soares_preg 3.09e+00
	415	15	11.7	399	27	AQ076038	CIT-HSP-2368G21.TF CIT 3.09e+00
	416	15	11.7	400	14	AA834887	oe02e04.s1 NCI_CGAP_Ov 3.09e+00
c	417	15	11.7	400	10	AA642620	nq71f04.s1 NCI_CGAP_Pr 3.09e+00
c	418	15	11.7	400	20	N88784	K5516F Fetal heart, La 3.09e+00
	419	15	11.7	401	7	HUM097C08B	Human fetal brain cDNA 3.09e+00
c	420	15	11.7	401	16	M85528	EST02044 Homo sapiens 3.09e+00
	421	15	11.7	403	29	AQ136362	HS_3063_A1_C05_MR CIT 3.09e+00
	422	15	11.7	403	28	AQ219829	HS_3238_B2_F12_MR CIT 3.09e+00
c	423	15	11.7	403	5	T52681	ya72e04.r1 Homo sapien 3.09e+00
c	424	15	11.7	403	21	W24418	zb60d12.r1 Soares feta 3.09e+00
	425	15	11.7	404	23	AA116856	mq25a04.r1 Barstead MP 3.09e+00
	426	15	11.7	404	16	H49323	yc21g09.s1 Homo sapien 3.09e+00
	427	15	11.7	404	13	AA760734	nz10a09.s1 NCI_CGAP_GC 3.09e+00
	428	15	11.7	404	23	AA065594	ml71f06.r1 Stratagene 3.09e+00
	429	15	11.7	404	19	N62531	yz74h02.s1 Homo sapien 3.09e+00
c	430	15	11.7	405	20	N79119	yz91h11.r1 Homo sapien 3.09e+00
	431	15	11.7	406	18	AI093684	qa31d08.s1 Soares_NhHM 3.09e+00
c	432	15	11.7	406	11	AA701279	zi65b01.s1 Soares feta 3.09e+00
c	433	15	11.7	406	22	AA181075	zp67e09.s1 Stratagene 3.09e+00
c	434	15	11.7	406	28	AQ092680	HS_3005_A2_B04_T7 CIT 3.09e+00
c	435	15	11.7	406	27	AQ091599	RPCI11-410I1.TF RPCI11 3.09e+00
c	436	15	11.7	407	11	AA076497	zm91e08.r1 Stratagene 3.09e+00
c	437	15	11.7	407	28	AQ117195	HS_2188_A1_G01_MF CIT 3.09e+00
	438	15	11.7	408	28	AQ220006	HS_3244_B2_E05_MR CIT 3.09e+00
	439	15	11.7	408	9	AA228300	nc38e03.r1 NCI_CGAP_Pr 3.09e+00
c	440	15	11.7	408	16	H45076	yo01e03.r1 Homo sapien 3.09e+00
c	441	15	11.7	408	21	W24898	zb65h05.r1 Soares feta 3.09e+00
c	442	15	11.7	408	8	AA099815	zl79a07.r1 Stratagene 3.09e+00
	443	15	11.7	410	6	AA389292	mp18c11.r1 Life Tech m 3.09e+00
	444	15	11.7	410	21	W60291	zd29c03.s1 Soares feta 3.09e+00
c	445	15	11.7	410	12	AA761100	ny11h11.s1 NCI_CGAP_GC 3.09e+00
	446	15	11.7	411	10	AA426461	zv61d02.s1 Soares test 3.09e+00
	447	15	11.7	411	16	H50552	yp83a10.s1 Homo sapien 3.09e+00
	448	15	11.7	412	27	AQ089798	HS_3008_B2_D12_T7 CIT 3.09e+00
c	449	15	11.7	412	5	T47851	yb17f01.r1 Homo sapien 3.09e+00
c	450	15	11.7	412	17	AI028217	ov96c09.x1 Soares_test 3.09e+00

c	451	15	11.7	412	27	AQ079831	CIT-HSP-2369G15.TF CIT 3.09e+00
	452	15	11.7	413	10	AA634400	zu79d01.s1 Soares test 3.09e+00
	453	15	11.7	413	16	R49644	yg61h07.s1 Homo sapien 3.09e+00
c	454	15	11.7	413	27	AQ051782	RPCI11-53J1.TJ RPCI11 3.09e+00
	455	15	11.7	414	8	AA565476	nk46a08.s1 NCI_CGAP_GC 3.09e+00
c	456	15	11.7	414	27	AQ024892	HS_2185_A2_C03_MF CIT 3.09e+00
c	457	15	11.7	414	29	AQ144089	HS_3074_A2_D06_MR CIT 3.09e+00
	458	15	11.7	415	22	W99324	ze03b07.s1 Soares feta 3.09e+00
c	459	15	11.7	415	15	AI061123	an35e09.x1 Gessler Wil 3.09e+00
c	460	15	11.7	415	12	AA725849	ai23e06.s1 Soares test 3.09e+00
	461	15	11.7	416	24	AA277219	va79d12.r1 Soares mous 3.09e+00
	462	15	11.7	417	23	AA174736	mt12d12.r1 Soares mous 3.09e+00
c	463	15	11.7	417	11	AA703951	ag79e05.r1 Stratagene 3.09e+00
	464	15	11.7	418	28	AQ200312	RPCI11-58L3.TJ RPCI11 3.09e+00
c	465	15	11.7	418	5	T36592	EST101556 Saccharomyce 3.09e+00
	466	15	11.7	419	23	AA152570	mr95d10.r1 Stratagene 3.09e+00
	467	15	11.7	419	20	W75696	me57a03.r1 Soares mous 3.09e+00
c	468	15	11.7	419	29	AQ240093	CIT-HSP-2384K24.TFB.1 3.09e+00
c	469	15	11.7	419	14	AA829151	od78e11.s1 NCI_CGAP_Ov 3.09e+00
	470	15	11.7	421	29	AQ145480	HS_3203_A2_C11_MR CIT 3.09e+00
	471	15	11.7	421	18	AI180495	uc66d12.r1 Soares mous 3.09e+00
	472	15	11.7	421	16	R82061	yj06f02.s1 Homo sapien 3.09e+00
c	473	15	11.7	421	23	AA121843	zn95g12.r1 Stratagene 3.09e+00
	474	15	11.7	422	5	T78176	yd79c08.r1 Homo sapien 3.09e+00
c	475	15	11.7	422	18	AI143921	qd99g02.x1 Soares_test 3.09e+00
	476	15	11.7	423	18	AI078455	oz05h07.x1 Soares_feta 3.09e+00
	477	15	11.7	423	5	R31370	yh75c06.r1 Homo sapien 3.09e+00
c	478	15	11.7	423	8	AA152480	zo07c03.r1 Stratagene 3.09e+00
c	479	15	11.7	424	28	AQ183345	HS_3138_B2_E10_MR CIT 3.09e+00
c	480	15	11.7	424	10	AA631864	np61d05.s1 NCI_CGAP_Br 3.09e+00
	481	15	11.7	425	12	AA753176	97AS2245 Rice Immature 3.09e+00
c	482	15	11.7	425	29	AQ246751	HS_2197_A2_A05_MR CIT 3.09e+00
c	483	15	11.7	425	15	AA974203	oq18f01.s1 NCI_CGAP_GC 3.09e+00
c	484	15	11.7	425	13	AA164784	zo99b10.s1 Stratagene 3.09e+00
c	485	15	11.7	426	16	H48631	yq80f05.r1 Homo sapien 3.09e+00
	486	15	11.7	427	27	AQ072324	HS_3011_A2_F01_T7 CIT 3.09e+00
	487	15	11.7	427	28	AQ189888	HS_3224_A2_A01_MR CIT 3.09e+00
c	488	15	11.7	427	12	AA777144	zf27e05.s1 Soares feta 3.09e+00
	489	15	11.7	428	27	GGA300120	Gallus gallus anonymou 3.09e+00
	490	15	11.7	428	19	H67779	yu54g09.s1 Homo sapien 3.09e+00
c	491	15	11.7	428	13	AA191168	zp86f12.r1 Stratagene 3.09e+00
c	492	15	11.7	428	11	AA082995	zn08c06.s1 Stratagene 3.09e+00
c	493	15	11.7	429	12	T76207	10985 Lambda-PRL2 Arab 3.09e+00
	494	15	11.7	430	14	AA900569	UI-R-E0-dn-a-05-0-UI.s 3.09e+00
c	495	15	11.7	430	14	AA964510	UI-R-C0-gv-f-03-0-UI.s 3.09e+00
c	496	15	11.7	431	6	W88760	zh74e05.s1 Soares feta 3.09e+00
c	497	15	11.7	431	17	AI070859	UI-R-C2-ms-a-05-0-UI.s 3.09e+00
c	498	15	11.7	431	29	AQ235251	HS_2053_B1_A10_MR CIT 3.09e+00
	499	15	11.7	432	5	T70413	yd13f03.r1 Homo sapien 3.09e+00
c	500	15	11.7	432	12	AA777178	zf29h08.s1 Soares feta 3.09e+00
c	501	15	11.7	432	26	AQ014493	CIT-HSP-2300D3.TR CIT- 3.09e+00

c	502	15	11.7	432	27	AQ031166	HS_2198_A2_B12_MR CIT 3.09e+00
	503	15	11.7	433	27	B83920	RPCI11-21C7.TV RPCI11 3.09e+00
c	504	15	11.7	433	6	AA437175	zv61d02.r1 Soares test 3.09e+00
c	505	15	11.7	433	12	AA769727	ob20b06.s1 NCI_CGAP_Ki 3.09e+00
c	506	15	11.7	433	6	AA043782	zk59d12.r1 Soares preg 3.09e+00
	507	15	11.7	434	14	AA910261	ok86d06.s1 NCI_CGAP_Ki 3.09e+00
c	508	15	11.7	434	20	W05145	za82f04.r1 Soares feta 3.09e+00
c	509	15	11.7	434	14	AA826044	od60a09.s1 NCI_CGAP_GC 3.09e+00
c	510	15	11.7	434	19	N26829	yw65d12.s1 Homo sapien 3.09e+00
	511	15	11.7	435	28	AQ113713	CIT-HSP-2371C10.TF CIT 3.09e+00
c	512	15	11.7	435	12	AA769750	ob20d06.s1 NCI_CGAP_Ki 3.09e+00
c	513	15	11.7	435	17	AU016786	Mus musculus 2-cell em 3.09e+00
	514	15	11.7	436	26	FR0009381	F.rubripes GSS sequenc 3.09e+00
	515	15	11.7	436	28	AQ167939	HS_2232_B1_D12_MR CIT 3.09e+00
c	516	15	11.7	436	21	W39666	zc19h03.r1 Soares para 3.09e+00
c	517	15	11.7	437	24	AA267404	ms96g11.r1 Soares mous 3.09e+00
c	518	15	11.7	437	29	AQ251687	HS_3002_A2_A03_MR CIT 3.09e+00
	519	15	11.7	438	14	AA982783	uh15e12.r1 Soares mous 3.09e+00
	520	15	11.7	438	23	AA177585	mt21f09.r1 Soares mous 3.09e+00
c	521	15	11.7	438	7	AA464853	zx44h02.s1 Soares tota 3.09e+00
c	522	15	11.7	439	21	W39174	zc16f02.r1 Soares para 3.09e+00
	523	15	11.7	440	13	AA866340	UI-R-A0-bm-g-09-0-UI.s 3.09e+00
	524	15	11.7	440	26	B67841	T24J21TR TAMU Arabidop 3.09e+00
	525	15	11.7	440	23	AA175441	ms87d04.r1 Soares mous 3.09e+00
c	526	15	11.7	441	21	W24976	zb66e08.r1 Soares feta 3.09e+00
	527	15	11.7	442	18	AI138632	qc31h07.x1 Soares_preg 3.09e+00
c	528	15	11.7	442	23	AA260565	va11a03.r1 Soares mous 3.09e+00
c	529	15	11.7	442	19	N44624	yy30h09.r1 Homo sapien 3.09e+00
	530	15	11.7	443	23	AA254916	ms80b10.r1 Soares mous 3.09e+00
	531	15	11.7	443	5	T71409	yd35d01.r1 Homo sapien 3.09e+00
c	532	15	11.7	444	17	AU016887	Mus musculus 2-cell em 3.09e+00
	533	15	11.7	445	9	AA553572	nk81h01.s1 NCI_CGAP_Sc 3.09e+00
	534	15	11.7	445	28	AQ111252	CIT-HSP-2371I22.TR CIT 3.09e+00
	535	15	11.7	446	17	AI126886	qb95d04.x1 Soares_feta 3.09e+00
c	536	15	11.7	446	26	B43544	HS-1057-B2-H05-MR.ab1 3.09e+00
c	537	15	11.7	446	21	W25193	zb67c09.r1 Soares feta 3.09e+00
	538	15	11.7	448	14	AA848772	EST191533 Normalized r 3.09e+00
	539	15	11.7	448	20	W59302	md41b11.r1 Soares mous 3.09e+00
	540	15	11.7	449	16	R91623	yq10f04.r1 Homo sapien 3.09e+00
	541	15	11.7	449	18	AI056616	oz10g11.x1 Soares_feta 3.09e+00
c	542	15	11.7	449	22	AA120852	zn83c09.r1 Stratagene 3.09e+00
c	543	15	11.7	449	10	AA417117	zu07f03.r1 Soares test 3.09e+00
c	544	15	11.7	449	27	B91845	CIT-HSP-2164I17.TR CIT 3.09e+00
c	545	15	11.7	449	18	AI081321	ox72g02.x1 Soares_NhHM 3.09e+00
c	546	15	11.7	449	10	AA155643	zo70e02.s1 Stratagene 3.09e+00
	547	15	11.7	450	16	R76192	yi71a11.s1 Homo sapien 3.09e+00
c	548	15	11.7	450	13	AA173815	zp01b06.r1 Stratagene 3.09e+00
c	549	15	11.7	450	18	AI141608	oy91e03.x1 Soares_feta 3.09e+00
	550	15	11.7	451	20	W50586	md04f08.r1 Soares mous 3.09e+00
c	551	15	11.7	451	20	W00703	za72f12.r1 Homo sapien 3.09e+00
	552	15	11.7	452	6	AA126906	zl87b03.s1 Stratagene 3.09e+00

	553	15	11.7	453	29	AQ147696	HS_3079_A1_C02_MR CIT 3.09e+00
	554	15	11.7	453	13	AA629762	ae63d04.s1 Stratagene 3.09e+00
	555	15	11.7	453	5	T98717	ye61f03.r1 Homo sapien 3.09e+00
	556	15	11.7	454	13	AA874857	UI-R-E0-cg-h-12-0-UI.s 3.09e+00
	557	15	11.7	455	11	AA734530	vt94h02.r1 Soares mous 3.09e+00
	558	15	11.7	455	23	AA254224	va09e01.r1 Soares mous 3.09e+00
c	559	15	11.7	455	22	AA194867	zp75h04.r1 Stratagene 3.09e+00
c	560	15	11.7	456	20	W01417	za73d03.r1 Soares feta 3.09e+00
c	561	15	11.7	457	18	AI185682	qe60f03.x1 Soares_feta 3.09e+00
	562	15	11.7	459	6	C19953	Rice cDNA, partial seq 3.09e+00
c	563	15	11.7	461	10	AA621225	zu81g04.s1 Soares test 3.09e+00
c	564	15	11.7	461	5	T91404	yd53g05.s1 Homo sapien 3.09e+00
	565	15	11.7	462	11	AA705102	zj94g10.s1 Soares feta 3.09e+00
	566	15	11.7	462	17	AI118621	uc12a01.r1 Soares mous 3.09e+00
	567	15	11.7	463	19	H98802	yx17e11.s1 Homo sapien 3.09e+00
	568	15	11.7	463	21	AA014442	mi64d06.r1 Soares mous 3.09e+00
c	569	15	11.7	463	13	AA810304	nw66g08.s1 NCI_CGAP_GC 3.09e+00
	570	15	11.7	464	29	AQ229812	HS_2027_A2_A09_MR CIT 3.09e+00
	571	15	11.7	465	28	AQ214405	HS_3012_B1_B02_T7 CIT 3.09e+00
	572	15	11.7	465	19	H61830	yu41g11.r1 Homo sapien 3.09e+00
c	573	15	11.7	465	9	AA523012	ni43g09.s1 NCI_CGAP_Lu 3.09e+00
c	574	15	11.7	465	6	AA026477	zj99b01.r1 Soares preg 3.09e+00
	575	15	11.7	467	17	AI091556	ow59b07.x1 Soares_NSF_ 3.09e+00
	576	15	11.7	468	6	AA040107	zk44b12.s1 Soares preg 3.09e+00
c	577	15	11.7	469	26	AQ021263	CIT-HSP-2308K21.TF CIT 3.09e+00
	578	15	11.7	471	28	AQ194944	RPCI11-59I17.TJ RPCI11 3.09e+00
	579	15	11.7	471	17	AI113560	GH09896.5prime GH Dros 3.09e+00
	580	15	11.7	471	11	AA677061	zj55a04.s1 Soares feta 3.09e+00
	581	15	11.7	471	15	AA937167	ok13a12.s1 Soares_NSF_ 3.09e+00
c	582	15	11.7	471	8	AA157683	zo46f06.r1 Stratagene 3.09e+00
c	583	15	11.7	471	28	AQ174597	HS_3208_A2_E05_MR CIT 3.09e+00
c	584	15	11.7	473	8	AA283596	zt15d08.s1 NCI_CGAP_GC 3.09e+00
c	585	15	11.7	473	22	AA195606	zq48g03.s1 Stratagene 3.09e+00
c	586	15	11.7	474	10	AA160288	zo87d11.r1 Stratagene 3.09e+00
	587	15	11.7	475	13	AA874807	UI-R-E0-bw-g-12-0-UI.s 3.09e+00
	588	15	11.7	475	11	T44617	7880 Lambda-PRL2 Arabi 3.09e+00
c	589	15	11.7	475	28	AQ217023	HS_3252_B2_D11_T7 CIT 3.09e+00
c	590	15	11.7	476	24	AA316162	EST187899 HCC cell lin 3.09e+00
c	591	15	11.7	476	20	W20417	zb22c03.r1 Soares feta 3.09e+00
c	592	15	11.7	478	17	AA966816	s9f03a1.f1 Aspergillus 3.09e+00
	593	15	11.7	479	18	AI167641	ok27c09.x1 Soares_NSF_ 3.09e+00
	594	15	11.7	479	28	AQ217365	HS_2190_A1_D06_MR CIT 3.09e+00
	595	15	11.7	480	16	H09648	yl99c07.s1 Homo sapien 3.09e+00
c	596	15	11.7	480	13	AA204765	zq83g03.r1 Stratagene 3.09e+00
c	597	15	11.7	480	8	AA225345	nc24b08.r1 NCI_CGAP_Pr 3.09e+00
	598	15	11.7	481	6	AA039317	zk39c07.s1 Soares preg 3.09e+00
	599	15	11.7	482	17	AI005254	ot98h12.x1 Soares_tota 3.09e+00
c	600	15	11.7	482	17	AI125395	qd88h08.x1 Soares_test 3.09e+00
c	601	15	11.7	482	13	AA610019	af18d12.s1 Soares test 3.09e+00
	602	15	11.7	483	20	W05278	za83d05.r1 Soares feta 3.09e+00
c	603	15	11.7	483	17	AA918307	ol42b05.s1 Soares_NFL_ 3.09e+00

c	604	15	11.7	483	21	W60382	zd29c03.r1 Soares feta 3.09e+00
c	605	15	11.7	484	11	AA701124	zg56b11.s1 Soares pine 3.09e+00
c	606	15	11.7	486	29	AQ141686	HS_3151_B1_E05_T7 CIT 3.09e+00
c	607	15	11.7	486	21	R84363	yq23h05.r1 Soares reti 3.09e+00
c	608	15	11.7	486	6	AA010624	zi09h09.s1 Soares feta 3.09e+00
	609	15	11.7	487	18	AI101895	EST211184 Normalized r 3.09e+00
	610	15	11.7	487	6	AA002220	zh85h02.r1 Soares feta 3.09e+00
	611	15	11.7	487	27	B74681	CIT-HSP-2034C13.TF CIT 3.09e+00
c	612	15	11.7	488	22	AA169736	zo84a07.r1 Stratagene 3.09e+00
c	613	15	11.7	488	13	AA165174	zo94b12.r1 Stratagene 3.09e+00
	614	15	11.7	489	27	B95201	CIT-HSP-2172H24.TF CIT 3.09e+00
	615	15	11.7	489	15	AI048279	ud68a05.y1 Sugano mous 3.09e+00
	616	15	11.7	489	19	N42851	yy13a05.r1 Homo sapien 3.09e+00
c	617	15	11.7	490	13	AA805687	ns42a01.s1 NCI_CGAP_GC 3.09e+00
	618	15	11.7	491	19	N51878	yz05b03.s1 Homo sapien 3.09e+00
	619	15	11.7	491	5	R48247	yj67b05.s1 Homo sapien 3.09e+00
	620	15	11.7	491	28	AQ121156	HS_3073_A1_A01_MF CIT 3.09e+00
c	621	15	11.7	491	10	AA169433	zo84g03.r1 Stratagene 3.09e+00
c	622	15	11.7	491	13	AA188140	zp71h08.r1 Stratagene 3.09e+00
	623	15	11.7	492	19	N57498	yy54d06.s1 Homo sapien 3.09e+00
c	624	15	11.7	492	13	AA630766	ac14b10.s1 Stratagene 3.09e+00
c	625	15	11.7	493	13	AA843188	aj15g06.s1 Soares para 3.09e+00
c	626	15	11.7	493	5	T67716	yc27d09.s1 Homo sapien 3.09e+00
c	627	15	11.7	493	14	C88719	Mus musculus early bla 3.09e+00
	628	15	11.7	494	9	AA042197	24518 CD4-13 Arabidops 3.09e+00
c	629	15	11.7	495	17	AU013778	Schizosaccharomyces po 3.09e+00
	630	15	11.7	497	10	AA616581	vn65h05.r1 Barstead mo 3.09e+00
c	631	15	11.7	497	19	W29110	yx47c01.r1 Homo sapien 3.09e+00
c	632	15	11.7	498	29	AQ235782	HS_2021_B2_B07_MR CIT 3.09e+00
	633	15	11.7	499	28	AQ177591	HS_2201_B1_H02_MF CIT 3.09e+00
c	634	15	11.7	500	11	AA718972	zh20h08.s1 Soares pine 3.09e+00
c	635	15	11.7	500	14	AA826045	od60a10.s1 NCI_CGAP_GC 3.09e+00
	636	15	11.7	501	26	AQ022182	CIT-HSP-2301H19.TF CIT 3.09e+00
c	637	15	11.7	501	8	AA283109	zt16b08.r1 NCI_CGAP_GC 3.09e+00
c	638	15	11.7	502	27	B94233	CIT-HSP-2173F20.TF CIT 3.09e+00
c	639	15	11.7	503	24	AA274207	vb92c07.r1 Soares mous 3.09e+00
	640	15	11.7	504	18	AI171789	EST217771 Normalized r 3.09e+00
c	641	15	11.7	504	17	AA912721	o141a11.s1 Soares_NFL_ 3.09e+00
	642	15	11.7	505	12	AA739190	vv50e04.r1 Soares 2NbM 3.09e+00
c	643	15	11.7	508	18	AI144218	qb93h06.x1 Soares_feta 3.09e+00
c	644	15	11.7	508	22	N74757	yv52d09.r1 Soares feta 3.09e+00
	645	15	11.7	510	8	AA280224	zt04c07.r1 NCI_CGAP_GC 3.09e+00
	646	15	11.7	510	27	AQ081996	RPCI11-54G1.TJ RPCI11 3.09e+00
	647	15	11.7	510	8	AA563298	v155c04.r1 Stratagene 3.09e+00
c	648	15	11.7	510	18	AI150550	qf41h02.x1 Soares_test 3.09e+00
c	649	15	11.7	510	15	AI026331	L0-550f3 Ice plant Lam 3.09e+00
c	650	15	11.7	510	20	W23475	zb33e01.r1 Soares para 3.09e+00
c	651	15	11.7	511	27	AQ029336	RPCI11-37L1.TV RPCI11 3.09e+00
c	652	15	11.7	512	21	N93004	zb59a02.s1 Soares feta 3.09e+00
	653	15	11.7	513	14	AA925495	UI-R-A1-ep-c-04-0-UI.s 3.09e+00
	654	15	11.7	514	28	AQ203724	HS_3103_A2_C05_MF CIT 3.09e+00

	655	15	11.7	516	15	AI060252	UI-R-C1-11-c-09-0-UI.s 3.09e+00
c	656	15	11.7	517	26	B43276	HS-1057-A2-H05-MR.abi 3.09e+00
c	657	15	11.7	518	27	AQ035078	CIT-HSP-2332M12.TF CIT 3.09e+00
	658	15	11.7	520	11	R80493	16848 Lambda-PRL2 Arab 3.09e+00
	659	15	11.7	523	10	AA426049	zw49g12.s1 Soares tota 3.09e+00
c	660	15	11.7	524	8	AA523163	ni66e02.s1 NCI_CGAP_Pr 3.09e+00
c	661	15	11.7	525	28	AQ187770	HS_3141_B1_B06_T7 CIT 3.09e+00
c	662	15	11.7	526	28	AQ193217	HS_3062_B1_G03_MF CIT 3.09e+00
c	663	15	11.7	529	19	N25018	yx17e11.r1 Homo sapien 3.09e+00
c	664	15	11.7	530	27	B69324	CIT-HSP-2053N12.TR CIT 3.09e+00
c	665	15	11.7	530	17	AI085959	oz86g07.x1 Soares_sene 3.09e+00
c	666	15	11.7	530	8	AA478784	zv20c04.s1 Soares NhHM 3.09e+00
c	667	15	11.7	530	28	AQ187538	HS_3132_A1_H01_T7 CIT 3.09e+00
c	668	15	11.7	530	28	AQ189592	HS_3219_B1_D03_T7 CIT 3.09e+00
	669	15	11.7	532	27	B80830	CIT-HSP-2022G6.TPB CIT 3.09e+00
c	670	15	11.7	533	8	AA547092	vk72g02.s1 Knowles Sol 3.09e+00
	671	15	11.7	536	11	H37618	15747 Lambda-PRL2 Arab 3.09e+00
	672	15	11.7	537	8	AA478903	zv20c04.r1 Soares NhHM 3.09e+00
	673	15	11.7	538	27	GGA160110	Gallus gallus anonymou 3.09e+00
	674	15	11.7	542	29	AQ233421	HS_2045_B2_A12_T7 CIT 3.09e+00
	675	15	11.7	543	11	AA075158	zm86d10.r1 Stratagene 3.09e+00
c	676	15	11.7	543	10	AA160485	zo73d07.s1 Stratagene 3.09e+00
c	677	15	11.7	545	8	AA214086	zn57h08.s1 Stratagene 3.09e+00
c	678	15	11.7	545	10	AA626713	zu89b12.s1 Soares test 3.09e+00
	679	15	11.7	547	28	AQ204106	HS_3107_A2_G04_T7 CIT 3.09e+00
	680	15	11.7	549	11	H37729	15858 Lambda-PRL2 Arab 3.09e+00
	681	15	11.7	550	28	AQ116266	RPCI11-42M2.TK.1 RPCI1 3.09e+00
c	682	15	11.7	552	19	N48370	yy43a05.r1 Homo sapien 3.09e+00
	683	15	11.7	553	14	AA896240	vy12g09.r1 Stratagene 3.09e+00
	684	15	11.7	554	5	T90031	ye13c01.s1 Homo sapien 3.09e+00
c	685	15	11.7	554	23	AA243749	zr67a10.s1 Soares NhHM 3.09e+00
	686	15	11.7	556	20	W65998	me09f04.r1 Soares mous 3.09e+00
c	687	15	11.7	557	11	N65187	20227 Lambda-PRL2 Arab 3.09e+00
	688	15	11.7	558	23	AA166345	ms71e02.r1 Soares mous 3.09e+00
	689	15	11.7	559	13	AA625532	af72g02.r1 Soares NhHM 3.09e+00
c	690	15	11.7	559	11	AA075159	zm86d10.s1 Stratagene 3.09e+00
	691	15	11.7	560	27	AQ057612	CIT-HSP-2345M10.TR CIT 3.09e+00
c	692	15	11.7	560	22	AA205666	zq68h02.s1 Stratagene 3.09e+00
	693	15	11.7	561	28	AQ194225	RPCI11-62J11.TK RPCI11 3.09e+00
c	694	15	11.7	561	29	AQ228546	HS_2023_B2_H10_MR CIT 3.09e+00
	695	15	11.7	562	13	AA819889	UI-R-A0-aq-c-02-0-UI.s 3.09e+00
c	696	15	11.7	563	12	AA777512	zj02g07.s1 Soares feta 3.09e+00
c	697	15	11.7	563	21	W52926	zc03g07.r1 Soares para 3.09e+00
	698	15	11.7	564	8	AA243670	zr67d12.r1 Soares NhHM 3.09e+00
c	699	15	11.7	566	26	FR0024970	F.rubripes GSS sequenc 3.09e+00
	700	15	11.7	568	15	AI064188	GH04471.5prime GH Dros 3.09e+00
c	701	15	11.7	568	20	W37666	zc12f06.r1 Soares para 3.09e+00
	702	15	11.7	571	28	AQ123690	HS_3104_B2_A02_T7 CIT 3.09e+00
	703	15	11.7	572	6	AA150269	zl07f01.s1 Soares preg 3.09e+00
	704	15	11.7	574	17	AU024535	Mus musculus unfertili 3.09e+00
	705	15	11.7	576	20	W03024	za02d08.r1 Soares mela 3.09e+00

c	706	15	11.7	580	26	B17112	344G14.TPB CIT978SKA1 3.09e+00
	707	15	11.7	581	23	AA243828	zr67a10.r1 Soares NhHM 3.09e+00
	708	15	11.7	582	27	AQ054289	CIT-HSP-2338I4.TR CIT- 3.09e+00
c	709	15	11.7	582	27	AQ004746	CIT-HSP-2157M8.TF CIT- 3.09e+00
c	710	15	11.7	584	15	C81211	Mus musculus 3.5-dpc b 3.09e+00
c	711	15	11.7	584	8	AA129358	zm84e01.r1 Stratagene 3.09e+00
c	712	15	11.7	586	27	B65270	CIT-HSP-2021D14.TR CIT 3.09e+00
c	713	15	11.7	588	10	AA624628	vn90b08.r1 Stratagene 3.09e+00
	714	15	11.7	589	11	AA075470	zm87c09.r1 Stratagene 3.09e+00
c	715	15	11.7	589	20	W07309	za97d09.r1 Soares feta 3.09e+00
	716	15	11.7	590	27	AQ074798	CIT-HSP-2351E8.TF CIT- 3.09e+00
c	717	15	11.7	590	23	AA053855	zf52f03.r1 Soares reti 3.09e+00
	718	15	11.7	594	27	B83084	RPCI11-20M21.TP RPCI11 3.09e+00
c	719	15	11.7	594	15	AI006551	ue14b11.y1 Sugano mous 3.09e+00
	720	15	11.7	597	26	B57478	CIT-HSP-2009M11.TF CIT 3.09e+00
c	721	15	11.7	600	20	W52964	zc94f12.r1 Pancreatic 3.09e+00
c	722	15	11.7	604	9	AA041159	24425 CD4-13 Arabidops 3.09e+00
c	723	15	11.7	605	27	B84356	RPCI11-24A5.TV RPCI11 3.09e+00
	724	15	11.7	606	22	AA166691	zo85d02.s1 Stratagene 3.09e+00
	725	15	11.7	608	19	H97933	yx04c04.s1 Homo sapien 3.09e+00
	726	15	11.7	608	27	B58686	CIT-HSP-2013L19.TF CIT 3.09e+00
c	727	15	11.7	608	28	AQ208310	HS_3219_A1_G09_T7 CIT 3.09e+00
c	728	15	11.7	608	27	FR0030302	Fugu rubripes GSS sequ 3.09e+00
	729	15	11.7	609	19	N45317	yz17d09.s1 Homo sapien 3.09e+00
	730	15	11.7	617	19	AA620012	vl76f06.r1 Knowles Sol 3.09e+00
c	731	15	11.7	617	22	AA205531	zq67a03.s1 Stratagene 3.09e+00
	732	15	11.7	621	28	AQ160628	mgxb0006018r CUGI Rice 3.09e+00
c	733	15	11.7	621	11	AA075556	zm88e07.r1 Stratagene 3.09e+00
c	734	15	11.7	622	10	AA632633	np87a08.s1 NCI_CGAP_Th 3.09e+00
c	735	15	11.7	623	22	AA166953	zo87h04.r1 Stratagene 3.09e+00
	736	15	11.7	627	29	AQ236897	RPCI11-71M10.TJ RPCI11 3.09e+00
	737	15	11.7	631	22	AA158264	zo76d10.s1 Stratagene 3.09e+00
	738	15	11.7	634	27	AQ025273	ep(3)3093 Drosophila m 3.09e+00
	739	15	11.7	635	27	AQ053094	RPCI11-53F15.TJ RPCI11 3.09e+00
	740	15	11.7	636	27	B83095	RPCI11-20G21.TP RPCI11 3.09e+00
c	741	15	11.7	638	20	W05775	za89c05.r1 Soares feta 3.09e+00
c	742	15	11.7	639	26	B20398	T22I5-T7 TAMU Arabidop 3.09e+00
c	743	15	11.7	639	26	AQ018479	CIT-HSP-2305B17.TR CIT 3.09e+00
	744	15	11.7	640	27	AQ025174	ep(3)1096 Drosophila m 3.09e+00
c	745	15	11.7	644	17	AA903468	ok58g12.s1 NCI_CGAP_Le 3.09e+00
c	746	15	11.7	644	27	AQ042884	CIT-HSP-2318O10.TF CIT 3.09e+00
c	747	15	11.7	645	15	C93921	Dictyostelium discoide 3.09e+00
c	748	15	11.7	652	11	AA075482	zm88g09.r1 Stratagene 3.09e+00
	749	15	11.7	654	28	AQ196497	CIT-HSP-2385C5.TR CIT- 3.09e+00
	750	15	11.7	661	27	AQ037909	CIT-HSP-2325G3.TV CIT- 3.09e+00
	751	15	11.7	670	10	AA166714	zo85f02.s1 Stratagene 3.09e+00
	752	15	11.7	673	23	AA108549	mp34e07.r1 Barstead MP 3.09e+00
c	753	15	11.7	674	27	AQ014402	RPCI11-24A5.TKBF RPCI1 3.09e+00
	754	15	11.7	676	27	AQ008149	RPCI11-20M21.TPD RPCI1 3.09e+00
	755	15	11.7	678	18	AI176546	EST220133 Normalized r 3.09e+00
	756	15	11.7	688	15	AI054608	coau0001105 Cotton Bol 3.09e+00

c	757	15	11.7	694	14	AA829764	od35g05.s1 NCI_CGAP_GC	3.09e+00
	758	15	11.7	701	9	HSZ99408	Homo sapiens mRNA; exp	3.09e+00
	759	15	11.7	718	27	AQ050737	nbxb0001cA02r CUGI Ric	3.09e+00
	760	15	11.7	718	26	B09259	T9i4-T7 TAMU Arabidops	3.09e+00
c	761	15	11.7	735	26	B01131	cSRL-126c8-u cSRL flow	3.09e+00
c	762	15	11.7	738	28	AQ157846	nbxb0010001r CUGI Rice	3.09e+00
c	763	15	11.7	739	28	AQ160201	mgxb0003B19r CUGI Rice	3.09e+00
	764	15	11.7	766	20	W25896	17c12 Human retina cDN	3.09e+00
c	765	15	11.7	779	28	AQ160143	mgxb0003C18r CUGI Rice	3.09e+00
c	766	15	11.7	782	26	B20572	T20J9-T7 TAMU Arabidop	3.09e+00
c	767	15	11.7	798	26	B20028	F6L3-T7 IGF Arabidopsi	3.09e+00
	768	15	11.7	812	26	B21120	F8H17-T7 IGF Arabidops	3.09e+00
	769	15	11.7	850	21	W29272	mc22f06.r1 Soares mous	3.09e+00
c	770	15	11.7	875	27	AQ026784	CIT-HSP-2322L5.TF CIT-	3.09e+00
	771	15	11.7	894	26	B09372	T23L24-Sp6 TAMU Arabid	3.09e+00
c	772	14	10.9	141	13	AA842139	MBAFCP3G01T3 Brugia ma	4.58e+01
c	773	14	10.9	149	7	AA512259	vj17d07.r1 Barstead mo	4.58e+01
c	774	14	10.9	156	23	AA247664	csh0268.seq.F Human fe	4.58e+01
c	775	14	10.9	201	20	ATTS0041	A. thaliana transcribe	4.58e+01
	776	14	10.9	202	12	AA763763	vp10d08.r1 Soares mous	4.58e+01
c	777	14	10.9	204	26	FR0015760	F.rubripes GSS sequenc	4.58e+01
c	778	14	10.9	219	17	AU012918	Schizosaccharomyces po	4.58e+01
c	779	14	10.9	226	25	AA376060	EST88429 HSC172 cells	4.58e+01
	780	14	10.9	232	17	AI112248	UI-R-Y0-mh-g-08-0-UI.s	4.58e+01
c	781	14	10.9	242	14	AA964046	UI-R-C0-gu-h-05-0-UI.s	4.58e+01
	782	14	10.9	247	23	AA254301	va09h06.r1 Soares mous	4.58e+01
	783	14	10.9	253	19	N74045	za58b07.s1 Homo sapien	4.58e+01
	784	14	10.9	259	20	N98743	yy65g08.r1 Homo sapien	4.58e+01
c	785	14	10.9	262	23	AA977886	7H08A03 Chromosome 7 H	4.58e+01
c	786	14	10.9	266	9	AA575487	vl94a11.r1 Knowles Sol	4.58e+01
c	787	14	10.9	269	25	AA376238	EST88653 HSC172 cells	4.58e+01
	788	14	10.9	275	8	AA483904	ne92a11.s1 NCI_CGAP_Ki	4.58e+01
c	789	14	10.9	276	20	W96522	zo43g01.r1 Soares reti	4.58e+01
	790	14	10.9	280	12	AA759800	vw57f01.r1 Soares mous	4.58e+01
	791	14	10.9	281	5	R22419	yh23c06.s1 Homo sapien	4.58e+01
c	792	14	10.9	290	14	AA887898	nq99f11.s1 NCI_CGAP_Co	4.58e+01
c	793	14	10.9	291	5	T39652	ya09f03.r3 Homo sapien	4.58e+01
	794	14	10.9	296	15	AA996655	UI-R-C0-hj-d-06-0-UI.s	4.58e+01
c	795	14	10.9	297	27	B93849	CIT-HSP-2163K5.TF CIT-	4.58e+01
c	796	14	10.9	300	7	SSC134	S.scrofa mRNA (clone C	4.58e+01
c	797	14	10.9	302	23	AA152578	mr95g11.r1 Stratagene	4.58e+01
	798	14	10.9	304	28	AQ198514	RPCI11-62C3.TK RPCI11	4.58e+01
	799	14	10.9	304	14	AA956002	UI-R-E1-fh-h-01-0-UI.s	4.58e+01
c	800	14	10.9	311	24	AA290451	vb18d03.r1 Soares mous	4.58e+01
c	801	14	10.9	311	23	AA117220	mo62f09.r1 Stratagene	4.58e+01
c	802	14	10.9	317	18	AI123863	qa74g08.x1 Soares_feta	4.58e+01
c	803	14	10.9	319	5	R38703	yc89h02.s1 Homo sapien	4.58e+01
	804	14	10.9	321	17	AA965420	i2c12a1.r1 Aspergillus	4.58e+01
	805	14	10.9	327	25	AA355214	EST63594 Jurkat T-cell	4.58e+01
	806	14	10.9	328	19	HSC30H051	H. sapiens partial cDN	4.58e+01
	807	14	10.9	329	5	R09608	yf23f03.r1 Homo sapien	4.58e+01

c	808	14	10.9	333	18	C99738	Arabidopsis thaliana Y	4.58e+01
	809	14	10.9	336	25	AA360385	EST69515 T-cell lympho	4.58e+01
	810	14	10.9	337	10	AA657694	nt82g02.s1 NCI_CGAP_Pr	4.58e+01
	811	14	10.9	338	18	AI157731	ud81g08.r2 Soares mous	4.58e+01
	812	14	10.9	338	17	AI093364	qb02g04.x1 Soares_preg	4.58e+01
	813	14	10.9	342	17	AA409595	EST01325 Mouse 7.5 dpc	4.58e+01
c	814	14	10.9	342	14	AA865321	og88c01.s1 NCI_CGAP_Ki	4.58e+01
	815	14	10.9	348	5	R39471	yh95a11.r1 Homo sapien	4.58e+01
c	816	14	10.9	348	28	AQ122315	HS_3085_A1_H03_MR CIT	4.58e+01
	817	14	10.9	349	23	W86837	zh59a04.s1 Soares feta	4.58e+01
	818	14	10.9	356	11	AA679665	ah15b12.s1 Gessler Wil	4.58e+01
	819	14	10.9	357	20	W97209	mf96b06.r1 Soares mous	4.58e+01
c	820	14	10.9	359	5	MUSP196A	Mus musculus expressed	4.58e+01
	821	14	10.9	360	5	R35209	yh87a10.s1 Homo sapien	4.58e+01
	822	14	10.9	360	9	C64239	C.elegans cDNA clone y	4.58e+01
	823	14	10.9	360	6	C13575	C.elegans cDNA clone y	4.58e+01
c	824	14	10.9	360	9	C42137	C.elegans cDNA clone y	4.58e+01
	825	14	10.9	361	18	AI151464	qc74f02.x1 Soares_plac	4.58e+01
	826	14	10.9	364	21	W72872	zd59f06.s1 Soares feta	4.58e+01
c	827	14	10.9	364	24	AA304781	EST17577 Aorta endothe	4.58e+01
c	828	14	10.9	365	15	AA879334	TENU0234 T.cruzi epima	4.58e+01
c	829	14	10.9	366	5	T38156	EST103536 Saccharomyce	4.58e+01
c	830	14	10.9	368	14	AA833998	qf24h06.s1 NCI_CGAP_Ki	4.58e+01
c	831	14	10.9	368	9	AA574822	vm31g06.r1 Knowles Sol	4.58e+01
c	832	14	10.9	371	18	AI039522	ox37b01.s1 Soares_tota	4.58e+01
	833	14	10.9	373	14	AA930882	vz71f09.s1 Soares mous	4.58e+01
c	834	14	10.9	374	16	H54354	yq93a09.r1 Homo sapien	4.58e+01
c	835	14	10.9	374	7	AA452230	zx30h06.r1 Soares tota	4.58e+01
	836	14	10.9	377	12	AA808277	oc41d03.s1 NCI_CGAP_GC	4.58e+01
c	837	14	10.9	378	28	AQ120788	HS_3076_A1_A01_MR CIT	4.58e+01
c	838	14	10.9	383	14	AA968028	uh07g09.r1 Soares mous	4.58e+01
c	839	14	10.9	386	24	HSPD04272	H.sapiens EST sequence	4.58e+01
c	840	14	10.9	386	22	N59697	yv55g07.r1 Soares feta	4.58e+01
	841	14	10.9	387	13	AA880960	vx34c12.r1 Stratagene	4.58e+01
c	842	14	10.9	389	7	SCMRAP0317	Schistosoma mansoni (c	4.58e+01
	843	14	10.9	392	28	AQ120877	HS_3073_B1_A07_MR CIT	4.58e+01
c	844	14	10.9	392	22	AA100549	zn72h11.s1 Stratagene	4.58e+01
c	845	14	10.9	394	28	AQ220786	HS_3247_B2_D11_MR CIT	4.58e+01
c	846	14	10.9	396	28	AQ169029	HS_3179_B2_H10_MR CIT	4.58e+01
c	847	14	10.9	396	13	AA812038	ob48a07.s1 NCI_CGAP_GC	4.58e+01
c	848	14	10.9	397	28	AQ221375	HS_2004_B2_A06_T7 CIT	4.58e+01
c	849	14	10.9	400	19	H68279	yr88a11.r1 Homo sapien	4.58e+01
c	850	14	10.9	401	10	AA626013	zu77a01.s1 Soares test	4.58e+01
	851	14	10.9	404	23	AA106962	mm30e08.r1 Stratagene	4.58e+01
	852	14	10.9	404	12	AA803623	GM13680.3prime GM Dros	4.58e+01
	853	14	10.9	405	17	AU020762	Mus musculus 8-cell em	4.58e+01
	854	14	10.9	406	28	AQ212052	HS_3241_A1_E04_T7 CIT	4.58e+01
c	855	14	10.9	407	16	H54253	yq89e12.s1 Homo sapien	4.58e+01
c	856	14	10.9	407	26	B75129	RPCI11-13J3.TP RPCI11	4.58e+01
c	857	14	10.9	407	9	AA595381	no36e08.s1 NCI_CGAP_Pr	4.58e+01
	858	14	10.9	409	17	AI112045	UI-R-Y0-mh-b-06-0-UI.s	4.58e+01

	859	14	10.9	409	27	AQ086768	HS_2175_B2_A03_MR CIT	4.58e+01
c	860	14	10.9	410	10	AA624484	vn82a04.r1 Knowles Sol	4.58e+01
	861	14	10.9	411	18	AI147501	qb21h05.x1 Soares_preg	4.58e+01
c	862	14	10.9	411	19	N57312	yv81c05.r1 Homo sapien	4.58e+01
c	863	14	10.9	411	16	H05751	yl73d06.s1 Homo sapien	4.58e+01
c	864	14	10.9	411	10	AA423205	ve37c09.r1 Soares mous	4.58e+01
	865	14	10.9	413	17	AA994075	ou37e05.s1 Soares_NFL_	4.58e+01
	866	14	10.9	415	6	AA193385	zr39h03.r1 Soares NhHM	4.58e+01
c	867	14	10.9	416	21	AA030341	mi28h05.r1 Soares mous	4.58e+01
c	868	14	10.9	416	20	W39378	zc77f07.r1 Pancreatic	4.58e+01
c	869	14	10.9	417	15	AI049820	an37f02.x1 Gessler Wil	4.58e+01
	870	14	10.9	418	27	AQ036384	CIT-HSP-2324P11.TF CIT	4.58e+01
	871	14	10.9	418	5	R54010	yj70g06.s1 Homo sapien	4.58e+01
	872	14	10.9	418	29	AQ139112	HS_3080_B2_A06_MR CIT	4.58e+01
	873	14	10.9	419	18	AI052381	os23c04.x1 Soares_tota	4.58e+01
	874	14	10.9	419	6	C17598	Human placenta cDNA 5'	4.58e+01
c	875	14	10.9	419	5	R49307	yg59f08.s1 Homo sapien	4.58e+01
c	876	14	10.9	419	20	W56878	zc01e04.r1 Soares para	4.58e+01
c	877	14	10.9	420	11	AA688093	nv14g06.s1 NCI_CGAP_Pr	4.58e+01
c	878	14	10.9	422	26	B46573	HS-1064-B1-E04-MR.abi	4.58e+01
c	879	14	10.9	423	13	AA806083	oc25a02.s1 NCI_CGAP_GC	4.58e+01
c	880	14	10.9	423	9	AA607207	vm84g01.r1 Knowles Sol	4.58e+01
	881	14	10.9	425	12	AA755275	vp92d06.r1 Stratagene	4.58e+01
	882	14	10.9	426	6	AA434484	zw13a03.r1 Soares ovar	4.58e+01
	883	14	10.9	426	14	AA912057	og22h11.s1 NCI_CGAP_PN	4.58e+01
	884	14	10.9	427	21	N95427	zb80c12.s1 Soares sene	4.58e+01
	885	14	10.9	427	19	N20637	yx40b04.s1 Homo sapien	4.58e+01
	886	14	10.9	430	9	AA535589	nf87a05.s1 NCI_CGAP_Co	4.58e+01
	887	14	10.9	431	13	AA859489	UI-R-E0-bv-g-01-0-UI.s	4.58e+01
	888	14	10.9	432	16	R63564	yi09d01.r1 Homo sapien	4.58e+01
	889	14	10.9	434	29	AQ139214	HS_3090_A1_D07_MR CIT	4.58e+01
c	890	14	10.9	434	15	AI049888	an30e11.x1 Gessler Wil	4.58e+01
	891	14	10.9	438	21	W18000	mb83a05.r1 Soares mous	4.58e+01
	892	14	10.9	441	10	AA593307	nn57c05.s1 NCI_CGAP_Ki	4.58e+01
	893	14	10.9	441	14	AA877742	nr02g03.s1 NCI_CGAP_Co	4.58e+01
c	894	14	10.9	446	8	AA215384	zr94f09.r1 NCI_CGAP_GC	4.58e+01
c	895	14	10.9	446	11	H36724	14392 Lambda-PRL2 Arab	4.58e+01
	896	14	10.9	449	15	C83183	Oryctolagus cuniculus	4.58e+01
	897	14	10.9	450	7	C24687	Dictyostelium discoide	4.58e+01
	898	14	10.9	450	20	W19409	zb26b02.r1 Soares feta	4.58e+01
	899	14	10.9	450	11	AA581490	nd17h08.s1 NCI_CGAP_Ov	4.58e+01
c	900	14	10.9	450	13	AA486716	ab18g07.s1 Stratagene	4.58e+01
	901	14	10.9	453	8	AA099099	zn45h02.r1 Stratagene	4.58e+01
	902	14	10.9	454	28	AQ119705	HS_2171_B2_H08_MR CIT	4.58e+01
	903	14	10.9	455	16	R62458	yg53d11.s1 Homo sapien	4.58e+01
	904	14	10.9	457	6	AA007319	zh98h08.r1 Soares feta	4.58e+01
c	905	14	10.9	458	29	AQ149560	HS_3165_B1_H02_MR CIT	4.58e+01
	906	14	10.9	461	22	H96662	yw10e10.s1 Soares mela	4.58e+01
	907	14	10.9	462	10	AA641547	nr79c01.s1 NCI_CGAP_Pr	4.58e+01
	908	14	10.9	463	19	H59119	yr03c08.s1 Homo sapien	4.58e+01
	909	14	10.9	465	19	N42905	yy08c09.r1 Homo sapien	4.58e+01

c	910	14	10.9	465	18	AI170274	EST216200 Normalized r	4.58e+01
	911	14	10.9	466	21	W47461	zc34g04.r1 Soares sene	4.58e+01
	912	14	10.9	469	19	N48811	yy76f04.s1 Homo sapien	4.58e+01
c	913	14	10.9	473	26	AQ012569	CIT-HSP-229BF13.TR CIT	4.58e+01
c	914	14	10.9	479	8	AA549877	0950m3 gmbPfHB3.1, G.	4.58e+01
	915	14	10.9	480	13	AA182668	zp47c01.r1 Stratagene	4.58e+01
	916	14	10.9	485	11	AA674313	vp96h12.r1 Stratagene	4.58e+01
	917	14	10.9	485	11	AA672209	vk90e03.s1 Knowles Sol	4.58e+01
c	918	14	10.9	485	29	AQ144226	HS_3085_A2_D03_MF CIT	4.58e+01
	919	14	10.9	490	22	AA024784	ze79a01.r1 Soares feta	4.58e+01
	920	14	10.9	492	19	N51241	yz13e07.s1 Homo sapien	4.58e+01
	921	14	10.9	492	17	AI118515	ue94a11.x1 Sugano mous	4.58e+01
	922	14	10.9	493	17	AI118113	ui45c11.x1 Sugano mous	4.58e+01
c	923	14	10.9	493	28	AQ209769	HS_3244_A1_A12_T7 CIT	4.58e+01
c	924	14	10.9	494	29	AQ141551	HS_3154_A1_E08_T7 CIT	4.58e+01
c	925	14	10.9	495	12	AA760345	vv75c12.r1 Stratagene	4.58e+01
	926	14	10.9	496	27	AQ059626	CIT-HSP-2347I16.TF CIT	4.58e+01
	927	14	10.9	498	6	AA127678	zk89f10.r1 Soares preg	4.58e+01
c	928	14	10.9	498	14	AA926013	UI-R-A1-es-c-12-0-UI.s	4.58e+01
	929	14	10.9	499	15	AA913349	ol28h10.s1 Soares_NFL_	4.58e+01
c	930	14	10.9	499	13	AA600304	ag04a09.s1 Gessler Wil	4.58e+01
c	931	14	10.9	499	13	AA714157	nv41f07.r1 NCI_CGAP_Ew	4.58e+01
	932	14	10.9	500	26	B62237	T20O18TR TAMU Arabidop	4.58e+01
c	933	14	10.9	500	11	AA727736	vp32b03.r1 Barstead mo	4.58e+01
c	934	14	10.9	506	7	AA445520	vf61b08.r1 Barstead MP	4.58e+01
c	935	14	10.9	509	11	AA710831	vt52b05.r1 Barstead mo	4.58e+01
c	936	14	10.9	510	20	W80067	me89c04.r1 Soares mous	4.58e+01
	937	14	10.9	512	8	AA243621	zs16d01.s1 NCI_CGAP_GC	4.58e+01
	938	14	10.9	512	15	AI011042	EST205493 Normalized r	4.58e+01
	939	14	10.9	512	12	AA803789	GM14279.5prime GM Dros	4.58e+01
c	940	14	10.9	512	11	AA689970	vt63h07.r1 Barstead mo	4.58e+01
c	941	14	10.9	517	21	W89377	mf73c03.r1 Soares mous	4.58e+01
c	942	14	10.9	518	21	AA037972	mi77e06.r1 Soares mous	4.58e+01
	943	14	10.9	528	17	AU023053	Mus musculus unfertili	4.58e+01
	944	14	10.9	531	7	AA453545	zx32b04.r1 Soares tota	4.58e+01
c	945	14	10.9	536	28	AQ172280	HS_3179_B2_F04_T7 CIT	4.58e+01
	946	14	10.9	542	18	AI154622	ud33e03.r1 Soares mous	4.58e+01
c	947	14	10.9	542	28	AQ118994	HS_3002_A2_C06_T7 CIT	4.58e+01
c	948	14	10.9	543	12	AA740880	nz03h05.s1 NCI_CGAP_GC	4.58e+01
	949	14	10.9	548	15	AI064547	GH05004.5prime GH Dros	4.58e+01
	950	14	10.9	550	23	AA217425	mu92e06.r1 Soares mous	4.58e+01
c	951	14	10.9	555	29	AQ142133	HS_3157_B1_D06_T7 CIT	4.58e+01
	952	14	10.9	557	14	AA898385	NCM6D4T7 Mycelial Neur	4.58e+01
	953	14	10.9	558	26	B01683	cSRL-136b3-u cSRL flow	4.58e+01
	954	14	10.9	563	22	AA165477	zo82d05.r1 Stratagene	4.58e+01
c	955	14	10.9	563	26	B78163	T31117TR TAMU Arabidop	4.58e+01
	956	14	10.9	566	15	C76026	Mus musculus 3.5-dpc b	4.58e+01
c	957	14	10.9	566	19	N63905	za27f09.s1 Homo sapien	4.58e+01
	958	14	10.9	572	17	AU020795	Mus musculus 8-cell em	4.58e+01
c	959	14	10.9	579	27	B74689	CIT-HSP-2034E11.TP CIT	4.58e+01
	960	14	10.9	580	15	C77885	Mus musculus 3.5-dpc b	4.58e+01

	961	14	10.9	581	9	AA522441	ng30g11.s1 NCI_CGAP_Co 4.58e+01
	962	14	10.9	581	15	C80026	Mus musculus 3.5-dpc b 4.58e+01
c	963	14	10.9	584	12	AA751864	96AS0717 Rice Immature 4.58e+01
	964	14	10.9	588	13	C88199	Mus musculus fertilize 4.58e+01
	965	14	10.9	591	27	FR0033301	Fugu rubripes GSS sequ 4.58e+01
	966	14	10.9	608	6	AA135510	sl09c01.s1 Soares preg 4.58e+01
	967	14	10.9	610	15	AI063822	GH03913.5prime GH Dros 4.58e+01
	968	14	10.9	610	28	AQ176899	HS_3207_A1_H03_T7 CIT 4.58e+01
	969	14	10.9	618	14	AA941983	LD27357.5prime LD Dros 4.58e+01
c	970	14	10.9	618	22	AA148053	zo51g11.r1 Stratagene 4.58e+01
	971	14	10.9	620	26	AG004960	Homo sapiens genomic D 4.58e+01
	972	14	10.9	620	27	AG011640	Homo sapiens genomic D 4.58e+01
	973	14	10.9	625	12	AA820886	LD24814.5prime LD Dros 4.58e+01
	974	14	10.9	629	15	C77452	Mus musculus 3.5-dpc b 4.58e+01
	975	14	10.9	629	28	AQ182845	HS_3110_B2_F02_MR CIT 4.58e+01
c	976	14	10.9	638	14	C92507	Dictyostelium discoide 4.58e+01
	977	14	10.9	645	11	AA694760	GM01488.5prime GM Dros 4.58e+01
c	978	14	10.9	648	14	C91107	Dictyostelium discoide 4.58e+01
	979	14	10.9	651	12	AA202616	LD02753.5prime LD Dros 4.58e+01
c	980	14	10.9	655	20	W26242	26g4 Human retina cDNA 4.58e+01
c	981	14	10.9	662	9	AA596275	vo38h06.r1 Barstead mo 4.58e+01
	982	14	10.9	664	12	AA735510	LD21433.5prime LD Dros 4.58e+01
c	983	14	10.9	665	26	B15703	34502.TP CIT978SKA1 Ho 4.58e+01
c	984	14	10.9	673	28	AQ156899	nbxb0008J13r CUGI Rice 4.58e+01
c	985	14	10.9	697	22	AA193941	rs07g09.r1 Sommer Pris 4.58e+01
	986	14	10.9	699	27	AQ036745	CIT-HSP-2334H13.TR CIT 4.58e+01
c	987	14	10.9	710	27	AQ008499	CIT-HSP-2282L24.TFB CI 4.58e+01
	988	14	10.9	713	14	AA942208	LD26289.5prime LD Dros 4.58e+01
	989	14	10.9	731	11	AA673192	vp46d09.r1 Knowles Sol 4.58e+01
c	990	14	10.9	732	28	AF010781	Homo sapiens chromosom 4.58e+01
c	991	14	10.9	740	26	B11649	F14J16-Sp6.3 IGF Arabi 4.58e+01
c	992	14	10.9	743	26	AG007718	Homo sapiens genomic D 4.58e+01
	993	14	10.9	745	12	AA817184	LD22357.5prime LD Dros 4.58e+01
	994	14	10.9	748	28	AQ158482	nbxb0011006f CUGI Rice 4.58e+01
	995	14	10.9	750	28	AQ201170	RPC111-46K24.TJ RPC111 4.58e+01
	996	14	10.9	773	28	AQ163066	mgxb0021P23r CUGI Rice 4.58e+01
	997	14	10.9	774	28	AQ163059	mgxb0021P07r CUGI Rice 4.58e+01
c	998	14	10.9	820	26	B05515	cSRL-64h12-u cSRL flow 4.58e+01
c	999	14	10.9	946	26	B11509	T12B13-T7 TAMU Arabido 4.58e+01
c	1000	14	10.9	1260	26	B12114	T1L15-Sp6.1 TAMU Arabi 4.58e+01

ALIGNMENTS

```
RESULT    1
LOCUS       HUMGS01470     128 bp    mRNA            EST       17-JUN-1996
DEFINITION  Human HL60 3'directed MboI cDNA, HUMGS01470, clone pm1222, mRNA
            sequence.
ACCESSION   D20496
NID         g501592
KEYWORDS    EST; EST(expressed sequence tag); HL60; gene signature(GS);
```



```
            granulocyte; macrophage; monocyte; promyelocyte.
SOURCE      Homo sapiens adult female promyelocyte cell_line:HL60 cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo
REFERENCE   1  (sites)
  AUTHORS   Okubo,K., Itoh,K., Fukushima,A., Yoshii,J. and Matsubara,K.
  TITLE     Monitoring cell physiology by expression profiles and discovering
            cell type-specific genes by compiled expression profiles
  JOURNAL   Genomics 30 (2), 178-186 (1995)
  MEDLINE   96163870
REFERENCE   2  (bases 1 to 128)
  AUTHORS   Okubo,K., Fukushima,A., Yoshii,J., Niiyama,T., Kojima,Y.,
            Yoshinari,H., Arimoto,J. and Matsubara,K.
  TITLE     Gene expression of human promyelocytic cell line HL60before and
            after induction of differentiation. A new application of
            3'directedcDNA sequencing
  JOURNAL   Unpublished (1993)
REFERENCE   3  (bases 1 to 128)
  AUTHORS   Okubo,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-SEP-1993) to the DDBJ/EMBL/GenBank databases. Kousaku
            Okubo, Osaka University, Institute for Molecular and Cellular Bio;
            1-3,Yamada-oka, Suita, Osaka Pref. 565, Japan
            (E-mail:kousaku@imcb.osaka-u.ac.jp, Tel:06-877-5111(ex.3315),
            Fax:06-877-1922)
COMMENT     Submitted (10-Sep-1993) to DDBJ by:
            Kousaku Okubo
            Institute for Molecular and Cellular Biology
            Osaka University
            3-1, Yamadaoka
            Suita, Osaka, 565
            Japan
            Phone: 06-877-5111
            Fax  : 06-875-1922.
FEATURES                Location/Qualifiers
     source             1. .128
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /cell_line="HL60"
                        /cell_type="promyelocyte"
                        /dev_stage="adult"
                        /sex="female"
BASE COUNT       41 a      22 c      18 g      47 t
ORIGIN

  Query Match          100.0%;  Score 128;  DB 7;  Length 128;
  Best Local Similarity 100.0%;  Pred. No. 2.46e-235;
  Matches  128;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```



```
Db      1 GATCAGAATTTTTAAAGCAGAATTTTGTCAAAAGGGTCATTTTTTTTGTCTACCCCTTTTA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      1 GATCAGAATTTTTAAAGCAGAATTTTGTCAAAAGGGTCATTTTTTTTGTCTACCCCTTTTA 60

Db     61 CACTTTTCAGATTCTCAAAGTGTCTCATCTCAACTTTTAAAAGAATAAAGAATATCTTGC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     61 CACTTTTCAGATTCTCAAAGTGTCTCATCTCAACTTTTAAAAGAATAAAGAATATCTTGC 120

Db    121 TGGGCAAA 128
          ||||||||
Qy    121 TGGGCAAA 128


RESULT    2
LOCUS       H78523         401 bp    mRNA            EST       09-NOV-1995
DEFINITION  yu13c02.r1 Homo sapiens cDNA clone 233666 5'.
ACCESSION   H78523
NID         g1056612
KEYWORDS    EST.
SOURCE      human clone=233666 primer=M13RP1 library=Soares fetal liver spleen
            1NFLS vector=pT7T3D (Pharmacia) with a modified polylinker
            host=DH10B (ampicillin resistant) Rsite1=Pac I Rsite2=Eco RI Liver
            and spleen from a 20 week-post conception male fetus. 1st strand
            cDNA was primed with a Pac I - oligo(dT) primer [5'
            AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTT 3'], double-stranded
            cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac
            I and cloned into the Pac I and Eco RI sites of the modified pT7T3
            vector.  Library went through one round of normalization. Library
            constructed by Bento Soares and M.Fatima Bonaldo.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 401)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 340
```



```
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES                Location/Qualifiers
     source             1. .401
                        /organism="Homo sapiens"
                        /clone="233666"
     mRNA               <1. .>401
BASE COUNT      104 a      68 c      60 g     167 t       2 others
ORIGIN

  Query Match          64.8%;  Score 83;  DB 19;  Length 401;
  Best Local Similarity 100.0%;  Pred. No. 6.54e-134;
  Matches   83;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    306 TTTGTCAAAAGGGTCATTTTTTTTGTCTACCCCTTTTACACTTTTCAGATTCTCAAAGTGT 365
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     24 TTTGTCAAAAGGGTCATTTTTTTTGTCTACCCCTTTTACACTTTTCAGATTCTCAAAGTGT 83

Db    366 CTCATCTCAACTTTTAAAAGAAT 388
          |||||||||||||||||||||||
Qy     84 CTCATCTCAACTTTTAAAAGAAT 106


RESULT    3
LOCUS       AA053439       380 bp    mRNA            EST       13-SEP-1996
DEFINITION  zl70b05.r1 Stratagene colon (#937204) Homo sapiens cDNA clone
            509937 5' similar to gb:X02492 INTERFERON-INDUCED PROTEIN 6-16
            PRECURSOR (HUMAN);, mRNA sequence.
ACCESSION   AA053439
NID         g1544094
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 380)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
            Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
```

```
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 342.
FEATURES              Location/Qualifiers
     source           1. .380
                      /organism="Homo sapiens"
                      /note="Organ: colon; Vector: pBluescript SK-; Site_1:
                      EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                      Oligo dT. T-84 colonic epithelial cell line.  Average
                      insert size: 1.0 kb; Uni-ZAP XR Vector; ~5' adaptor
                      sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor sequence: 5'
                      CTCGAGTTTTTTTTTTTTTTTT 3'"
                      /db_xref="taxon:9606"
                      /clone="509937"
                      /clone_lib="Stratagene colon (#937204)"
                      /lab_host="SOLR cells (kanamycin resistant)"
     mRNA             <1. .>380
BASE COUNT      114 a     79 c     80 g    104 t      3 others
ORIGIN

  Query Match           61.7%;  Score 79;  DB 21;  Length 380;
  Best Local Similarity 100.0%;  Pred. No. 4.24e-125;
  Matches    79;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db       1 TCTACCCCTTTTACACTTTTCAGATTCTCAAAGTGTCTCATCTCAACTTTTAAAAGAATA 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      48 TCTACCCCTTTTACACTTTTCAGATTCTCAAAGTGTCTCATCTCAACTTTTAAAAGAATA 107

Db      61 AAGAATATCTTGCTGGGCA 79
           ||||||||||||||||||
Qy     108 AAGAATATCTTGCTGGGCA 126


RESULT    4
LOCUS       AA461465      356 bp    mRNA            EST       09-JUN-1997
DEFINITION  zx68e06.s1 Soares total fetus Nb2HF8 9w Homo sapiens cDNA clone
            796642 3', mRNA sequence.
ACCESSION   AA461465
NID         g2185329
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo
REFERENCE   1  (bases 1 to 356)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J.,
            Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
```



```
  TITLE     WashU-Merck EST Project 1997
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -41m13 fwd. ET from Amersham
            High quality sequence stop: 192.
FEATURES              Location/Qualifiers
     source           1. .356
                      /organism="Homo sapiens"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                      was prepared from mRNA obtained from pooled 8-9 week
                      (total) fetus material with a Not I - oligo(dT) primer (5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCTTAATTTTTTTTTTTTTTTT 3']).
                      Double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Not I and cloned into the Not I
                      and Eco RI sites of the modified pT7T3 vector. Library
                      went through one round of normalization, and was
                      constructed  by Bento Soares and M. Fatima Bonaldo."
                      /db_xref="taxon:9606"
                      /clone="796642"
                      /clone_lib="Soares total fetus Nb2HF8 9w"
                      /dev_stage="8-9 weeks"
                      /lab_host="DH10B"
     mRNA             complement(<1. .>356)
                      /db_xref="GDB:6041025"
BASE COUNT      105 a     76 c     56 g    119 t
ORIGIN

  Query Match           15.6%;  Score 20;  DB 7;  Length 356;
  Best Local Similarity 100.0%;  Pred. No. 4.21e-07;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      30 AGATATTCTTTATTCTTTTA 49
           ||||||||||||||||||||
Cp     117 AGATATTCTTTATTCTTTTA 98


RESULT    5
LOCUS       AQ017960      363 bp    DNA             GSS       08-JUN-1998
DEFINITION  CIT-HSP-2314N16.TF CIT-HSP Homo sapiens genomic clone 2314N16,
            genomic survey sequence.
ACCESSION   AQ017960
```



```
NID         g3196696
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 363)
  AUTHORS   Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K.,
            Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
            Simon,M. and Venter,J.C.
  TITLE     Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13-21
            Class: BAC ends.
FEATURES              Location/Qualifiers
     source           1. .363
                      /organism="Homo sapiens"
                      /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                      HindIII"
                      /db_xref="taxon:9606"
                      /clone="2314N16"
                      /clone_lib="CIT-HSP"
                      /sex="Male"
                      /cell_type="Sperm"
BASE COUNT      126 a     70 c     51 g    116 t
ORIGIN

  Query Match           14.8%;  Score 19;  DB 26;  Length 363;
  Best Local Similarity 100.0%;  Pred. No. 1.29e-05;
  Matches    19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     171 CTTTATTCTTTTAAAAGTT 189
           |||||||||||||||||||
Cp     110 CTTTATTCTTTTAAAAGTT 92


RESULT    6
LOCUS       W68051        528 bp    mRNA            EST       16-OCT-1996
DEFINITION  zd42c09.s1 Soares fetal heart NbHH19W Homo sapiens cDNA clone
```



```
            343312 3' similar to contains Alu repetitive element;contains
            element PTR5 repetitive element ;, mRNA sequence.
ACCESSION   W68051
NID         g1376940
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 528)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 605   Std Error: 0.00
            Seq primer: ETPrimer
            High quality sequence stop: 377.
FEATURES              Location/Qualifiers
     source           1. .528
                      /organism="Homo sapiens"
                      /note="Organ: heart; Vector: pT7T3D (Pharmacia) with a
                      modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                      strand cDNA was primed with a Not I - oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCATCTTTTTTTTTTTTTTTT 3'],
                      double-stranded cDNA was size selected, ligated to Eco RI
                      adapters (Pharmacia), digested with Not I and cloned into
                      the Not I and Eco RI sites of a modified pT7T3 vector
                      (Pharmacia). Library went through one round of
                      normalization to a Cot = 5. Library constructed by
                      M.Fatima Bonaldo. This library was constructed from the
                      same fetus as the fetal lung library, Soares fetal lung
                      NbHL19W."
                      /db_xref="taxon:9606"
                      /clone="343312"
                      /clone_lib="Soares fetal heart NbHH19W"
                      /sex="unknown"
                      /dev_stage="19 weeks"
                      /lab_host="DH10B (ampicillin resistant)"
```

```
     mRNA            complement(<1. .>528)
BASE COUNT      116 a    142 c    127 g    141 t      2 others
ORIGIN

  Query Match          14.8%;  Score 19;  DB 21;  Length 528;
  Best Local Similarity 100.0%;  Pred. No. 1.29e-05;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    306 AGATATTCTTTATTCTTTT 324
          ||||||||||||||||||||
Cp    117 AGATATTCTTTATTCTTTT 99


RESULT    7
LOCUS       B76117        358 bp    DNA             GSS       06-JUN-1998
DEFINITION  RPCI11-13K22.TP RPCI11 Homo sapiens genomic clone R-13K22, genomic
            survey sequence.
ACCESSION   B76117
NID         g2771804
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 358)
  AUTHORS   Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K.,
            Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences for Sequence-Ready Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: RPCI11-13K22.TV
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            For clone availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: SP6
            Class: BAC ends.
FEATURES              Location/Qualifiers
     source           1. .358
                      /organism="Homo sapiens"
                      /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                      RPCI11 Human Male BAC Library"
                      /db_xref="taxon:9606"
                      /clone="R-13K22"
```



```
                      /clone_lib="RPCI11"
                      /sex="Male"
                      /cell_type="Lymphocytes"
BASE COUNT      131 a     41 c     49 g    137 t
ORIGIN

  Query Match          14.1%;  Score 18;  DB 26;  Length 358;
  Best Local Similarity 100.0%;  Pred. No. 3.50e-04;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    166 TCATCTCAACTTTTAAAA 183
          ||||||||||||||||||
Qy     85 TCATCTCAACTTTTAAAA 102


RESULT    8
LOCUS       AI154735      374 bp    mRNA            EST       30-SEP-1998
DEFINITION  ud36h11.r1 Soares mouse mammary gland NbMMG Mus musculus cDNA clone
            1448037 5', mRNA sequence.
ACCESSION   AI154735
NID         g3683204
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 374)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 266 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:921353
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 359.
FEATURES              Location/Qualifiers
     source           1. .374
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
```



```
                      /note="Organ: mammary gland; Vector: pT7T3D-Pac
                      (Pharmacia) with a modified polylinker; Site_1: Not I;
                      Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                      oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATGGTTTTTTTTTTTTTTTTTTTTTT
                      T 3']; double-stranded cDNA was ligated to Eco RI
                      adaptors (Pharmacia), digested with Not I and cloned into
                      the Not I and Eco RI sites of the modified pT7T3 vector.
                      RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
                      constructed and normalized by Bento Soares and M.Fatima
                      Bonaldo."
                      /db_xref="taxon:10090"
                      /clone="1448037"
                      /clone_lib="Soares mouse mammary gland NbMMG"
                      /sex="male"
                      /tissue_type="mammary gland"
                      /dev_stage="4 weeks"
                      /lab_host="DH10B"
BASE COUNT      122 a     69 c     71 g    112 t
ORIGIN

  Query Match          14.1%;  Score 18;  DB 18;  Length 374;
  Best Local Similarity 100.0%;  Pred. No. 3.50e-04;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     58 ACTTTTCAGATTCTCAAA 75
          ||||||||||||||||||
Qy     62 ACTTTTCAGATTCTCAAA 79


RESULT    9
LOCUS       R92292        377 bp    mRNA            EST       25-AUG-1995
DEFINITION  yq03b03.r1 Homo sapiens cDNA clone 195821 5'.
ACCESSION   R92292
NID         g959832
KEYWORDS    EST.
SOURCE      human clone=195821 library=Soares fetal liver spleen 1NFLS
            vector=pT7T3D (Pharmacia) with a modified polylinker host=DH10B
            (ampicillin resistant) primer=M13RP1 Rsite1=Pac I Rsite2=Eco RI
            Liver and spleen from a 20 week-post conception male fetus. 1st
            strand cDNA was primed with a Pac I - oligo(dT) primer [5'
            AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTT 3'], double-stranded
            cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac
            I and cloned into the Pac I and Eco RI sites of the modified pT7T3
            vector. Library went through one round of normalization. Library
            constructed by Bento Soares and M.Fatima Bonaldo.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
```



```
REFERENCE   1  (bases 1 to 377)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 346
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES              Location/Qualifiers
     source           1. .377
                      /organism="Homo sapiens"
                      /clone="195821"
     mRNA             <1. .>377
BASE COUNT      125 a     67 c     63 g    120 t      2 others
ORIGIN

  Query Match          14.1%;  Score 18;  DB 16;  Length 377;
  Best Local Similarity 100.0%;  Pred. No. 3.50e-04;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     23 TCTCAACTTTTAAAAGAA 40
          ||||||||||||||||||
Qy     88 TCTCAACTTTTAAAAGAA 105


RESULT   10
LOCUS       AQ181367      408 bp    DNA             GSS       09-SEP-1998
DEFINITION  HS_3216_A2_E03_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3216 Col=6 Row=I, genomic survey
            sequence.
ACCESSION   AQ181367
NID         g3578734
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 408)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Furlong,J. and
```

```
            Shaker,R.
  TITLE     Construction of a Characterized Clone Resource for Genomic
            Sequencing
  JOURNAL   Unpublished (1998)
COMMENT
            Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3216  row: I  column: 6
            Class: BAC ends
            High quality sequence stop: 408.
FEATURES              Location/Qualifiers
     source           1. .408
                      /organism="Homo sapiens"
                      /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                      E-Coli DH10B"
                      /db_xref="taxon:9606"
                      /clone="Plate=3216 Col=6 Row=I"
                      /clone_lib="CIT Approved Human Genomic Sperm Library D"
                      /sex="male"
BASE COUNT      141 a     47 c     58 g    160 t      2 others
ORIGIN

  Query Match           14.1%;  Score 18;  DB 28;  Length 408;
  Best Local Similarity 100.0%;  Pred. No. 3.50e-04;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    189 TCATCTCAACTTTTAAAA 206
          ||||||||||||||||||
Qy     85 TCATCTCAACTTTTAAAA 102


RESULT   11
LOCUS       R37679        411 bp    mRNA            EST       04-MAY-1995
DEFINITION  yf61g04.s1 Homo sapiens cDNA clone 26789 3'.
ACCESSION   R37679
NID         g795135
KEYWORDS    EST.
SOURCE      human clone=26789 library=Soares infant brain 1NIB vector=Lafmid BA
            host=DH10B (ampicillin resistant) primer=-21m13 Rsite1=Not I
            Rsite2=Hind III Whole brain from a 73 days post natal female. 1st
            strand cDNA was primed with a Not I - oligo(dT) primer [5'
            AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTTT 3']; double-stranded
            cDNA was ligated to Hind III adaptors (Pharmacia), digested with
            Not I and directionally cloned into the Not I and Hind III sites of
            the Lafmid BA vector. Library went through one round of
```



```
            normalization. Library constructed by Bento Soares and M.Fatima
            Bonaldo.
  ORGANISM  Homo sapiens
            Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 411)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            GDB: G00-399-136
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 368
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES              Location/Qualifiers
     source           1. .411
                      /organism="Homo sapiens"
                      /clone="26789"
BASE COUNT      150 a     62 c     74 g    121 t      4 others
ORIGIN

  Query Match           14.1%;  Score 18;  DB 5;  Length 411;
  Best Local Similarity 100.0%;  Pred. No. 3.50e-04;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    371 AAAATTCTGCTTTAAAAA 388
          ||||||||||||||||||
Cp     26 AAAATTCTGCTTTAAAAA 9


RESULT   12
LOCUS       B83406        439 bp    DNA             GSS       20-AUG-1998
DEFINITION  RPCI11-13K22.TPB RPCI11 Homo sapiens genomic clone R-13K22, genomic
            survey sequence.
ACCESSION   B83406
NID         g2870429
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
```



```
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 439)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
            Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences for Sequence-Ready Map Building (1998)
  JOURNAL   Unpublished (1998)
COMMENT     Other_GSSs: RPCI11-13K22.TVB RPCI11-13K22.TP RPCI11-13K22.TV
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            For clone availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: SP6
            Class: BAC ends.
FEATURES              Location/Qualifiers
     source           1. .439
                      /organism="Homo sapiens"
                      /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                      RPCI11 Human Male BAC Library"
                      /db_xref="taxon:9606"
                      /clone="R-13K22"
                      /clone_lib="RPCI11"
                      /sex="Male"
                      /cell_type="Lymphocytes"
BASE COUNT      151 a     57 c     64 g    167 t
ORIGIN

  Query Match           14.1%;  Score 18;  DB 27;  Length 439;
  Best Local Similarity 100.0%;  Pred. No. 3.50e-04;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    178 TCATCTCAACTTTTAAAA 195
          ||||||||||||||||||
Qy     85 TCATCTCAACTTTTAAAA 102


RESULT   13
LOCUS       R93755        510 bp    mRNA            EST       29-AUG-1995
DEFINITION  yg35b06.r1 Homo sapiens cDNA clone 197747 5'.
ACCESSION   R93755
NID         g967921
KEYWORDS    EST.
SOURCE      human clone=197747 library=Soares fetal liver spleen 1NFLS
```



```
            vector=pT7T3D (Pharmacia) with a modified polylinker host=DH10B
            (ampicillin resistant) primer=M13RP1 Rsite1=Pac I Rsite2=Eco RI
            Liver and spleen from a 20 week-post conception male fetus. 1st
            strand cDNA was primed with a Pac I - oligo(dT) primer [5'
            AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTT 3'], double-stranded
            cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac
            I and cloned into the Pac I and Eco RI sites of the modified pT7T3
            vector.  Library went through one round of normalization. Library
            constructed by Bento Soares and M.Fatima Bonaldo.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 510)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 326
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES              Location/Qualifiers
     source           1. .510
                      /organism="Homo sapiens"
                      /clone="197747"
     mRNA             <1. .>510
BASE COUNT      177 a     76 c     71 g    180 t      6 others
ORIGIN

  Query Match           14.1%;  Score 18;  DB 16;  Length 510;
  Best Local Similarity  90.0%;  Pred. No. 3.50e-04;
  Matches   18;  Conservative    0;  Mismatches   2;  Indels   0;  Gaps   0;

Db    429 AAAATTCNGCTTTAAAAANT 448
          ||||||| |||||||||| |
Cp     26 AAAATTCTGCTTTAAAAATT 7
```

```
RESULT   14
LOCUS       AQ245173      536 bp    DNA             GSS       02-OCT-1998
DEFINITION  HS_2008_A1_E10_T7 CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=2008 Col=19 Row=I, genomic survey
            sequence.
ACCESSION   AQ245173
NID         g3691747
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 536)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Furlong,J.,
            Shaker,R., Schmidt,S., Traicoff,R. and Hood,L.E.
  TITLE     Construction of a Characterized Clone Resource for Genomic
            Sequencing
  JOURNAL   Unpublished (1998)
COMMENT
            Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 2008  row: I  column: 19
            Class: BAC ends
            High quality sequence stop: 536.
FEATURES             Location/Qualifiers
     source          1. .536
                     /organism="Homo sapiens"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
                     /db_xref="taxon:9606"
                     /clone="Plate=2008 Col=19 Row=I"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
BASE COUNT      135 a    119 c    112 g    161 t      9 others
ORIGIN

  Query Match          14.1%;  Score 18;  DB 29;  Length 536;
  Best Local Similarity 100.0%;  Pred. No. 3.50e-04;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    472 GATGAGACACTTTGAGAA 489
          ||||||||||||||||||
Cp     89 GATGAGACACTTTGAGAA 72
```



```
RESULT   15
LOCUS       AI023497      111 bp    mRNA            EST       27-AUG-1998
DEFINITION  ov70h11.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1642725
            3', mRNA sequence.
ACCESSION   AI023497
NID         g3238541
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 111)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 533   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 107.
FEATURES             Location/Qualifiers
     source          1. .111
                     /organism="Homo sapiens"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from mRNA obtained from Clontech
                     Laboratories, Inc., and primed with  a Not I - oligo(dT)
                     primer (5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3').
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization to Cot5, and was
                     constructed  by Bento Soares and M. Fatima Bonaldo."
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1642725"
                     /clone_lib="Soares_testis_NHT"
                     /sex="male"
                     /lab_host="DH10B"
BASE COUNT       42 a     25 c     12 g     32 t
```



```
ORIGIN

  Query Match          13.3%;  Score 17;  DB 17;  Length 111;
  Best Local Similarity 100.0%;  Pred. No. 8.40e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      2 TTTTAAAAGAATAAAGA 18
          |||||||||||||||||
Qy     95 TTTTAAAAGAATAAAGA 111


RESULT   16
LOCUS       AA496694      152 bp    mRNA            EST       05-MAR-1998
DEFINITION  ae30e05.s1 Gessler Wilms tumor Homo sapiens cDNA clone 897344 3',
            mRNA sequence.
ACCESSION   AA496694
NID         g2230015
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 152)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
            Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
            Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1245   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 137.
FEATURES             Location/Qualifiers
     source          1. .152
                     /organism="Homo sapiens"
                     /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; RNA
                     was prepared from a pool of 6 anonymous Wilms' tumor RNAs.
                      RNA was prepared by acid-phenol, followed by one round of
                     oligo dT selection.   cDNA library preparation was with
                     the BRL/Life Tech. Superscript Plasmid system.  An
                     oligo-dT NotI primer for first strand synthesis generated
                     gcggccgccc(t)n at the 3' end of the clones.  A 5' SalI
```



```
                     adaptor was used  with sequence 5'-gtcgacccacgcgtccg-3'.
                     Resulting cDNAs were size selected (average size 2 kb),
                     NotI digested, and ligated into NotI/SalI-cut pSPORT1.
                     Library was constructed by Dr. Manfred Gessler."
                     /db_xref="taxon:9606"
                     /clone="897344"
                     /clone_lib="Gessler Wilms tumor"
                     /sex="pooled (6)"
                     /lab_host="DH10B"
BASE COUNT       41 a     16 c     34 g     61 t
ORIGIN

  Query Match          13.3%;  Score 17;  DB 13;  Length 152;
  Best Local Similarity 100.0%;  Pred. No. 8.40e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     27 TATTCTTTATTCTTTTA 43
          |||||||||||||||||
Cp    114 TATTCTTTATTCTTTTA 98


RESULT   17
LOCUS       AA513115      169 bp    mRNA            EST       20-AUG-1997
DEFINITION  nh78d10.s1 NCI_CGAP_Br1.1 Homo sapiens cDNA clone IMAGE:964627,
            mRNA sequence.
ACCESSION   AA513115
NID         g2251527
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 169)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
```

```
            Insert Length: 3429   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 132.
FEATURES              Location/Qualifiers
     source           1. .169
                      /organism="Homo sapiens"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; 1st strand cDNA was prepared from pooled bulk
                      breast tumor tissue, and was then primed with a Not I -
                      oligo(dT) primer. Double-stranded cDNA was ligated to Eco
                      RI adaptors (Pharmacia), digested with Not I and cloned
                      into the Not I and Eco RI sites of the modified pT7T3
                      vector.  Library is not normalized. (The normalized
                      version of this library is NCI_CGAP_Br2.) Library was
                      constructed by Bento Soares and M. Fatima Bonaldo."
                      /db_xref="taxon:9606"
                      /clone="IMAGE:964627"
                      /clone_lib="NCI_CGAP_Br1.1"
                      /sex="female, pooled"
                      /tissue_type="breast"
                      /lab_host="DH10B"
     mRNA             <1. .>169
BASE COUNT       75 a     23 c     24 g     47 t
ORIGIN

  Query Match           13.3%;  Score 17;  DB 8;  Length 169;
  Best Local Similarity 100.0%;  Pred. No. 8.40e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    135 AAAAGAATAAAGAATAT 151
          |||||||||||||||||
Qy     99 AAAAGAATAAAGAATAT 115


RESULT   18
LOCUS       AA025396      172 bp    mRNA            EST       01-FEB-1997
DEFINITION  ze90b03.s1 Soares fetal heart NbHH19W Homo sapiens cDNA clone
            366221 3', mRNA sequence.
ACCESSION   AA025396
NID         gl491562
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 172)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
            Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     WashU-Merck EST Project
```



```
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 755   Std Error: 0.00
            Seq primer: -40M13 fwd. from Amersham.
FEATURES              Location/Qualifiers
     source           1. .172
                      /organism="Homo sapiens"
                      /note="Organ: heart; Vector: pT7T3D (Pharmacia) with a
                      modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                      strand cDNA was primed with a Not I - oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCATCTTTTTTTTTTTTTTTT 3'],
                      double-stranded cDNA was size selected, ligated to Eco RI
                      adapters (Pharmacia), digested with Not I and cloned into
                      the Not I and Eco RI sites of a modified pT7T3 vector
                      (Pharmacia). Library went through one round of
                      normalization to a Cot = 5. Library constructed by
                      M.Fatima Bonaldo. This library was constructed from the
                      same fetus as the fetal lung library, Soares fetal lung
                      NbHL19W."
                      /db_xref="taxon:9606"
                      /clone="366221"
                      /clone_lib="Soares fetal heart NbHH19W"
                      /sex="unknown"
                      /dev_stage="19 weeks"
                      /lab_host="DH10B (ampicillin resistant)"
     mRNA             complement(<1. .>172)
BASE COUNT       76 a     23 c     23 g     49 t      1 others
ORIGIN

  Query Match           13.3%;  Score 17;  DB 22;  Length 172;
  Best Local Similarity 100.0%;  Pred. No. 8.40e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    146 AAAAGAATAAAGAATAT 162
          |||||||||||||||||
Qy     99 AAAAGAATAAAGAATAT 115


RESULT   19
LOCUS       AI185506      190 bp    mRNA            EST       09-OCT-1998
DEFINITION  qe54e09.x1 Soares_fetal_lung_NbHL19W Homo sapiens cDNA clone
            IMAGE:1742824 3', mRNA sequence.
```



```
ACCESSION   AI185506
NID         g3736144
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 190)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40UP from Giboo
            High quality sequence stop: 156.
FEATURES              Location/Qualifiers
     source           1. .190
                      /organism="Homo sapiens"
                      /note="Organ: lung; Vector: pT7T3D (Pharmacia) with a
                      modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                      strand cDNA was primed with a Not I - oligo(dT) primer
                      [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCAATTTTTTTTTTTTTTTT-3'],
                      double-stranded cDNA was size selected, ligated to Eco RI
                      adapters (Pharmacia), digested with Not I and cloned into
                      the Not I and Eco RI sites of a modified pT7T3 vector
                      (Pharmacia). Library went through one round of
                      normalization to a Cot = 5. Library constructed by Bento
                      Soares and M.Fatima Bonaldo. This library was constructed
                      from the same fetus as the fetal heart library, Soares
                      fetal heart NbHH19W."
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1742824"
                      /clone_lib="Soares_fetal_lung_NbHL19W"
                      /dev_stage="19 weeks"
                      /lab_host="DH10B (ampicillin resistant)"
BASE COUNT       77 a     25 c     24 g     64 t
ORIGIN

  Query Match           13.3%;  Score 17;  DB 18;  Length 190;
  Best Local Similarity 100.0%;  Pred. No. 8.40e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    153 AAAAGAATAAAGAATAT 169
          |||||||||||||||||
Qy     99 AAAAGAATAAAGAATAT 115
```



```
RESULT   20
LOCUS       AA775160      202 bp    mRNA            EST       05-FEB-1998
DEFINITION  ac78e09.s1 Stratagene lung (#937210) Homo sapiens cDNA clone 868744
            3', mRNA sequence.
ACCESSION   AA775160
NID         g2834494
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 202)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
            Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
            Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 147.
FEATURES              Location/Qualifiers
     source           1. .202
                      /organism="Homo sapiens"
                      /note="Organ: lung; Vector: pBluescript SK-; Site_1:
                      EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                      Oligo dT. normal lung. Average insert size: 1.0 kb;
                      Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCACGAG
                      3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTTT 3'"
                      /db_xref="taxon:9606"
                      /clone="868744"
                      /clone_lib="Stratagene lung (#937210)"
                      /sex="male"
                      /dev_stage="72 years"
                      /lab_host="SOLR cells (kanamycin resistant)"
BASE COUNT       91 a     26 c     30 g     55 t
ORIGIN

  Query Match           13.3%;  Score 17;  DB 12;  Length 202;
  Best Local Similarity 100.0%;  Pred. No. 8.40e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```

```
Db    135 AAAAGAATAAAGAATAT 151
          |||||||||||||||||
Qy     99 AAAAGAATAAAGAATAT 115


RESULT   21
LOCUS       AA702752      241 bp    mRNA            EST       19-DEC-1997
DEFINITION  zi82g01.s1 Soares fetal liver spleen 1NFLS S1 Homo sapiens cDNA
            clone 447312 3', mRNA sequence.
ACCESSION   AA702752
NID         g2705865
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 241)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
            Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
            Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 107.
FEATURES             Location/Qualifiers
     source          1. .241
                     /organism="Homo sapiens"
                     /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                     with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                     This is a subtracted version of the original Soares fetal
                     liver spleen 1NFLS library.  1st strand cDNA was primed
                     with a Pac I - oligo(dT) primer [5'
                     AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Pac I and cloned into the Pac I
                     and Eco RI sites of the modified pT7T3 vector.  Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
                     /db_xref="GDB:1351569"
                     /db_xref="taxon:9606"
                     /clone="447312"
```



```
                     /clone_lib="Soares fetal liver spleen 1NFLS S1"
                     /sex="male"
                     /dev_stage="20 week-post conception fetus"
                     /lab_host="DH10B (ampicillin resistant)"
BASE COUNT      101 a      38 c      37 g      65 t
ORIGIN

  Query Match           13.3%;  Score 17;  DB 11;  Length 241;
  Best Local Similarity 100.0%;  Pred. No. 8.40e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    137 AAAAGAATAAAGAATAT 153
          |||||||||||||||||
Qy     99 AAAAGAATAAAGAATAT 115


RESULT   22
LOCUS       AA779558      250 bp    mRNA            EST       05-FEB-1998
DEFINITION  af24g10.s1 Soares total fetus Nb2HF8 9w Homo sapiens cDNA clone
            1032642 3', mRNA sequence.
ACCESSION   AA779558
NID         g2838889
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 250)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
            Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
            Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 231.
FEATURES             Location/Qualifiers
     source          1. .250
                     /organism="Homo sapiens"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from mRNA obtained from pooled 8-9 week
```



```
                     (total) fetus material with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCGCTTAATTTTTTTTTTTTTTTT 3'].
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed  by Bento Soares and M. Fatima Bonaldo."
                     /db_xref="taxon:9606"
                     /clone="1032642"
                     /clone_lib="Soares total fetus Nb2HF8 9w"
                     /dev_stage="8-9 weeks"
                     /lab_host="DH10B"
BASE COUNT       90 a      46 c      49 g      65 t
ORIGIN

  Query Match           13.3%;  Score 17;  DB 12;  Length 250;
  Best Local Similarity 100.0%;  Pred. No. 8.40e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     15 AAAAGAATAAAGAATAT 31
          |||||||||||||||||
Qy     99 AAAAGAATAAAGAATAT 115


RESULT   23
LOCUS       AA552090      306 bp    mRNA            EST       04-SEP-1997
DEFINITION  ng48d01.s1 NCI_CGAP_Co3 Homo sapiens cDNA clone IMAGE:938017, mRNA
            sequence.
ACCESSION   AA552090
NID         g2322342
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 306)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck,
            M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arraying: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
```



```
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 659  Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 292.
FEATURES             Location/Qualifiers
     source          1. .306
                     /organism="Homo sapiens"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from 12 pooled bulk tumor samples and primed
                     with a Not I - oligo(dT) primer. Double-stranded cDNA was
                     ligated to Eco RI adaptors (Pharmacia), digested with Not
                     I and cloned into the Not I and Eco RI sites of the
                     modified pT7T3 vector. Library went through one round of
                     normalization."
                     /db_xref="taxon:9606"
                     /clone="IMAGE:938017"
                     /clone_lib="NCI_CGAP_Co3"
                     /sex="pooled"
                     /tissue_type="colon"
                     /lab_host="DH10B"
     mRNA            <1. .>306
BASE COUNT      120 a      50 c      53 g      83 t
ORIGIN

  Query Match           13.3%;  Score 17;  DB 9;  Length 306;
  Best Local Similarity 100.0%;  Pred. No. 8.40e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    138 AAAAGAATAAAGAATAT 154
          |||||||||||||||||
Qy     99 AAAAGAATAAAGAATAT 115


RESULT   24
LOCUS       AA579359      309 bp    mRNA            EST       12-SEP-1997
DEFINITION  nf29c03.s1 NCI_CGAP_Pr1 Homo sapiens cDNA clone IMAGE:915172, mRNA
            sequence.
ACCESSION   AA579359
NID         g2357543
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 309)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
```

```
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui,
            M.D., Michael Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 1122   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham.
FEATURES              Location/Qualifiers
     source           1. .309
                      /organism="Homo sapiens"
                      /note="Vector: pAMP10; Site_1: NotI; Site_2: EcoRI; 1st
                      strand cDNA was primed with oligo(dT)17 on 50 ng of
                      DNAse-treated, total cellular RNA obtained from
                      5,000-10,000 microdissected, histologically normal
                      prostate epithelial cells. Double-stranded cDNA was
                      ligated to EcoRI adaptors, 5 cycles of PCR applied to the
                      cDNA with an adaptor-specific primer, and the resulting
                      PCR product subcloned into pAMP10 by the UDG-cloning
                      method (Life Technologies).  Average insert size is 600
                      bp. NOTE: Not directionally cloned.  This library was
                      constructed by David Krizman."
                      /db_xref="taxon:9606"
                      /clone="IMAGE:915172"
                      /clone_lib="NCI_CGAP_Pr1"
                      /sex="Male"
                      /dev_stage="45 years old"
                      /lab_host="DH10B"
     mRNA             <1. .>309
BASE COUNT      120 a     49 c     50 g     90 t
ORIGIN

  Query Match           13.3%;  Score 17;  DB 9;  Length 309;
  Best Local Similarity 100.0%;  Pred. No. 8.40e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    158 AAAAGAATAAAGAATAT 174
          |||||||||||||||||
Qy     99 AAAAGAATAAAGAATAT 115


RESULT   25
LOCUS       AA007422      315 bp    mRNA            EST       09-MAY-1997
DEFINITION  zh98g03.s1 Soares fetal liver spleen 1NFLS S1 Homo sapiens cDNA
```



```
            clone 429364 3', mRNA sequence.
ACCESSION   AA007422
NID         g1463408
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 315)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
            Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1038   Std Error: 0.00
            Seq primer: mob.REGA+ET
            High quality sequence stop: 257.
FEATURES              Location/Qualifiers
     source           1. .315
                      /organism="Homo sapiens"
                      /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                      with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                      This is a subtracted version of the original Soares fetal
                      liver spleen 1NFLS library.  1st strand cDNA was primed
                      with a Pac I - oligo(dT) primer [5'
                      AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTT 3'],
                      double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Pac I and cloned into the Pac I
                      and Eco RI sites of the modified pT7T3 vector.  Library
                      went through one round of normalization. Library
                      constructed by Bento Soares and M.Fatima Bonaldo."
                      /db_xref="taxon:9606"
                      /clone="429364"
                      /clone_lib="Soares fetal liver spleen 1NFLS S1"
                      /sex="male"
                      /dev_stage="20 week-post conception fetus"
                      /lab_host="DH10B (ampicillin resistant)"
     mRNA             complement(<1. .>315)
                      /db_xref="GDB:1329133"
```



```
BASE COUNT      124 a     48 c     57 g     85 t      1 others
ORIGIN

  Query Match           13.3%;  Score 17;  DB 6;  Length 315;
  Best Local Similarity 100.0%;  Pred. No. 8.40e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    144 AAAAGAATAAAGAATAT 160
          |||||||||||||||||
Qy     99 AAAAGAATAAAGAATAT 115


RESULT   26
LOCUS       AA804853      336 bp    mRNA            EST       19-FEB-1998
DEFINITION  of44g06.s1 NCI_CGAP_CNS1 Homo sapiens cDNA clone IMAGE:1427098,
            mRNA sequence.
ACCESSION   AA804853
NID         g2876180
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 336)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            unknown library type
            Insert Length: 1136   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 329.
FEATURES              Location/Qualifiers
     source           1. .336
                      /organism="Homo sapiens"
                      /note="Organ: brain; Vector: pCMV-SPORT4; Site_1: SalI;
                      Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                      Average insert size 1.0 kb."
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1427098"
                      /clone_lib="NCI_CGAP_CNS1"
                      /tissue_type="substantia nigra"
                      /lab_host="DH10B"
BASE COUNT      128 a     51 c     56 g     101 t
ORIGIN

  Query Match           13.3%;  Score 17;  DB 13;  Length 336;
```



```
  Best Local Similarity 100.0%;  Pred. No. 8.40e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    146 AAAAGAATAAAGAATAT 162
          |||||||||||||||||
Qy     99 AAAAGAATAAAGAATAT 115


RESULT   27
LOCUS       AA114829      341 bp    mRNA            EST       11-MAY-1997
DEFINITION  zk88e05.s1 Soares pregnant uterus NbHPU Homo sapiens cDNA clone
            489920 3', mRNA sequence.
ACCESSION   AA114829
NID         g1669951
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 341)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
            Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 889   Std Error: 0.00
            Seq primer: -40M13 fwd. from Amersham
            High quality sequence stop: 258.
FEATURES              Location/Qualifiers
     source           1. .341
                      /organism="Homo sapiens"
                      /note="Organ: uterus; Vector: pT7T3-Pac; Site_1: Not I;
                      Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                      oligo(dT) primer [5'
                      AACTGGAAGAATTCGCGGCCGCCTTTTTTTTTTTTTTTTT 3'],
                      double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Not I and cloned into the Not I
                      and Eco RI sites of the modified pT7T3 vector. Library
                      went through one round of normalization.  Library
```

```
                    constructed by M. Fatima Bonaldo."
                    /db_xref="taxon:9606"
                    /clone="489920"
                    /clone_lib="Soares pregnant uterus NbHPU"
                    /sex="female"
                    /dev_stage="adult"
                    /lab_host="DH10B"
     mRNA           complement(<1. .>341)
                    /db_xref="GDB:3804507"
BASE COUNT      130 a     56 c     61 g     92 t      2 others
ORIGIN

  Query Match           13.3%;  Score 17;  DB 6;  Length 341;
  Best Local Similarity 100.0%;  Pred. No. 8.40e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    137 AAAAGAATAAAGAATAT 153
          |||||||||||||||||
Qy     99 AAAAGAATAAAGAATAT 115


RESULT   28
LOCUS       AI004353      341 bp    mRNA            EST       27-AUG-1998
DEFINITION  ot95b09.x1 Soares_total_fetus_Nb2HF8_9w Homo sapiens cDNA clone
            IMAGE:1624505 3', mRNA sequence.
ACCESSION   AI004353
NID         g3213863
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 341)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 2102   Std Error: 0.00
            Seq primer: primer name ambiguous
            High quality sequence stop: 281.
FEATURES            Location/Qualifiers
     source         1. .341
                    /organism="Homo sapiens"
                    /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                    polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
```



```
                    was prepared from mRNA obtained from pooled 8-9 week
                    (total) fetus material with a Not I - oligo(dT) primer [5'
                    TGTTACCAATCTGAAGTGGGAGCGGCCGCTTAATTTTTTTTTTTTTTTT 3'].
                    Double-stranded cDNA was ligated to Eco RI adaptors
                    (Pharmacia), digested with Not I and cloned into the Not I
                    and Eco RI sites of the modified pT7T3 vector. Library
                    went through one round of normalization, and was
                    constructed  by Bento Soares and M. Fatima Bonaldo."
                    /db_xref="taxon:9606"
                    /clone="IMAGE:1624505"
                    /clone_lib="Soares_total_fetus_Nb2HF8_9w"
                    /dev_stage="8-9 weeks"
                    /lab_host="DH10B"
BASE COUNT      127 a     52 c     57 g    105 t
ORIGIN

  Query Match           13.3%;  Score 17;  DB 17;  Length 341;
  Best Local Similarity 100.0%;  Pred. No. 8.40e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    149 AAAAGAATAAAGAATAT 165
          |||||||||||||||||
Qy     99 AAAAGAATAAAGAATAT 115


RESULT   29
LOCUS       R22948        343 bp    mRNA            EST       20-APR-1995
DEFINITION  yh31a06.s1 Homo sapiens cDNA clone 131314 3'.
ACCESSION   R22948
NID         g777836
KEYWORDS    EST.
SOURCE      human clone=131314 library=Soares placenta Nb2HP vector=pT7T3D
            (Pharmacia) with a modified polylinker host=DH10B (ampicillin
            resistant) primer=-21m13 Rsite1=Not I Rsite2=Eco RI Female placenta
            obtained at birth (full term). 1st strand cDNA was primed with a
            Not I - oligo(dT) primer [5'
            AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTTT 3'], double-stranded
            cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not
            I and cloned into the Not I and Eco RI sites of the modified pT7T3
            vector. Library went through one round of normalization. Library
            constructed by Bento Soares and M.Fatima Bonaldo.
  ORGANISM  Homo sapiens
            Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 343)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
```



```
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 173
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES            Location/Qualifiers
     source         1. .343
                    /organism="Homo sapiens"
                    /clone="131314"
BASE COUNT      119 a     55 c     73 g     93 t      3 others
ORIGIN

  Query Match           13.3%;  Score 17;  DB 5;  Length 343;
  Best Local Similarity 100.0%;  Pred. No. 8.40e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    147 AAAAGAATAAAGAATAT 163
          |||||||||||||||||
Qy     99 AAAAGAATAAAGAATAT 115


RESULT   30
LOCUS       AA485269      343 bp    mRNA            EST       11-AUG-1997
DEFINITION  ab07g06.s1 Stratagene lung (#937210) Homo sapiens cDNA clone 840154
            3', mRNA sequence.
ACCESSION   AA485269
NID         g2214488
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 343)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J.,
            Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-Merck EST Project 1997
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Wilson RK
            Washington University School of Medicine
```



```
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -41m13 fwd. ET from Amersham
            High quality sequence stop: 338.
FEATURES            Location/Qualifiers
     source         1. .343
                    /organism="Homo sapiens"
                    /note="Organ: lung; Vector: pBluescript SK-; Site_1:
                    EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                    Oligo dT, normal lung.  Average insert size: 1.0 kb;
                    Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCACGAG
                    3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
                    /db_xref="taxon:9606"
                    /clone="840154"
                    /clone_lib="Stratagene lung (#937210)"
                    /sex="male"
                    /dev_stage="72 years"
                    /lab_host="SOLR cells (kanamycin resistant)"
     mRNA           complement(<1. .>343)
BASE COUNT      129 a     51 c     57 g    106 t
ORIGIN

  Query Match           13.3%;  Score 17;  DB 8;  Length 343;
  Best Local Similarity 100.0%;  Pred. No. 8.40e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    155 AAAAGAATAAAGAATAT 171
          |||||||||||||||||
Qy     99 AAAAGAATAAAGAATAT 115


RESULT   31
LOCUS       AA200434      353 bp    mRNA            EST       19-FEB-1997
DEFINITION  mu35g12.r1 Soares 2NbMT Mus musculus cDNA clone 641446 5', mRNA
            sequence.
ACCESSION   AA200434
NID         g1796704
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Mus.
REFERENCE   1  (bases 1 to 353)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
```

```
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:393438
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 318.
FEATURES              Location/Qualifiers
     source           1. .353
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                      was primed with a Not I - oligo(dT) primer [5'
                      TGTTACAATCTGAAGTGGGAGCGGCCGCGTTTTTTTTTTTTTTTTTTTTTT
                      3']; double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Not I and cloned into the Not I
                      and Eco RI sites of the modified pT7T3 vector. RNA
                      provided by Dr. Bertrand Jordan. Library went through two
                      rounds of normalization, and was constructed by Bento
                      Soares and M.Fatima Bonaldo."
                      /db_xref="taxon:10090"
                      /clone="641446"
                      /clone_lib="Soares 2NbMT"
                      /sex="male"
                      /tissue_type="Thymus"
                      /dev_stage="4 weeks"
                      /lab_host="DH10B"
     mRNA             <1. .>353
BASE COUNT      101 a     67 c     84 g    101 t
ORIGIN

  Query Match           13.3%;  Score 17;  DB 23;  Length 353;
  Best Local Similarity 100.0%;  Pred. No. 8.40e-03;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    276 TTTATTCTTTTAAAAGT 292
          |||||||||||||||||
Cp    109 TTTATTCTTTTAAAAGT 93
```



```
RESULT   32
LOCUS       AQ098400      360 bp    DNA             GSS       26-AUG-1998
DEFINITION  HS_3046_A2_E12_MF CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3046 Col=24 Row=I, genomic survey
            sequence.
ACCESSION   AQ098400
NID         g3469429
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 360)
  AUTHORS   Mahairas,G.G., Zackrone,K.D., Smith,T., Tipton,S., Schmidt,S.,
            Traicoff,R., Abajian,C., Blanchard,A., West,A. and Hood,L.E.
  TITLE     Construction of a Characterized Clone Resource for Genomic
            Sequencing: Generation and Preliminary Analysis of 20,000 Sequence
            Tagged Connectors
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3046  row: I  column: 24
            Class: BAC ends
            High quality sequence stop: 360.
FEATURES              Location/Qualifiers
     source           1. .360
                      /organism="Homo sapiens"
                      /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                      E-Coli DH10B"
                      /db_xref="taxon:9606"
                      /clone="Plate=3046 Col=24 Row=I"
                      /clone_lib="CIT Approved Human Genomic Sperm Library D"
                      /sex="male"
BASE COUNT       78 a     96 c     74 g    112 t
ORIGIN

  Query Match           13.3%;  Score 17;  DB 28;  Length 360;
  Best Local Similarity 100.0%;  Pred. No. 8.40e-03;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    252 TCATTTTTTTGTCTACC 268
          |||||||||||||||||
Qy     37 TCATTTTTTTGTCTACC 53
```



```
RESULT   33
LOCUS       AA693353      366 bp    mRNA            EST       16-DEC-1997
DEFINITION  ah19e01.s1 Gessler Wilms tumor Homo sapiens cDNA clone 1157112 3',
            mRNA sequence.
ACCESSION   AA693353
NID         g2694291
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 366)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krisman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
            Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
            Theising,B., White,T., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40m13 fwd. ET from Amersham.
FEATURES              Location/Qualifiers
     source           1. .366
                      /organism="Homo sapiens"
                      /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; RNA
                      was prepared from a pool of 6 anonymous Wilms' tumor RNAs.
                       RNA was prepared by acid-phenol, followed by one round of
                      oligo dT selection.  cDNA library preparation was with
                      the BRL/Life Tech. Superscript Plasmid system.  An
                      oligo-dT NotI primer for first strand synthesis generated
                      gcggccgccc(t)n at the 3' end of the clones.  A 5' SalI
                      adaptor was used  with sequence 5'-gtcgacccacgcgtccg-3'.
                      Resulting cDNAs were size selected (average size 2 kb),
                      NotI digested, and ligated into NotI/SalI-cut pSPORT1.
                      Library was constructed by Dr. Manfred Gessler."
                      /db_xref="taxon:9606"
                      /clone="1157112"
                      /clone_lib="Gessler Wilms tumor"
                      /sex="pooled (6)"
                      /lab_host="DH10B"
BASE COUNT      139 a     61 c     62 g    104 t
ORIGIN
```



```
  Query Match           13.3%;  Score 17;  DB 11;  Length 366;
  Best Local Similarity 100.0%;  Pred. No. 8.40e-03;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    145 AAAAGAATAAAGAATAT 161
          |||||||||||||||||
Qy     99 AAAAGAATAAAGAATAT 115


RESULT   34
LOCUS       AA953621      368 bp    mRNA            EST       07-MAY-1998
DEFINITION  oq65a08.s1 NCI_CGAP_Kid6 Homo sapiens cDNA clone IMAGE:1591190 3',
            mRNA sequence.
ACCESSION   AA953621
NID         g3117768
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 368)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: Stratagene, Inc.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 328.
FEATURES              Location/Qualifiers
     source           1. .368
                      /organism="Homo sapiens"
                      /note="Organ: kidney; Vector: Bluescript SK-; Site_1:
                      EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                      Oligo dT. Pooled kidney tumors. 5' adaptor sequence: 5'
                      GAATTCGGCACGAG 3' 3' adaptor sequence: 5'
                      CTCGAGTTTTTTTTTTTTTTTT 3' Average insert size:  1.0 kb."
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1591190"
                      /clone_lib="NCI_CGAP_Kid6"
```

```
                     /sex="mixed"
                     /tissue_type="kidney tumor"
                     /lab_host="SOLR (kanamycin resistant)"
BASE COUNT       140 a      60 c      66 g     102 t
ORIGIN

  Query Match           13.3%;  Score 17;  DB 14;  Length 368;
  Best Local Similarity 100.0%;  Pred. No. 8.40e-03;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    135 AAAAGAATAAAGAATAT 151
          |||||||||||||||||
Qy     99 AAAAGAATAAAGAATAT 115


RESULT   35
LOCUS       AI056065      376 bp    mRNA            EST       20-JUL-1998
DEFINITION  ox46d08.s1 Soares_total_fetus_Nb2HF8_9w Homo sapiens cDNA clone
            IMAGE:1659375 3', mRNA sequence.
ACCESSION   AI056065
NID         g3329931
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 376)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 334.
FEATURES             Location/Qualifiers
     source          1. .376
                     /organism="Homo sapiens"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from mRNA obtained from pooled 8-9 week
                     (total) fetus material with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCTTAATTTTTTTTTTTTTTTT 3']).
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
```



```
                     constructed  by Bento Soares and M. Fatima Bonaldo."
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1659375"
                     /clone_lib="Soares_total_fetus_Nb2HF8_9w"
                     /dev_stage="8-9 weeks"
                     /lab_host="DH10B"
BASE COUNT       139 a      61 c      63 g     113 t
ORIGIN

  Query Match           13.3%;  Score 17;  DB 15;  Length 376;
  Best Local Similarity 100.0%;  Pred. No. 8.40e-03;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    148 AAAAGAATAAAGAATAT 164
          |||||||||||||||||
Qy     99 AAAAGAATAAAGAATAT 115


RESULT   36
LOCUS       AA864358      377 bp    mRNA            EST       13-MAY-1998
DEFINITION  oh54f06.s1 NCI_CGAP_GC4 Homo sapiens cDNA clone IMAGE:1470467 3',
            mRNA sequence.
ACCESSION   AA864358
NID         g2958671
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 377)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 1143   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .377
```



```
                     /organism="Homo sapiens"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; 1st strand cDNA was prepared from 3 pooled
                     germ cell tumors, and was then primed with a Not I -
                     oligo(dT) primer. Double-stranded cDNA was ligated to Eco
                     RI adaptors (Pharmacia), digested with Not I and cloned
                     into the Not I and Eco RI sites of the modified pT7T3
                     vector.  Library is normalized. Library was constructed by
                     Bento Soares and M. Fatima Bonaldo."
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1470467"
                     /clone_lib="NCI_CGAP_GC4"
                     /tissue_type="pooled germ cell tumors"
                     /lab_host="DH10B"
BASE COUNT       140 a      64 c      68 g     105 t
ORIGIN

  Query Match           13.3%;  Score 17;  DB 14;  Length 377;
  Best Local Similarity 100.0%;  Pred. No. 8.40e-03;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    137 AAAAGAATAAAGAATAT 153
          |||||||||||||||||
Qy     99 AAAAGAATAAAGAATAT 115


RESULT   37
LOCUS       AA648400      378 bp    mRNA            EST       13-NOV-1997
DEFINITION  ns22a01.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1184328,
            mRNA sequence.
ACCESSION   AA648400
NID         g2574829
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 378)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            unknown library type
            Insert Length: 461   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .378
```



```
                     /organism="Homo sapiens"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from human tonsillar cells enriched for
                     germinal center B cells by flow sorting (CD20+, IgD-),
                     provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                     (NCI) and Dr. Gerald Marti (CBER).  cDNA synthesis was
                     primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTT-
                     3']. Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo."
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1184328"
                     /clone_lib="NCI_CGAP_GCB1"
                     /tissue_type="germinal center B cell"
                     /lab_host="DH10B"
BASE COUNT       135 a      60 c      63 g     120 t
ORIGIN

  Query Match           13.3%;  Score 17;  DB 10;  Length 378;
  Best Local Similarity 100.0%;  Pred. No. 8.40e-03;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    159 AAAAGAATAAAGAATAT 175
          |||||||||||||||||
Qy     99 AAAAGAATAAAGAATAT 115


RESULT   38
LOCUS       AA766793      381 bp    mRNA            EST       08-FEB-1998
DEFINITION  oa35f06.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1306979,
            mRNA sequence.
ACCESSION   AA766793
NID         g2818031
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 381)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
```

```
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
            Ph.D., Gerald Marti, M.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 461   Std Error: 0.00
            Seq primer: -40ml3 fwd. ET from Amersham
            High quality sequence stop: 380.
  FEATURES              Location/Qualifiers
       source           1. .381
                        /organism="Homo sapiens"
                        /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                        polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                        was prepared from human tonsillar cells enriched for
                        germinal center B cells by flow sorting (CD20+, IgD-),
                        provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                        (NCI) and Dr. Gerald Marti (CBER).  cDNA synthesis was
                        primed with a Not I - oligo(dT) primer
                        [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTT-
                        3']. Double-stranded cDNA was ligated to Eco RI adaptors
                        (Pharmacia), digested with Not I and cloned into the Not I
                        and Eco RI sites of the modified pT7T3 vector. Library
                        went through one round of normalization, and was
                        constructed by Bento Soares and M. Fatima Bonaldo."
                        /db_xref="taxon:9606"
                        /clone="IMAGE:1306979"
                        /clone_lib="NCI_CGAP_GCB1"
                        /tissue_type="germinal center B cell"
                        /lab_host="DH10B"
  BASE COUNT      137 a     61 c     63 g    120 t
  ORIGIN

    Query Match           13.3%;  Score 17;  DB 12;  Length 381;
    Best Local Similarity 100.0%;  Pred. No. 8.40e-03;
    Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

  Db   159 AAAAGAATAAAGAATAT 175
           |||||||||||||||||
  Qy    99 AAAAGAATAAAGAATAT 115


  RESULT   39
  LOCUS       AI034146      383 bp    mRNA            EST       28-AUG-1998
  DEFINITION  ow15c06.x1 Soares_parathyroid_tumor_NbHPA Homo sapiens cDNA clone
              IMAGE:1646890 3', mRNA sequence.
  ACCESSION   AI034146
```



```
  NID         g3255099
  KEYWORDS    EST.
  SOURCE      human.
    ORGANISM  Homo sapiens
              Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
              Primates; Catarrhini; Hominidae; Homo.
  REFERENCE   1  (bases 1 to 383)
    AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
    TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
              Tumor Gene Index
    JOURNAL   Unpublished (1997)
  COMMENT
              Contact: Robert Strausberg, Ph.D.
              Tel: (301) 496-1550
              Email: Robert_Strausberg@nih.gov
              cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
              Bonaldo, Ph.D.
               cDNA Library Arrayed by: Greg Lennon, Ph.D.
               DNA Sequencing by: Washington University Genome Sequencing Center
               Clone distribution: NCI-CGAP clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              www-bio.llnl.gov/bbrp/image/image.html

              Insert Length: 887   Std Error: 0.00
              Seq primer: -40ml3 fwd. ET from Amersham
              High quality sequence stop: 369.
  FEATURES              Location/Qualifiers
       source           1. .383
                        /organism="Homo sapiens"
                        /note="Organ: parathyroid gland; Vector: pT7T3D
                        (Pharmacia) with a modified polylinker; Site_1: Not I;
                        Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                        oligo(dT) primer
                        [5'-
                        TGTTACCAATCTGAAGTGGGAGCGGCCGCACCAATTTTTTTTTTTTTTTTTTTTT
                        T-3'], double-stranded cDNA was size selected, ligated to
                        Eco RI adapters (Pharmacia), digested with Not I and
                        cloned into the Not I and Eco RI sites of a modified pT7T3
                        vector (Pharmacia). Library went through one round of
                        normalization to a Cot = 5. Library constructed by Bento
                        Soares and M.Fatima Bonaldo. RNA from sporadic parathyroid
                        adenomas was kindly provided by Dr. Stephen Marx, National
                        Institute of Diabetes and Digestive and Kidney Diseases,
                        NIH."
                        /db_xref="taxon:9606"
                        /clone="IMAGE:1646890"
                        /clone_lib="Soares_parathyroid_tumor_NbHPA"
                        /tissue_type="parathyroid tumor"
                        /dev_stage="adult"
                        /lab_host="DH10B (ampicillin resistant)"
  BASE COUNT      143 a     62 c     67 g    111 t
```



```
  ORIGIN

    Query Match           13.3%;  Score 17;  DB 17;  Length 383;
    Best Local Similarity 100.0%;  Pred. No. 8.40e-03;
    Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

  Db   151 AAAAGAATAAAGAATAT 167
           |||||||||||||||||
  Qy    99 AAAAGAATAAAGAATAT 115


  RESULT   40
  LOCUS       AI006803      386 bp    mRNA            EST       12-JUN-1998
  DEFINITION  ua74b07.r1 Soares 2NbMT Mus musculus cDNA clone 1363189 5', mRNA
              sequence.
  ACCESSION   AI006803
  NID         g3216360
  KEYWORDS    EST.
  SOURCE      house mouse.
    ORGANISM  Mus musculus
              Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
              Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  REFERENCE   1  (bases 1 to 386)
    AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
              Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
              Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
              Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
              Waterston,R.
    TITLE     The WashU-HHMI Mouse EST Project
    JOURNAL   Unpublished (1996)
  COMMENT
              Contact: Marra M/Mouse EST Project
              WashU-HHMI Mouse EST Project
              Washington University School of MedicineP
              4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
              Tel: 314 286 1800
              Fax: 314 286 1810
              Email: mouseest@watson.wustl.edu
              This clone is available royalty-free through LLNL ; contact the
              IMAGE Consortium (info@image.llnl.gov) for further information.
              MGI:896409
              Seq primer: -28ml3 rev2 ET from Amersham
              High quality sequence stop: 351.
  FEATURES              Location/Qualifiers
       source           1. .386
                        /organism="Mus musculus"
                        /strain="C57BL/6J"
                        /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                        polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                        was primed with a Not I - oligo(dT) primer [5'
                        TGTTACCAATCTGAAGTGGGAGCGGCCGCGTTTTTTTTTTTTTTTTTTTTTTT
```



```
                        3']; double-stranded cDNA was ligated to Eco RI adaptors
                        (Pharmacia), digested with Not I and cloned into the Not I
                        and Eco RI sites of the modified pT7T3 vector. RNA
                        provided by Dr. Bertrand Jordan. Library went through two
                        rounds of normalization, and was constructed by Bento
                        Soares and M.Fatima Bonaldo."
                        /db_xref="taxon:10090"
                        /clone="1363189"
                        /clone_lib="Soares 2NbMT"
                        /sex="male"
                        /tissue_type="Thymus"
                        /dev_stage="4 weeks"
                        /lab_host="DH10B"
  BASE COUNT      105 a     73 c    100 g    108 t
  ORIGIN

    Query Match           13.3%;  Score 17;  DB 15;  Length 386;
    Best Local Similarity 100.0%;  Pred. No. 8.40e-03;
    Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

  Db   348 TTTATTCTTTTAAAAGT 364
           |||||||||||||||||
  Cp   109 TTTATTCTTTTAAAAGT 93


  RESULT   41
  LOCUS       AA910013      388 bp    mRNA            EST       26-AUG-1998
  DEFINITION  ol40a12.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
              IMAGE:1525918 3', mRNA sequence.
  ACCESSION   AA910013
  NID         g3049303
  KEYWORDS    EST.
  SOURCE      human.
    ORGANISM  Homo sapiens
              Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
              Primates; Catarrhini; Hominidae; Homo.
  REFERENCE   1  (bases 1 to 388)
    AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
    TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
              Tumor Gene Index
    JOURNAL   Unpublished (1997)
  COMMENT
              Contact: Robert Strausberg, Ph.D.
              Tel: (301) 496-1550
              Email: Robert_Strausberg@nih.gov
              This clone is available royalty-free through LLNL ; contact the
              IMAGE Consortium (info@image.llnl.gov) for further information.
              Insert Length: 756   Std Error: 0.00
              Seq primer: -40ml3 fwd. ET from Amersham
              High quality sequence stop: 342.
  FEATURES              Location/Qualifiers
```

```
     source          1. .388
                     /organism="Homo sapiens"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from three normalized
                     libraries (fetal lung NbHL19W, testis NHT, and B-cell
                     NCI_CGAP_GCB1) were mixed, and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR-amplified cDNAs from pools of 5,000 clones made
                     from the same 3 libraries. The pools consisted of
                     I.M.A.G.E. clones 297480-302087, 682632-687239,
                     726408-728711, and 729096-731399. Subtraction by Bento
                     Soares and M. Fatima Bonaldo."
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1525918"
                     /clone_lib="Soares_NFL_T_GBC_S1"
                     /lab_host="DH10B"
BASE COUNT      109 a     85 c     90 g    104 t
ORIGIN

  Query Match           13.3%;  Score 17;  DB 17;  Length 388;
  Best Local Similarity 100.0%;  Pred. No. 8.40e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     23 TATTCTTTATTCTTTTA 39
          |||||||||||||||||
Cp    114 TATTCTTTATTCTTTTA 98


RESULT   42
LOCUS       W02793         401 bp    mRNA            EST       25-NOV-1996
DEFINITION  zc64c10.s1 Soares fetal heart NbHH19W Homo sapiens cDNA clone
            327090 3', mRNA sequence.
ACCESSION   W02793
NID         g1274771
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 401)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
```



```
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 621   Std Error: 0.00
            Seq primer: mob.REGA+ET
            High quality sequence stop: 283.
FEATURES                Location/Qualifiers
     source             1. .401
                        /organism="Homo sapiens"
                        /note="Organ: heart; Vector: pT7T3D (Pharmacia) with a
                        modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                        strand cDNA was primed with a Not I - oligo(dT) primer [5'
                        TGTTACCAATCTGAAGTGGGAGCGGCCGCATCTTTTTTTTTTTTTTTTT 3'],
                        double-stranded cDNA was size selected, ligated to Eco RI
                        adapters (Pharmacia), digested with Not I and cloned into
                        the Not I and Eco RI sites of a modified pT7T3 vector
                        (Pharmacia). Library went through one round of
                        normalization to a Cot = 5. Library constructed by
                        M.Fatima Bonaldo. This library was constructed from the
                        same fetus as the fetal lung library, Soares fetal lung
                        NbHL19W."
                        /db_xref="taxon:9606"
                        /clone="327090"
                        /clone_lib="Soares fetal heart NbHH19W"
                        /sex="unknown"
                        /dev_stage="19 weeks"
                        /lab_host="DH10B (ampicillin resistant)"
     mRNA               complement(<1. .>401)
BASE COUNT      141 a     71 c     71 g    118 t
ORIGIN

  Query Match           13.3%;  Score 17;  DB 22;  Length 401;
  Best Local Similarity 100.0%;  Pred. No. 8.40e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    144 AAAAGAATAAAGAATAT 160
          |||||||||||||||||
Qy     99 AAAAGAATAAAGAATAT 115


RESULT   43
LOCUS       AA486288       410 bp    mRNA            EST       11-AUG-1997
DEFINITION  ab35h08.s1 Stratagene HeLa cell s3 937216 Homo sapiens cDNA clone
            842847 3', mRNA sequence.
ACCESSION   AA486288
NID         g2216504
```



```
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 410)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J.,
            Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-Merck EST Project 1997
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -41m13 fwd. ET from Amersham
            High quality sequence stop: 310.
FEATURES                Location/Qualifiers
     source             1. .410
                        /organism="Homo sapiens"
                        /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                        XhoI; Cloned unidirectionally.  Primer: Oligo dT. HeLa S3
                        epithelioid carcinoma cells grown to semi-confluency
                        without induction. Average insert size: 1.5 kb; Uni-ZAP XR
                        Vector. -5' adaptor sequence: 5' GAATTCGGCACGAG 3' -3'
                        adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
                        /db_xref="taxon:9606"
                        /clone="842847"
                        /clone_lib="Stratagene HeLa cell s3 937216"
                        /sex="female"
                        /dev_stage="HeLa S3 cell line"
                        /lab_host="SOLR (kanamycin resistant)"
     mRNA               complement(<1. .>410)
BASE COUNT      140 a     69 c     69 g    132 t
ORIGIN

  Query Match           13.3%;  Score 17;  DB 8;  Length 410;
  Best Local Similarity 100.0%;  Pred. No. 8.40e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    155 AAAAGAATAAAGAATAT 171
          |||||||||||||||||
Qy     99 AAAAGAATAAAGAATAT 115
```



```
RESULT   44
LOCUS       AI025564       413 bp    mRNA            EST       27-AUG-1998
DEFINITION  ov87c05.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1644296
            3' similar to WP:F45G2.8 CE16051 ;, mRNA sequence.
ACCESSION   AI025564
NID         g3241177
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 413)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 966   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 412.
FEATURES                Location/Qualifiers
     source             1. .413
                        /organism="Homo sapiens"
                        /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                        polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                        was prepared from mRNA obtained from Clontech
                        Laboratories, Inc., and primed with  a Not I - oligo(dT)
                        primer [5'
                        TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3'].
                        Double-stranded cDNA was ligated to Eco RI adaptors
                        (Pharmacia), digested with Not I and cloned into the Not I
                        and Eco RI sites of the modified pT7T3 vector. Library
                        went through one round of normalization to Cot5, and was
                        constructed  by Bento Soares and M. Fatima Bonaldo."
                        /db_xref="taxon:9606"
                        /clone="IMAGE:1644296"
                        /clone_lib="Soares_testis_NHT"
                        /sex="male"
                        /lab_host="DH10B"
```

```
BASE COUNT       83 a    116 c    100 g    114 t
ORIGIN

  Query Match           13.3%;  Score 17;  DB 17;  Length 413;
  Best Local Similarity 100.0%;  Pred. No. 8.40e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    250 GACACTTTGAGAATCTG 266
          |||||||||||||||||
Cp     84 GACACTTTGAGAATCTG 68


RESULT   45
LOCUS       H10401        416 bp    mRNA            EST       23-JUN-1995
DEFINITION  ym08e07.s1 Homo sapiens cDNA clone 47085 3'.
ACCESSION   H10401
NID         g875223
KEYWORDS    EST.
SOURCE      human clone=47085 library=Soares infant brain 1NIB vector=Lafmid BA
            host=DH10B (ampicillin resistant) primer=Promega -21m13 Rsite1=Not
            I Rsite2=Hind III Whole brain from a 73 days post natal female. 1st
            strand cDNA was primed with a Not I - oligo(dT) primer [5'
            AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTT 3']; double-stranded
            cDNA was ligated to Hind III adaptors (Pharmacia), digested with
            Not I and directionally cloned into the Not I and Hind III sites of
            the Lafmid BA vector. Library went through one round of
            normalization. Library constructed by Bento Soares and M.Fatima
            Bonaldo.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 416)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            GDB: G00-419-626
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 323
```



```
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES             Location/Qualifiers
     source          1..416
                     /organism="Homo sapiens"
                     /clone="47085"
BASE COUNT      136 a     75 c     81 g    123 t      1 others
ORIGIN

  Query Match           13.3%;  Score 17;  DB 16;  Length 416;
  Best Local Similarity 100.0%;  Pred. No. 8.40e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    145 AAAAGAATAAAGAATAT 161
          |||||||||||||||||
Qy     99 AAAAGAATAAAGAATAT 115


RESULT   46
LOCUS       N57697        419 bp    mRNA            EST       28-JAN-1997
DEFINITION  yv55b06.s1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone
            246611 3' similar to contains Alu repetitive element;, mRNA
            sequence.
ACCESSION   N57697
NID         g1201587
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 419)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
            Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1087  Std Error: 0.00
            Seq primer: m13 -40 forward
            High quality sequence stop: 325.
```



```
FEATURES             Location/Qualifiers
     source          1..419
                     /organism="Homo sapiens"
                     /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                     with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                     1st strand cDNA was primed with a Pac I - oligo(dT) primer
                     [5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Pac I and cloned into the Pac I
                     and Eco RI sites of the modified pT7T3 vector.  Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
                     /db_xref="taxon:9606"
                     /clone="246611"
                     /clone_lib="Soares fetal liver spleen 1NFLS"
                     /sex="male"
                     /dev_stage="20 week-post conception fetus"
                     /lab_host="DH10B (ampicillin resistant)"
     mRNA            complement(<1..>419)
BASE COUNT       99 a     97 c     85 g    138 t
ORIGIN

  Query Match           13.3%;  Score 17;  DB 22;  Length 419;
  Best Local Similarity 100.0%;  Pred. No. 8.40e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    152 TTTTAAAGCAGAATTTT 168
          |||||||||||||||||
Qy     10 TTTTAAAGCAGAATTTT 26


RESULT   47
LOCUS       AA490320      420 bp    mRNA            EST       15-AUG-1997
DEFINITION  aa51c09.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:824464 3',
            mRNA sequence.
ACCESSION   AA490320
NID         g2219493
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 420)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
```



```
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
            Ph.D., Gerald Marti, M.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -41m13 fwd. ET from Amersham
            High quality sequence stop: 419.
FEATURES             Location/Qualifiers
     source          1..420
                     /organism="Homo sapiens"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from human tonsillar cells enriched for
                     germinal center B cells by flow sorting (CD20+, IgD-),
                     provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                     (NCI) and Dr. Gerald Marti (CBER).  cDNA synthesis was
                     primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCTCATTTTTTTTTTTTTTTTT-
                     3']. Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo."
                     /db_xref="taxon:9606"
                     /clone="IMAGE:824464"
                     /clone_lib="NCI_CGAP_GCB1"
                     /tissue_type="germinal center B cell"
                     /lab_host="DH10B"
     mRNA            complement(<1..>420)
BASE COUNT      147 a     75 c     71 g    127 t
ORIGIN

  Query Match           13.3%;  Score 17;  DB 8;  Length 420;
  Best Local Similarity 100.0%;  Pred. No. 8.40e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    146 AAAAGAATAAAGAATAT 162
          |||||||||||||||||
Qy     99 AAAAGAATAAAGAATAT 115


RESULT   48
LOCUS       AA535623      422 bp    mRNA            EST       21-AUG-1997
DEFINITION  nf87f03.s1 NCI_CGAP_Co3 Homo sapiens cDNA clone IMAGE:926909, mRNA
            sequence.
```

```
ACCESSION   AA535623
NID         g2279876
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 422)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck,
            M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arraying: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 575   Std Error: 0.00
            Seq primer: -40ml3 fwd. ET from Amersham
            High quality sequence stop: 383.
FEATURES            Location/Qualifiers
     source         1. .422
                    /organism="Homo sapiens"
                    /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                    polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                    was prepared from 12 pooled bulk tumor samples and primed
                    with a Not I - oligo(dT) primer. Double-stranded cDNA was
                    ligated to Eco RI adaptors (Pharmacia), digested with Not
                    I and cloned into the Not I and Eco RI sites of the
                    modified pT7T3 vector. Library went through one round of
                    normalization."
                    /db_xref="taxon:9606"
                    /clone="IMAGE:926909"
                    /clone_lib="NCI_CGAP_Co3"
                    /sex="pooled"
                    /tissue_type="colon"
                    /lab_host="DH10B"
     mRNA           <1. .>422
BASE COUNT      144 a     86 c     80 g    112 t
ORIGIN

  Query Match           13.3%;  Score 17;  DB 9;  Length 422;
```



```
  Best Local Similarity 100.0%;  Pred. No. 8.40e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     38 AAAAGAATAAAGAATAT 54
          |||||||||||||||||
Qy     99 AAAAGAATAAAGAATAT 115


RESULT   49
LOCUS       AA424324       423 bp    mRNA            EST       16-OCT-1997
DEFINITION  zv82a06.s1 Soares total fetus Nb2HF8 9w Homo sapiens cDNA clone
            760114 3', mRNA sequence.
ACCESSION   AA424324
NID         g2103294
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 423)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J.,
            Moore,B., Schellenberg,K., Steptoe,M., Tan,P., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-Merck EST Project 1997
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -41ml3 fwd. ET from Amersham
            High quality sequence stop: 301.
FEATURES            Location/Qualifiers
     source         1. .423
                    /organism="Homo sapiens"
                    /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                    polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                    was prepared from mRNA obtained from pooled 8-9 week
                    (total) fetus material with a Not I - oligo(dT) primer (5'
                    TGTTACCAATCTGAAGTGGGAGCGGCCGCTTAATTTTTTTTTTTTTTTT 3').
                    Double-stranded cDNA was ligated to Eco RI adaptors
                    (Pharmacia), digested with Not I and cloned into the Not I
                    and Eco RI sites of the modified pT7T3 vector. Library
                    went through one round of normalization, and was
                    constructed  by Bento Soares and M. Fatima Bonaldo."
                    /db_xref="taxon:9606"
```



```
                    /clone="760114"
                    /clone_lib="Soares total fetus Nb2HF8 9w"
                    /dev_stage="8-9 weeks"
                    /lab_host="DH10B"
BASE COUNT      142 a     75 c     70 g    136 t
ORIGIN

  Query Match           13.3%;  Score 17;  DB 10;  Length 423;
  Best Local Similarity 100.0%;  Pred. No. 8.40e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    159 AAAAGAATAAAGAATAT 175
          |||||||||||||||||
Qy     99 AAAAGAATAAAGAATAT 115


RESULT   50
LOCUS       W04711         424 bp    mRNA            EST       10-OCT-1996
DEFINITION  zb94d02.s1 Soares parathyroid tumor NbHPA Homo sapiens cDNA clone
            320451 3', mRNA sequence.
ACCESSION   W04711
NID         g1277431
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 424)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 2630   Std Error: 0.00
            Seq primer: mob.REGA+ET
            High quality sequence stop: 340.
FEATURES            Location/Qualifiers
     source         1. .424
                    /organism="Homo sapiens"
```



```
                    /note="Organ: parathyroid gland; Vector: pT7T3D
                    (Pharmacia) with a modified polylinker; Site_1: Not I;
                    Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                    oligo(dT) primer
                    [5'-
                    TGTTACCAATCTGAAGTGGGAGCGGCCGCACCAATTTTTTTTTTTTTTTTTTTTT
                    T-3'], double-stranded cDNA was size selected, ligated to
                    Eco RI adapters (Pharmacia), digested with Not I and
                    cloned into the Not I and Eco RI sites of a modified pT7T3
                    vector (Pharmacia). Library went through one round of
                    normalization to a Cot = 5. Library constructed by Bento
                    Soares and M.Fatima Bonaldo. RNA from sporadic parathyroid
                    adenomas was kindly provided by Dr. Stephen Marx, National
                    Institute of Diabetes and Digestive and Kidney Diseases,
                    NIH."
                    /db_xref="taxon:9606"
                    /clone="320451"
                    /clone_lib="Soares parathyroid tumor NbHPA"
                    /tissue_type="parathyroid tumor"
                    /dev_stage="adult"
                    /lab_host="DH10B (ampicillin resistant)"
     mRNA           complement(<1. .>424)
BASE COUNT      149 a     78 c     77 g    120 t
ORIGIN

  Query Match           13.3%;  Score 17;  DB 21;  Length 424;
  Best Local Similarity 100.0%;  Pred. No. 8.40e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    137 AAAAGAATAAAGAATAT 153
          |||||||||||||||||
Qy     99 AAAAGAATAAAGAATAT 115


Search completed: Tue Dec 22 21:04:25 1998
Job time : 331 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Tue Dec 22 21:04:45 1998; MasPar time 33.03 Seconds
527.168 Million cell updates/sec

Tabular output not generated.

Title: >US-08-530-112A-1320
Description: (1-128) from US08530112A.seq
Perfect Score: 128
N.A. Sequence: 1 GATCAGAATTTTTAAAGCAG.....AGAATATCTTGCTGGGCAAA 128
Comp: CTAGTCTTAAAAATTTCGTC.....TCTTATAGAACGACCCGTTT

Scoring table: TABLE jmetric
Gap 60

Rmatch STD : Dbase 0; Query 0

Searched: 188442 seqs, 68026449 bases x 2

Post-processing: Minimum Match 0%
Listing first 1000 summaries

Database: n-geneseq32
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39 40:part40

Statistics: Mean 6.106; Variance 3.097; scale 1.971

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
	1	128	100.0	128	19	T20320	Human gene signature	7.55e-86
	2	19	14.8	1523	37	T99822	Human steroid 5alpha-	1.52e-01
c	3	18	14.1	2206	15	Q92782	Human thymopoietin ge	6.13e-01
c	4	18	14.1	2509	15	Q92776	Human thymopoietin ge	6.13e-01
c	5	17	13.3	1289	16	Q87016	Clone B2 encodes func	2.40e+00
c	6	17	13.3	4937	37	T85307	Arabidopsis thaliana	2.40e+00
	7	16	12.5	125	20	T03611	Protocadherin clone H	9.07e+00
	8	16	12.5	125	12	Q68987	Human-410 cadherin-re	9.07e+00
c	9	16	12.5	645	8	Q48470	Bacillus subtilis PYR	9.07e+00
c	10	16	12.5	645	9	Q50811	Bacillus subtilis pcp	9.07e+00
c	11	16	12.5	1146	40	V24778	H. pylori ORF 04gpl12	9.07e+00
	12	15	11.7	168	15	Q94401	Tyrosine kinase-like	3.30e+01
c	13	15	11.7	276	22	T22939	Human gene signature	3.30e+01
c	14	15	11.7	301	19	T20572	Human gene signature	3.30e+01
	15	15	11.7	378	8	Q59533	Human brain Expressed	3.30e+01
c	16	15	11.7	401	8	Q60061	Human brain Expressed	3.30e+01
	17	15	11.7	449	31	T43489	ATM gene exon 57.	3.30e+01
c	18	15	11.7	650	3	Q20588	Sequence from potato	3.30e+01
c	19	15	11.7	703	3	Q20587	Sequence from potato	3.30e+01
c	20	15	11.7	718	3	Q20586	Sequence from potato	3.30e+01
c	21	15	11.7	767	1	N82022	Gene encoding 25 kD s	3.30e+01
c	22	15	11.7	767	7	Q43503	Sequence encoding the	3.30e+01
	23	15	11.7	861	21	T13643	AcNPV ORF 27, residue	3.30e+01
c	24	15	11.7	1047	13	Q78152	Tuber specific class	3.30e+01
	25	15	11.7	1390	31	T76777	Staphylococcus aureus	3.30e+01
c	26	15	11.7	1431	1	Q03256	Sequence encoding bet	3.30e+01
	27	15	11.7	1657	31	T68767	ATM mutant 497del7514	3.30e+01
	28	15	11.7	1807	3	Q20243	Mycoplasma-specific D	3.30e+01
c	29	15	11.7	1969	1	N93087	Human basic fibroblas	3.30e+01
	30	15	11.7	2128	11	Q64158	Partial coding sequen	3.30e+01
	31	15	11.7	2255	10	Q56758	Chitinase derivative	3.30e+01
c	32	15	11.7	2781	40	V04155	Blood group antigen b	3.30e+01
	33	15	11.7	3157	15	Q92521	Human colonic adenoca	3.30e+01
	34	15	11.7	3157	35	T93784	CCK-2, a human mammar	3.30e+01
	35	15	11.7	3157	15	Q92523	Human colonic adenoca	3.30e+01
	36	15	11.7	3372	4	Q27254	Growth Factor Recepto	3.30e+01
	37	15	11.7	3372	19	T03198	Human GRB-1 genomic D	3.30e+01
	38	15	11.7	3852	3	N50094	DNA encoding Factor-V	3.30e+01
c	39	15	11.7	4698	1	N90844	Tobacco basic beta-1,	3.30e+01
c	40	15	11.7	4699	1	Q03254	Sequence of clone BSG	3.30e+01
	41	15	11.7	5912	37	T45423	Partial ATM gene tran	3.30e+01
	42	15	11.7	5912	31	T43519	ATM gene clone 7-9.	3.30e+01

	43	15	11.7	6300	18	T03571	Factor-VIII cDNA.	3.30e+01
	44	15	11.7	6763	34	T79728	Beta-lactoglobulin cl	3.30e+01
	45	15	11.7	7053	33	T51357	Factor VIII:C coding	3.30e+01
	46	15	11.7	7056	10	Q66615	Sequence of human fac	3.30e+01
	47	15	11.7	7059	1	N81543	Human Factor VIII-C a	3.30e+01
	48	15	11.7	7440	3	N50375	DNA sequence encoding	3.30e+01
	49	15	11.7	7440	2	N60689	Sequence encoding hum	3.30e+01
	50	15	11.7	8241	1	N81096	cDNA sequence encodin	3.30e+01
	51	15	11.7	8241	1	N81439	Factor VIII cDNA inse	3.30e+01
	52	15	11.7	8341	31	T68772	ATM mutant 4777del830	3.30e+01
	53	15	11.7	8767	31	T68735	ATM mutant 8269del403	3.30e+01
	54	15	11.7	8768	37	T73812	ATM gene variant 8269	3.30e+01
	55	15	11.7	8799	31	T68758	ATM mutant 2467del372	3.30e+01
	56	15	11.7	8816	31	T68750	ATM mutant 5320del355	3.30e+01
	57	15	11.7	8873	31	T68740	ATM mutant 7630del298	3.30e+01
	58	15	11.7	8967	21	T31031	Factor-VIII full-leng	3.30e+01
	59	15	11.7	8970	31	T68771	ATM mutant 1407del201	3.30e+01
	60	15	11.7	8975	3	N50054	Human factor VIII gen	3.30e+01
	61	15	11.7	8996	31	T68783	ATM mutant 4443del175	3.30e+01
	62	15	11.7	8996	31	T68754	ATM mutant 4437del175	3.30e+01
	63	15	11.7	8997	31	T68756	ATM mutant 3403del174	3.30e+01
	64	15	11.7	8997	31	T68774	ATM mutant 3403del174	3.30e+01
	65	15	11.7	9006	31	T68775	ATM mutant 4612del165	3.30e+01
	66	15	11.7	9006	31	T68753	ATM mutant 4612del165	3.30e+01
	67	15	11.7	9009	9	Q50185	Human Factor VIII cDN	3.30e+01
	68	15	11.7	9009	33	T61548	Factor VIII:C (Arg168	3.30e+01
	69	15	11.7	9009	40	V12112	Homo sapiens factor V	3.30e+01
	70	15	11.7	9012	31	T68741	ATM mutant 7630del159	3.30e+01
	71	15	11.7	9018	31	T68736	ATM mutant 8269del153	3.30e+01
	72	15	11.7	9022	31	T68745	ATM mutant 6199del149	3.30e+01
	73	15	11.7	9029	31	T68752	ATM mutant 5178del142	3.30e+01
	74	15	11.7	9032	31	T68739	ATM mutant 7789del139	3.30e+01
	75	15	11.7	9044	31	T68755	ATM mutant 4110del127	3.30e+01
	76	15	11.7	9046	31	T68770	ATM mutant 2125del125	3.30e+01
	77	15	11.7	9046	31	T68761	ATM mutant 2125del125	3.30e+01
	78	15	11.7	9066	31	T68744	ATM mutant 6348del105	3.30e+01
	79	15	11.7	9081	31	T68759	ATM mutant 2377del90	3.30e+01
	80	15	11.7	9088	31	T68757	ATM mutant 2839del83	3.30e+01
	81	15	11.7	9160	31	T68749	ATM mutant 5539del11	3.30e+01
	82	15	11.7	9162	31	T68777	ATM mutant 7636del9 c	3.30e+01
	83	15	11.7	9164	31	T68751	ATM mutant 5320del7 c	3.30e+01
	84	15	11.7	9165	31	T68778	ATM mutant 7279del6 c	3.30e+01
	85	15	11.7	9165	31	T68785	ATM mutant 7278del6 c	3.30e+01
	86	15	11.7	9166	31	T68738	ATM mutant 7863del5 c	3.30e+01
	87	15	11.7	9166	31	T68746	ATM mutant 5979del5 c	3.30e+01
	88	15	11.7	9166	31	T68743	ATM mutant 6573del5 c	3.30e+01
	89	15	11.7	9167	31	T68742	ATM mutant 7517del4 c	3.30e+01
	90	15	11.7	9167	31	T68769	ATM mutant 7240del4 c	3.30e+01
	91	15	11.7	9168	31	T68776	ATM mutant 8578del3 c	3.30e+01
	92	15	11.7	9168	31	T68780	ATM mutant 5435del3 c	3.30e+01
	93	15	11.7	9169	31	T68766	ATM mutant 755delGT c	3.30e+01

	94	15	11.7	9169	31	T68760	ATM mutant 2284delCT	3.30e+01
	95	15	11.7	9169	31	T68763	ATM mutant 1563delAG	3.30e+01
	96	15	11.7	9169	37	T73808	ATM gene variant 9001	3.30e+01
	97	15	11.7	9169	37	T73811	ATM gene variant 8283	3.30e+01
	98	15	11.7	9169	31	T68731	ATM mutant 9001delAG	3.30e+01
	99	15	11.7	9169	31	T68734	ATM mutant 8283delTC	3.30e+01
	100	15	11.7	9170	31	T68762	ATM mutant 2113delT c	3.30e+01
	101	15	11.7	9171	37	T73810	ATM gene variant G830	3.30e+01
	102	15	11.7	9171	31	T68768	ATM mutant A538T codi	3.30e+01
	103	15	11.7	9171	31	T68765	ATM mutant C1240T cod	3.30e+01
	104	15	11.7	9171	31	T68764	ATM mutant C1339T cod	3.30e+01
	105	15	11.7	9171	31	T68773	ATM mutant C9139T cod	3.30e+01
	106	15	11.7	9171	37	T43497	ATM open reading fram	3.30e+01
	107	15	11.7	9171	31	T68782	ATM mutant T128C codi	3.30e+01
	108	15	11.7	9171	31	T68733	ATM mutant G8307A cod	3.30e+01
	109	15	11.7	9171	37	T45419	Ataxia-telangiectasia	3.30e+01
	110	15	11.7	9171	31	T68781	ATM mutant A8711G cod	3.30e+01
	111	15	11.7	9171	31	T68737	ATM mutant C8140T cod	3.30e+01
	112	15	11.7	9172	31	T68747	ATM mutant 5712insA c	3.30e+01
	113	15	11.7	9172	31	T68748	ATM mutant 5554insC c	3.30e+01
	114	15	11.7	9172	31	T68732	ATM mutant 8946insA c	3.30e+01
	115	15	11.7	9180	31	T68779	ATM mutant 5319ins9 c	3.30e+01
	116	15	11.7	9385	31	T72683	Cell cycle checkpoint	3.30e+01
c	117	15	11.7	10288	31	T71322	Plasmid pCB5i encodin	3.30e+01
	118	15	11.7	10807	34	T79725	Ovine beta-lactoglobu	3.30e+01
	119	15	11.7	10807	22	T03855	Sheep beta-lactoglobu	3.30e+01
	120	15	11.7	12778	31	T68784	ATM mutant G9170C cod	3.30e+01
c	121	15	11.7	13414	31	T71321	Plasmid pCB50 encodin	3.30e+01
	122	15	11.7	15155	3	N50107	DNA encoding Factor-V	3.30e+01
c	123	15	11.7	22481	24	T11658	PEDF full length sequ	3.30e+01
	124	15	11.7	133894	17	T13635	AcNPV genomic DNA clo	3.30e+01
	125	15	11.7	580073	27	T58040	Mycoplasma genitalium	3.30e+01
c	126	14	10.9	108	19	T21068	Human gene signature	1.15e+02
c	127	14	10.9	146	2	N60861	Fragment of plasmid P	1.15e+02
c	128	14	10.9	213	37	T92822	Flea saliva protein f	1.15e+02
c	129	14	10.9	270	37	T92821	Flea saliva protein f	1.15e+02
c	130	14	10.9	288	2	N60862	Fragment of plasmid P	1.15e+02
c	131	14	10.9	305	24	T34481	Neurite cDNA clone (C	1.15e+02
	132	14	10.9	455	6	Q39736	Expressed Sequence Ta	1.15e+02
	133	14	10.9	455	8	Q59148	Human brain Expressed	1.15e+02
c	134	14	10.9	595	37	T92820	Flea saliva protein f	1.15e+02
c	135	14	10.9	699	40	V24767	H. pylori ORF 01gel02	1.15e+02
	136	14	10.9	762	30	T67811	H. pylori cytoplasmic	1.15e+02
	137	14	10.9	762	30	T77491	H. pylori cytoplasmic	1.15e+02
	138	14	10.9	768	30	T68119	H. pylori cytoplasmic	1.15e+02
c	139	14	10.9	784	2	N60857	Sequence encoding phy	1.15e+02
	140	14	10.9	822	20	Q90765	PGau/B31/K48 outer su	1.15e+02
	141	14	10.9	822	20	Q90763	PGau/PBo outer surfac	1.15e+02
	142	14	10.9	822	20	Q90726	B31/PGau outer surfac	1.15e+02
	143	14	10.9	822	20	Q90711	B. burgdorferi strain	1.15e+02
	144	14	10.9	850	34	T71703	Staphylococcus aureus	1.15e+02

	145	14	10.9	1050	36	T85058	Myceliophthora thermo 1.15e+02
c	146	14	10.9	1158	40	V24720	H. pylori ORF 06ep103 1.15e+02
	147	14	10.9	1301	22	T27045	Telomerase component- 1.15e+02
	148	14	10.9	1356	4	Q25166	VP6 protein of strain 1.15e+02
	149	14	10.9	1356	24	T42059	Human rotavirus VP6 p 1.15e+02
	150	14	10.9	1379	1	N81591	Encodes Tobacco GapB 1.15e+02
	151	14	10.9	1498	14	Q83837	B. burgdorferi OspE-O 1.15e+02
	152	14	10.9	1816	38	T90895	Von Hippel-Lindau dis 1.15e+02
	153	14	10.9	1816	13	Q78896	VHL disease gene g7. 1.15e+02
	154	14	10.9	1886	13	Q75116	Composite sequence de 1.15e+02
	155	14	10.9	2152	39	V15092	Guinea pig sperm surf 1.15e+02
	156	14	10.9	2152	4	Q25301	cDNA encoding guinea 1.15e+02
	157	14	10.9	2152	9	Q54632	Guinea pig PH-20. 1.15e+02
	158	14	10.9	2748	10	Q44474	Lactococcus lactis DN 1.15e+02
c	159	14	10.9	2932	11	Q66636	Human Activin recepto 1.15e+02
c	160	14	10.9	2932	16	T06031	Human ALK-3 cDNA. 1.15e+02
	161	14	10.9	3245	7	Q40730	Mouse eps8. 1.15e+02
c	162	14	10.9	3568	24	T28638	RNase L inhibitor *RL 1.15e+02
c	163	14	10.9	4247	13	Q75303	Human astrovirus sero 1.15e+02
	164	14	10.9	4370	28	T51356	Bacillus subtilis Pst 1.15e+02
c	165	14	10.9	4948	25	T42134	I2C-1 gene encoding F 1.15e+02
	166	14	10.9	5042	38	T89493	Porcine heart-fatty a 1.15e+02
	167	14	10.9	5156	35	T90505	Arabidopsis thaliana 1.15e+02
	168	14	10.9	5285	40	T80199	Phosphatidyl inositol 1.15e+02
	169	14	10.9	5924	13	Q77790	Cobra CVF1 coding seq 1.15e+02
	170	14	10.9	6270	24	T45728	Cyclotella cryptica a 1.15e+02
	171	14	10.9	6469	2	Q11690	Entire sequence of La 1.15e+02
	172	14	10.9	6790	24	T45727	Cyclotella cryptica a 1.15e+02
c	173	14	10.9	6828	13	Q75302	Genomic RNA of human 1.15e+02
c	174	14	10.9	7277	35	T85178	Equine rhinovirus 1 (1.15e+02
	175	14	10.9	7824	16	Q74052	Human interleukin-1-b 1.15e+02
	176	14	10.9	8220	32	T72897	Plasmodium Proj3 gene 1.15e+02
	177	14	10.9	8220	14	Q83529	P. falciparum Proj3 g 1.15e+02
	178	14	10.9	15894	7	Q40480	Attenuated measles vi 1.15e+02
c	179	14	10.9	18177	1	N90490	DNA of human retinobl 1.15e+02
	180	14	10.9	19124	32	T72882	Plasmodium var-7 gene 1.15e+02
c	181	14	10.9	580073	27	T58840	Mycoplasma genitalium 1.15e+02
c	182	13	10.2	17	33	T81491	Human c-myb hammerhea 3.80e+02
c	183	13	10.2	17	33	T81493	Human c-myb hammerhea 3.80e+02
c	184	13	10.2	17	33	T81492	Human c-myb hammerhea 3.80e+02
	185	13	10.2	26	40	V25349	Primer for groE opero 3.80e+02
c	186	13	10.2	53	5	Q34096	Sequence upstream of 3.80e+02
	187	13	10.2	54	24	T28167	Senescence-related ge 3.80e+02
	188	13	10.2	106	15	Q85542	DNA probe 48 detects 3.80e+02
	189	13	10.2	115	1	N70586	Bacillus subtilis pro 3.80e+02
	190	13	10.2	123	20	T23097	Human gene signature 3.80e+02
c	191	13	10.2	128	24	T27155	Human Machado-Joseph 3.80e+02
c	192	13	10.2	129	24	T27153	Human Machado-Joseph 3.80e+02
c	193	13	10.2	145	18	T19713	Human gene signature 3.80e+02
c	194	13	10.2	167	19	T20882	Human gene signature 3.80e+02
	195	13	10.2	170	18	T21033	Human gene signature 3.80e+02

c	196	13	10.2	184	22	T26622	Human gene signature 3.80e+02
c	197	13	10.2	206	18	T19423	Human gene signature 3.80e+02
c	198	13	10.2	225	37	T92829	Flea saliva protein f 3.80e+02
	199	13	10.2	228	20	T23993	Human gene signature 3.80e+02
c	200	13	10.2	243	20	T22224	Human gene signature 3.80e+02
c	201	13	10.2	286	12	Q77205	Human genome fragment 3.80e+02
c	202	13	10.2	286	24	T34480	Neurite cDNA clone (C 3.80e+02
c	203	13	10.2	293	10	Q77581	Human genome fragment 3.80e+02
	204	13	10.2	299	19	T21323	Human gene signature 3.80e+02
c	205	13	10.2	305	8	Q59657	Human brain Expressed 3.80e+02
	206	13	10.2	309	40	V24824	H. pylori ORF hple105 3.80e+02
	207	13	10.2	327	8	Q60406	Human brain Expressed 3.80e+02
c	208	13	10.2	334	21	T24327	Human gene signature 3.80e+02
	209	13	10.2	335	8	Q60303	Human brain Expressed 3.80e+02
	210	13	10.2	336	29	T67340	H. pylori cytoplasmic 3.80e+02
	211	13	10.2	354	29	T67374	H. pylori surface mem 3.80e+02
c	212	13	10.2	375	8	Q59519	Human brain Expressed 3.80e+02
c	213	13	10.2	380	8	Q59917	Human brain Expressed 3.80e+02
	214	13	10.2	396	20	T21861	Human gene signature 3.80e+02
	215	13	10.2	404	18	T19078	Human gene signature 3.80e+02
	216	13	10.2	423	33	T80664	Type II topoisomerase 3.80e+02
c	217	13	10.2	510	35	T70134	Max 14 coding sequenc 3.80e+02
c	218	13	10.2	537	6	Q38720	Max cDNA. 3.80e+02
c	219	13	10.2	537	35	T70133	Max 11 coding sequenc 3.80e+02
c	220	13	10.2	565	37	T92831	Flea saliva protein f 3.80e+02
	221	13	10.2	634	11	Q63690	Partial cDNA encoding 3.80e+02
	222	13	10.2	637	31	T43455	ATM gene exon 15. 3.80e+02
	223	13	10.2	661	28	T36256	CD28 5'-untranslated 3.80e+02
	224	13	10.2	678	30	T67356	H. pylori cytoplasmic 3.80e+02
	225	13	10.2	687	1	Q04000	Sequence complementar 3.80e+02
c	226	13	10.2	768	18	T16240	VR-2332 ORF 2. 3.80e+02
c	227	13	10.2	771	20	T14423	PRRSV VR 2385 ORF-2. 3.80e+02
	228	13	10.2	773	2	N71219	Sequence of the 5' fl 3.80e+02
	229	13	10.2	773	4	N70211	Sequence of 5' flanki 3.80e+02
	230	13	10.2	773	4	N70212	Sequence of 5' flanki 3.80e+02
	231	13	10.2	773	2	N71220	Sequence of the 5' fl 3.80e+02
c	232	13	10.2	789	7	Q44255	Partial bovine VEGF-1 3.80e+02
c	233	13	10.2	789	2	Q10792	Bovine vascular endot 3.80e+02
	234	13	10.2	827	1	N82034	amylolytic enzyme exp 3.80e+02
	235	13	10.2	828	30	T67928	H. pylori surface or 3.80e+02
	236	13	10.2	843	38	V06118	Viral infection gene 3.80e+02
	237	13	10.2	847	22	T35102	Up-regulated senescen 3.80e+02
	238	13	10.2	879	40	V03314	Methanococcus thermol 3.80e+02
	239	13	10.2	880	38	V06095	Viral infection gene 3.80e+02
	240	13	10.2	891	30	T67951	H. pylori cytoplasmic 3.80e+02
	241	13	10.2	897	27	T49878	Porcine trypsinogen c 3.80e+02
c	242	13	10.2	909	19	T11233	Coding region for pro 3.80e+02
	243	13	10.2	930	30	T67570	H. pylori cytoplasmic 3.80e+02
c	244	13	10.2	932	15	Q83174	Lettuce infectious ye 3.80e+02
	245	13	10.2	939	30	T67457	H. pylori cytoplasmic 3.80e+02
	246	13	10.2	990	30	T67878	H. pylori cytoplasmic 3.80e+02

c	247	13	10.2	1005	30	T68260	H. pylori cytoplasmic 3.80e+02
	248	13	10.2	1057	40	V17148	Insecticidal toxin su 3.80e+02
	249	13	10.2	1058	22	T32432	Wasp venom BrhTX-1 su 3.80e+02
	250	13	10.2	1063	24	N95000	Bacillus subtilis lic 3.80e+02
c	251	13	10.2	1078	17	T01287	Rhipicephalus appendi 3.80e+02
	252	13	10.2	1127	23	T37331	Wasp venom 30 kDa ins 3.80e+02
	253	13	10.2	1194	2	N70395	Sequence encoding rat 3.80e+02
c	254	13	10.2	1212	1	Q03739	Arginine deaminase en 3.80e+02
	255	13	10.2	1236	38	V06459	Esex sequence used fo 3.80e+02
	256	13	10.2	1278	30	T63617	Rice tapetum specific 3.80e+02
	257	13	10.2	1311	30	T67502	H. pylori cell envelo 3.80e+02
c	258	13	10.2	1317	11	Q66179	Seven transmembrane r 3.80e+02
	259	13	10.2	1338	30	T68089	H. pylori cytoplasmic 3.80e+02
	260	13	10.2	1344	30	T67909	H. pylori transporter 3.80e+02
	261	13	10.2	1356	30	T67851	H. pylori protein ORF 3.80e+02
	262	13	10.2	1356	15	Q90035	Human SMP30 gene. 3.80e+02
	263	13	10.2	1375	30	T67082	Hop latent virus coat 3.80e+02
	264	13	10.2	1388	14	Q81473	HMG2 promoter I. 3.80e+02
	265	13	10.2	1404	29	T64457	Coat protein gene RNA 3.80e+02
	266	13	10.2	1404	29	T64458	Coat protein gene cDN 3.80e+02
c	267	13	10.2	1408	16	Q91761	Mouse Rad51 gene, enc 3.80e+02
	268	13	10.2	1449	32	T72162	N-acetylglucosamine 1 3.80e+02
c	269	13	10.2	1469	11	Q66476	K. marxianus URA3 gen 3.80e+02
c	270	13	10.2	1494	29	T59907	Yeast transcription r 3.80e+02
c	271	13	10.2	1497	37	V01915	Human GSPT1-TK gene. 3.80e+02
	272	13	10.2	1506	22	T36068	Mullerian inhibiting 3.80e+02
	273	13	10.2	1506	30	T67506	H. pylori cytoplasmic 3.80e+02
	274	13	10.2	1506	8	Q49766	Misr4. 3.80e+02
c	275	13	10.2	1581	24	T31712	Campylobacter fetus g 3.80e+02
c	276	13	10.2	1603	35	T89757	Streptococcus bovis a 3.80e+02
	277	13	10.2	1631	40	V15737	Gamma-sarcoglycane cD 3.80e+02
	278	13	10.2	1637	29	T45357	Human ceramide glucos 3.80e+02
c	279	13	10.2	1731	1	N80161	Human papilloma virus 3.80e+02
c	280	13	10.2	1733	11	Q73300	HIV-type virus MVP518 3.80e+02
	281	13	10.2	1734	30	T68960	Maltose responsive pr 3.80e+02
c	282	13	10.2	1755	11	Q64088	Mouse gene participat 3.80e+02
	283	13	10.2	1769	30	T63616	Rice tapetum specific 3.80e+02
	284	13	10.2	1794	14	Q83534	Sequence encoding ser 3.80e+02
c	285	13	10.2	1799	20	T14422	PRRSV VR 2385 ORF 2, 3.80e+02
c	286	13	10.2	1834	10	Q44495	Noggin clone A3. 3.80e+02
c	287	13	10.2	1867	7	Q40424	Bacillus protease 164 3.80e+02
	288	13	10.2	1883	2	N70154	Hydantoinase gene con 3.80e+02
	289	13	10.2	1929	14	Q82865	Retinoblastoma bindin 3.80e+02
c	290	13	10.2	1944	37	T91276	Streptococcus pneumon 3.80e+02
c	291	13	10.2	1972	2	Q10322	EcoR1-BamH1 fragment 3.80e+02
	292	13	10.2	1974	21	Q94448	Bacterial transferrin 3.80e+02
	293	13	10.2	1974	38	T49506	Transferrin receptor 3.80e+02
c	294	13	10.2	1975	33	T51739	Orange pectin methyle 3.80e+02
c	295	13	10.2	1985	33	T80073	Merozoite apical-end 3.80e+02
c	296	13	10.2	1985	1	Q06329	Sequence encoding mer 3.80e+02
	297	13	10.2	1992	21	T12563	S.aureus topoisomeras 3.80e+02

	298	13	10.2	1996	34	T90830	cDNA encoding glucoam 3.80e+02
c	299	13	10.2	2057	37	V01916	Human GSPT1-TK gene. 3.80e+02
c	300	13	10.2	2083	38	V02472	Bacillus amyloliquefa 3.80e+02
c	301	13	10.2	2084	16	Q95032	Bacillus amyloliquefa 3.80e+02
c	302	13	10.2	2084	15	Q88067	Bacillus amyloliquefa 3.80e+02
	303	13	10.2	2092	1	Q04091	2.1 kb XbaI-BamHI fra 3.80e+02
c	304	13	10.2	2126	7	Q41229	Clone GP14 encoding m 3.80e+02
	305	13	10.2	2144	40	V05531	Rhesus monkey neurope 3.80e+02
	306	13	10.2	2144	40	V05530	Rhesus monkey neurope 3.80e+02
	307	13	10.2	2157	38	V06493	Heat-resistant beta-g 3.80e+02
c	308	13	10.2	2190	7	Q41225	Clone PC14 encoding m 3.80e+02
c	309	13	10.2	2192	13	Q74634	POU domain transcript 3.80e+02
c	310	13	10.2	2205	13	Q74635	POU domain transcript 3.80e+02
c	311	13	10.2	2207	39	V13304	35S CaMV antisense tr 3.80e+02
c	312	13	10.2	2207	23	T33142	Potato trehalase cDNA 3.80e+02
c	313	13	10.2	2207	32	T69900	Potato tuber trehalas 3.80e+02
c	314	13	10.2	2207	36	T94452	Trehalase gene. 3.80e+02
c	315	13	10.2	2251	21	T17141	Tilapia prolactin rec 3.80e+02
c	316	13	10.2	2287	2	Q13802	Non-receptor linked p 3.80e+02
	317	13	10.2	2308	11	Q66638	Human Activin recepto 3.80e+02
c	318	13	10.2	2334	39	V09695	M. hyopneumoniae 2.3k 3.80e+02
c	319	13	10.2	2352	33	T60799	ORFs 2-5 of PRRSV iso 3.80e+02
c	320	13	10.2	2352	33	T60798	ORFs 2-5 of PRRSV iso 3.80e+02
c	321	13	10.2	2352	33	T60797	ORFs 2-5 of PRRSV iso 3.80e+02
c	322	13	10.2	2352	33	T60800	ORFs 2-5 of PRRSV iso 3.80e+02
c	323	13	10.2	2352	33	T60801	ORFs 2-5 of PRRSV iso 3.80e+02
c	324	13	10.2	2352	33	T60795	ORFs 2-5 of PRRSV iso 3.80e+02
	325	13	10.2	2368	4	Q23585	Anther specific promo 3.80e+02
	326	13	10.2	2368	2	Q13372	Potato wun1 gene with 3.80e+02
	327	13	10.2	2368	1	Q04444	Sequence of wun1-85 g 3.80e+02
	328	13	10.2	2385	12	Q70812	Protein kinase (CK1al 3.80e+02
	329	13	10.2	2385	16	Q92947	Human HRR25-like case 3.80e+02
	330	13	10.2	2385	15	Q99884	CKI-alpha-3Hu isoform 3.80e+02
c	331	13	10.2	2403	19	T11081	V.harveyi luxAB genes 3.80e+02
c	332	13	10.2	2404	31	T73282	S. cerevisiae origin 3.80e+02
c	333	13	10.2	2404	15	Q91289	ORC4 gene encoding a 3.80e+02
c	334	13	10.2	2470	36	T72944	Phaffia carotenoid bi 3.80e+02
	335	13	10.2	2487	2	Q13635	Rabbit ATHERO-ELAM. 3.80e+02
c	336	13	10.2	2529	25	T42637	Class A starch branch 3.80e+02
c	337	13	10.2	2531	25	T17267	Class A starch branch 3.80e+02
	338	13	10.2	2535	8	Q48580	NHS-5 gene. 3.80e+02
	339	13	10.2	2535	8	Q48408	Cystathione beta-synt 3.80e+02
c	340	13	10.2	2546	36	T72941	Phaffia carotenoid bi 3.80e+02
	341	13	10.2	2547	30	T67908	H. pylori cytoplasmic 3.80e+02
c	342	13	10.2	2576	25	T42636	Class A starch branch 3.80e+02
c	343	13	10.2	2578	25	T42631	Class A starch branch 3.80e+02
c	344	13	10.2	2628	37	V01060	Human PKR gene. 3.80e+02
c	345	13	10.2	2633	2	Q10351	pLUXSPAFc-2 encoding 3.80e+02
c	346	13	10.2	2642	2	Q10350	pLUXSPAFc-1 encoding 3.80e+02
	347	13	10.2	2712	13	Q85032	Mesquite allergenic p 3.80e+02
c	348	13	10.2	2730	32	T58778	Human eotaxin genomic 3.80e+02

c	349	13	10.2	2740	14	Q85355	Prosaposin cDNA.	3.80e+02
	350	13	10.2	2745	34	T90831	Glucoamylase P gene.	3.80e+02
	351	13	10.2	2749	30	T66652	Housefly Hermes eleme	3.80e+02
c	352	13	10.2	2753	38	V02953	S. diastaticus STA1 g	3.80e+02
c	353	13	10.2	2753	2	N60684	Sequence encoding ext	3.80e+02
c	354	13	10.2	2764	2	N60834	Sequence of plasmid p	3.80e+02
	355	13	10.2	2784	30	T68270	H. pylori cytoplasmic	3.80e+02
c	356	13	10.2	2814	7	Q40201	Sequence of JFBIg10 e	3.80e+02
c	357	13	10.2	2861	19	T13334	Epidermal-specific Bl	3.80e+02
c	358	13	10.2	2949	11	Q66198	H. influenzae b Ca D1	3.80e+02
c	359	13	10.2	2953	11	Q66200	H. influenzae b Minn	3.80e+02
c	360	13	10.2	2975	25	T42635	Class A starch branch	3.80e+02
c	361	13	10.2	2984	11	Q66199	H. influenzae b Eagen	3.80e+02
c	362	13	10.2	2989	11	Q66202	H. influenzae PAK 120	3.80e+02
c	363	13	10.2	3003	25	T42634	Class A starch branch	3.80e+02
	364	13	10.2	3012	3	Q21188	Sequence encoding a 3	3.80e+02
c	365	13	10.2	3033	25	T42630	Class A starch branch	3.80e+02
	366	13	10.2	3034	18	T08715	Calcitonin gene-relat	3.80e+02
	367	13	10.2	3035	19	T16876	cDNA encoding a human	3.80e+02
	368	13	10.2	3054	7	Q42406	Mouse lactoferrin gen	3.80e+02
c	369	13	10.2	3074	31	T69587	Potato starch branchi	3.80e+02
c	370	13	10.2	3077	1	N90068	Ricin D gene	3.80e+02
	371	13	10.2	3113	2	Q13870	Cruciferin A gene.	3.80e+02
	372	13	10.2	3132	20	T17140	Chicken caldesmon gen	3.80e+02
	373	13	10.2	3189	28	T31315	Adenovirus E4 and par	3.80e+02
c	374	13	10.2	3225	40	V20467	Human c-myb oncogene.	3.80e+02
c	375	13	10.2	3231	25	T42632	Class A starch branch	3.80e+02
	376	13	10.2	3249	12	Q71367	E.coli/S.cerevisiae s	3.80e+02
c	377	13	10.2	3252	10	Q54375	Human BDNF vector.	3.80e+02
	378	13	10.2	3315	19	T18535	Receptor tyrosine kin	3.80e+02
	379	13	10.2	3321	11	Q64897	HER4 N-terminal trunc	3.80e+02
c	380	13	10.2	3345	35	T39793	Mouse SH3P12 gene.	3.80e+02
c	381	13	10.2	3358	18	T16238	VR-2332 viral sequenc	3.80e+02
c	382	13	10.2	3360	3	Q15130	DNA contg. Sau3AI met	3.80e+02
	383	13	10.2	3386	35	T78861	Pichia methanolica AU	3.80e+02
	384	13	10.2	3386	30	T60656	AUG1 alcohol utilisat	3.80e+02
	385	13	10.2	3386	40	V20595	Pichia methanolica al	3.80e+02
	386	13	10.2	3386	39	V12855	AUG1 gene.	3.80e+02
	387	13	10.2	3389	16	T01031	Human leucine zipper	3.80e+02
	388	13	10.2	3389	35	T89349	Human leucine-zipper	3.80e+02
	389	13	10.2	3400	12	Q71366	E.coli/S.cerevisiae s	3.80e+02
	390	13	10.2	3542	15	Q94624	Brush-1 cDNA probe.	3.80e+02
c	391	13	10.2	3550	36	T72946	Phaffia carotenoid bi	3.80e+02
	392	13	10.2	3637	1	N91475	Human topoisomerase I	3.80e+02
c	393	13	10.2	3641	1	Q05752	Microspore-specific c	3.80e+02
c	394	13	10.2	3641	28	T48866	Brassica napus micros	3.80e+02
	395	13	10.2	3645	16	Q91432	Variant human DNA top	3.80e+02
c	396	13	10.2	3661	20	T05902	Human IL-2 gene (p2IL	3.80e+02
c	397	13	10.2	3701	13	Q79914	Cytochrome-P450-oxido	3.80e+02
c	398	13	10.2	3706	23	T36392	Lactococcus lactis su	3.80e+02
	399	13	10.2	3798	7	Q42424	ADA2 DNA.	3.80e+02

c	400	13	10.2	3881	19	T13335	Blec4 gene and promot	3.80e+02
	401	13	10.2	4157	30	T60665	Human CD100 antigen c	3.80e+02
c	402	13	10.2	4322	35	T93461	Protein phosphatase 1	3.80e+02
	403	13	10.2	4359	15	Q94625	Brush-1 cDNA.	3.80e+02
	404	13	10.2	4360	25	T44068	Rat von Ebner's gland	3.80e+02
c	405	13	10.2	4487	2	N60943	Sequence encoding glu	3.80e+02
c	406	13	10.2	4698	16	T01176	P815A antigen precurs	3.80e+02
c	407	13	10.2	4698	5	Q32370	P1A genomic DNA.	3.80e+02
c	408	13	10.2	4698	19	T14692	MAGE genomic DNA.	3.80e+02
c	409	13	10.2	4698	13	Q72475	Entire P1A genomic DN	3.80e+02
c	410	13	10.2	4724	24	T35514	Human restrictin cDNA	3.80e+02
c	411	13	10.2	4724	35	T94725	Recombinant human res	3.80e+02
c	412	13	10.2	4724	34	T70489	Human restrictin cDNA	3.80e+02
c	413	13	10.2	4779	25	T58751	Nucleolar/endosomal a	3.80e+02
	414	13	10.2	5099	38	T49503	Transferrin receptor	3.80e+02
	415	13	10.2	5099	21	Q94445	Bacterial transferrin	3.80e+02
	416	13	10.2	5244	28	T48869	IamH expression vecto	3.80e+02
	417	13	10.2	5501	11	Q64896	HER4.	3.80e+02
	418	13	10.2	5501	19	T18532	Receptor tyrosine kin	3.80e+02
c	419	13	10.2	5737	10	Q58034	Interleukin-2 gene.	3.80e+02
c	420	13	10.2	5737	16	Q74053	Human interleukin-2.	3.80e+02
	421	13	10.2	5873	35	T61701	Human prostatic speci	3.80e+02
	422	13	10.2	5904	6	Q37813	Sequence encoding the	3.80e+02
	423	13	10.2	5981	8	Q46174	TSL1 gene.	3.80e+02
	424	13	10.2	6075	4	Q27451	Type A human platelet	3.80e+02
c	425	13	10.2	6075	4	Q27451	Type A human platelet	3.80e+02
	426	13	10.2	6112	28	T48868	IamS expression vecto	3.80e+02
	427	13	10.2	6412	1	Q05989	TR4 cDNA clone relate	3.80e+02
c	428	13	10.2	6412	1	Q05989	TR4 cDNA clone relate	3.80e+02
	429	13	10.2	6464	8	Q48772	Arylamine N-acetyl-tr	3.80e+02
	430	13	10.2	6464	8	Q48770	Arylamine N-acetyl-tr	3.80e+02
	431	13	10.2	6464	8	Q48768	Arylamine N-acetyl-tr	3.80e+02
	432	13	10.2	6575	15	Q84655	Human neuronal calciu	3.80e+02
c	433	13	10.2	6658	34	T79882	Tomato immunity 2 gen	3.80e+02
	434	13	10.2	6725	15	Q87834	Human neuronal calciu	3.80e+02
	435	13	10.2	6785	3	Q20065	Plasmid pAZ112.	3.80e+02
	436	13	10.2	7320	30	T68081	H. pylori cytoplasmic	3.80e+02
c	437	13	10.2	7625	31	T71323	Hybrid activation vec	3.80e+02
	438	13	10.2	7653	9	Q64745	ced-3 (C7020T) gene.	3.80e+02
	439	13	10.2	7653	9	Q64743	ced-3 (C6485T) gene.	3.80e+02
	440	13	10.2	7653	9	Q64739	ced-3 (G6297A) gene.	3.80e+02
	441	13	10.2	7653	9	Q64740	ced-3 (C6322T) gene.	3.80e+02
	442	13	10.2	7653	9	Q54666	ced-3 gene.	3.80e+02
	443	13	10.2	7653	9	Q64744	ced-3 (G6536A) gene.	3.80e+02
	444	13	10.2	7653	9	Q54401	ced-3 gene.	3.80e+02
	445	13	10.2	7653	9	Q64738	ced-3 (C5940T) gene.	3.80e+02
	446	13	10.2	7653	9	Q64741	ced-3 (G6372A) gene.	3.80e+02
	447	13	10.2	7653	9	Q64742	ced-3 (C6434T) gene.	3.80e+02
	448	13	10.2	7653	24	T38196	Nematode Ced-3 gene.	3.80e+02
	449	13	10.2	7653	9	Q64737	ced-3 (G5757A) gene.	3.80e+02
	450	13	10.2	7653	9	Q64736	ced-3 (G2487A) gene.	3.80e+02

c	451	13	10.2	7859	2	Q12154	Shuttle vector pSW6.	3.80e+02
c	452	13	10.2	7859	7	Q44265	pSW6 for expression o	3.80e+02
c	453	13	10.2	7984	4	Q25185	pSW6 expression vecto	3.80e+02
	454	13	10.2	8043	14	Q85924	Human protein tyrosin	3.80e+02
c	455	13	10.2	8083	15	Q90132	Plasmid pBHVtkex-3::p	3.80e+02
c	456	13	10.2	8135	15	Q90134	Plasmid pBHVtkex-1::g	3.80e+02
c	457	13	10.2	8135	15	Q90136	Plasmid pBHVtkex-3::g	3.80e+02
c	458	13	10.2	8149	15	Q90135	Plasmid pBHVtkex-3::B	3.80e+02
c	459	13	10.2	8149	15	Q90133	Plasmid pBHVtkex-1::g	3.80e+02
c	460	13	10.2	8395	1	Q05056	BIV 106 viral genome.	3.80e+02
c	461	13	10.2	8482	1	Q03642	BIV 127 viral genome.	3.80e+02
c	462	13	10.2	8878	22	T03852	Human fibrinogen B-be	3.80e+02
c	463	13	10.2	8920	11	Q62924	Carbamoyl-phosphate-s	3.80e+02
	464	13	10.2	8937	3	Q20602	Nf1 gene.	3.80e+02
c	465	13	10.2	9215	2	Q10203	Sequence of simian im	3.80e+02
	466	13	10.2	9370	13	Q75165	AP-4 transcript (cDNA	3.80e+02
	467	13	10.2	9391	13	Q75164	AP-4 transcript (cDNA	3.80e+02
	468	13	10.2	9714	39	T73135	Porcine transmissible	3.80e+02
c	469	13	10.2	9793	10	Q58974	HIV-type virus MVP518	3.80e+02
	470	13	10.2	10288	31	T71322	Plasmid pCB51 encodin	3.80e+02
	471	13	10.2	10607	6	Q37894	Beta-casein coding se	3.80e+02
	472	13	10.2	10706	29	T46941	Human neurofibromatos	3.80e+02
c	473	13	10.2	11298	34	T86756	Human high affinity I	3.80e+02
c	474	13	10.2	11357	9	Q51024	Human FcERI beta chai	3.80e+02
	475	13	10.2	12923	3	N90338	Sequence of human mus	3.80e+02
	476	13	10.2	13414	31	T71321	Plasmid pCB50 encodin	3.80e+02
	477	13	10.2	13807	32	T62328	Exopolysaccharide ope	3.80e+02
c	478	13	10.2	14237	39	V05384	Human cathepsin K gen	3.80e+02
c	479	13	10.2	14507	32	T73568	Expression augmenting	3.80e+02
	480	13	10.2	14602	31	T73236	Streptococcus thermop	3.80e+02
	481	13	10.2	17327	7	Q44278	Serglycin - proteogly	3.80e+02
	482	13	10.2	29392	40	V15422	Mouse poly Ig recepto	3.80e+02
	483	13	10.2	29544	17	T12460	Ovine adenovirus geno	3.80e+02
	484	13	10.2	30967	23	T32454	Calpain large subunit	3.80e+02
	485	13	10.2	32026	36	T60559	Recombinant adenoviru	3.80e+02
	486	13	10.2	35000	36	T60557	Recombinant cis-actin	3.80e+02
	487	13	10.2	35408	31	T59272	Recombinant adenoviru	3.80e+02
	488	13	10.2	36538	36	T60558	Recombinant trans-act	3.80e+02
	489	13	10.2	40352	40	V02032	MAGE-B cluster DNA se	3.80e+02
c	490	13	10.2	40875	33	T80043	Insert from cosmid 10	3.80e+02
c	491	13	10.2	133894	17	T13635	AcNPV genomic DNA clo	3.80e+02
	492	13	10.2	134525	1	Q04525	Total base sequence o	3.80e+02
c	493	13	10.2	134525	1	Q04525	Total base sequence o	3.80e+02
	494	12	9.4	20	39	V05968	GAPDH upper primer fo	1.19e+03
c	495	12	9.4	20	39	T94234	Primer PI 8 for regul	1.19e+03
c	496	12	9.4	21	39	V13200	Antisense primer for	1.19e+03
	497	12	9.4	26	40	V03039	Sheep beta globin gen	1.19e+03
c	498	12	9.4	26	40	V03041	Sheep beta globin gen	1.19e+03
c	499	12	9.4	26	12	Q87829	GAPDH gene antisense	1.19e+03
c	500	12	9.4	29	12	Q67761	H6 promoter/TNF-alpha	1.19e+03
c	501	12	9.4	29	6	Q35349	PCR primer H65PH.	1.19e+03

c	502	12	9.4	29	5	Q35626	Primer H65PH.	1.19e+03
c	503	12	9.4	29	18	T00898	pEIVC5L amplification	1.19e+03
c	504	12	9.4	30	6	Q35339	PCR primer VV6K5P.	1.19e+03
	505	12	9.4	42	22	T26129	Human gene signature	1.19e+03
	506	12	9.4	50	5	Q33878	Sequence upstream fro	1.19e+03
	507	12	9.4	55	22	T26454	Human gene signature	1.19e+03
	508	12	9.4	63	5	Q33718	Upstream sequence of	1.19e+03
	509	12	9.4	71	19	T19849	Human gene signature	1.19e+03
	510	12	9.4	72	6	Q38300	T cell receptor V reg	1.19e+03
c	511	12	9.4	80	9	Q54845	Transcription termina	1.19e+03
c	512	12	9.4	80	18	T19474	Human gene signature	1.19e+03
	513	12	9.4	82	10	Q54961	groE promoter top str	1.19e+03
	514	12	9.4	96	31	T49278	HVR1 region of E2 gen	1.19e+03
c	515	12	9.4	103	20	T23272	Human gene signature	1.19e+03
	516	12	9.4	109	21	T26577	Human gene signature	1.19e+03
	517	12	9.4	110	34	T87403	EST of human mitochon	1.19e+03
	518	12	9.4	118	19	T20349	Human gene signature	1.19e+03
	519	12	9.4	119	20	T22335	Human gene signature	1.19e+03
	520	12	9.4	124	10	Q76443	Human genome fragment	1.19e+03
	521	12	9.4	125	22	T25161	Human gene signature	1.19e+03
	522	12	9.4	146	39	V12588	VH251/gammal mutant s	1.19e+03
	523	12	9.4	165	19	T20240	Human gene signature	1.19e+03
c	524	12	9.4	172	18	T19296	Human gene signature	1.19e+03
c	525	12	9.4	173	22	T25108	Human gene signature	1.19e+03
	526	12	9.4	174	2	Q10537	Non-A non-B hepatitis	1.19e+03
	527	12	9.4	174	2	Q10534	Non-A non-B hepatitis	1.19e+03
	528	12	9.4	179	19	T09732	Tomato genomic DNA, D	1.19e+03
	529	12	9.4	183	10	Q76962	Human genome fragment	1.19e+03
c	530	12	9.4	188	13	Q72528	Osteoclast-specific/r	1.19e+03
	531	12	9.4	191	11	Q77315	Human genome fragment	1.19e+03
	532	12	9.4	200	22	T23016	Human gene signature	1.19e+03
	533	12	9.4	207	22	T25992	Human gene signature	1.19e+03
	534	12	9.4	208	9	Q50639	Sequence encoding VR1	1.19e+03
	535	12	9.4	218	22	T25015	Human gene signature	1.19e+03
	536	12	9.4	222	14	Q83723	Optimised human beta-	1.19e+03
c	537	12	9.4	231	6	Q64546	Sequence obtained usi	1.19e+03
	538	12	9.4	240	35	T69356	Murine metastatic nuc	1.19e+03
c	539	12	9.4	241	12	Q81804	CZP2(1-65).	1.19e+03
c	540	12	9.4	244	6	Q39920	Expressed Sequence Ta	1.19e+03
c	541	12	9.4	246	8	Q60679	Human brain Expressed	1.19e+03
c	542	12	9.4	249	21	T13710	AcNPV ORF 120, residu	1.19e+03
c	543	12	9.4	249	30	T68241	H. pylori cytoplasmic	1.19e+03
	544	12	9.4	255	20	T23523	Human gene signature	1.19e+03
c	545	12	9.4	260	10	Q77485	Human genome fragment	1.19e+03
	546	12	9.4	264	29	T67371	H. pylori secreted or	1.19e+03
	547	12	9.4	265	22	T22948	Human gene signature	1.19e+03
	548	12	9.4	271	22	T26324	Human gene signature	1.19e+03
	549	12	9.4	276	40	V24752	H. pylori ORF hp5p156	1.19e+03
c	550	12	9.4	290	10	Q76614	Human genome fragment	1.19e+03
	551	12	9.4	294	22	T24898	Human gene signature	1.19e+03
	552	12	9.4	295	22	T25063	Human gene signature	1.19e+03

c	553	12	9.4	295	20	T23733	Human gene signature	1.19e+03
	554	12	9.4	306	20	T22170	Human gene signature	1.19e+03
	555	12	9.4	308	40	V09118	3' nucleotide sequenc	1.19e+03
c	556	12	9.4	309	30	T67833	H. pylori secreted or	1.19e+03
	557	12	9.4	311	8	Q60880	Human brain Expressed	1.19e+03
	558	12	9.4	315	40	V24779	H. pylori ORF 29ge101	1.19e+03
	559	12	9.4	315	30	T68193	H. pylori secreted or	1.19e+03
	560	12	9.4	321	8	Q60159	Human brain Expressed	1.19e+03
c	561	12	9.4	337	8	Q60565	Human brain Expressed	1.19e+03
	562	12	9.4	351	19	T19975	Human gene signature	1.19e+03
	563	12	9.4	354	31	T69212	Lawsonia intracellula	1.19e+03
	564	12	9.4	356	8	Q60654	Human brain Expressed	1.19e+03
	565	12	9.4	360	2	N60182	Sequence of the regul	1.19e+03
	566	12	9.4	386	21	T22532	Human gene signature	1.19e+03
	567	12	9.4	391	22	T30754	Mouse cryptdin 16 cDN	1.19e+03
c	568	12	9.4	391	19	T20524	Human gene signature	1.19e+03
	569	12	9.4	394	8	Q60760	Human brain Expressed	1.19e+03
	570	12	9.4	403	22	T30756	Mouse cryptdin cDNA c	1.19e+03
	571	12	9.4	419	22	T30727	Mouse cryptdin 1 gene	1.19e+03
	572	12	9.4	419	22	T30731	Mouse cryptdin 6 gene	1.19e+03
	573	12	9.4	420	22	T30745	Mouse cryptdin 7 cDNA	1.19e+03
	574	12	9.4	422	22	T30740	Mouse cryptdin 2 cDNA	1.19e+03
	575	12	9.4	422	22	T30741	Mouse cryptdin 3 cDNA	1.19e+03
	576	12	9.4	422	22	T30739	Mouse cryptdin 1 cDNA	1.19e+03
	577	12	9.4	422	22	T30744	Mouse cryptdin 6 cDNA	1.19e+03
c	578	12	9.4	423	33	T80663	Type II topoisomerase	1.19e+03
	579	12	9.4	424	20	T23275	Human gene signature	1.19e+03
c	580	12	9.4	432	19	T20800	Human gene signature	1.19e+03
	581	12	9.4	433	34	T87423	EST from an individua	1.19e+03
	582	12	9.4	433	29	T63705	PAF-AH exon 4 and par	1.19e+03
c	583	12	9.4	467	39	V05735	Internal nucleotide s	1.19e+03
c	584	12	9.4	475	31	T43475	ATM gene exon 40.	1.19e+03
	585	12	9.4	477	19	T20498	Human gene signature	1.19e+03
c	586	12	9.4	478	8	Q60781	Human brain Expressed	1.19e+03
	587	12	9.4	479	18	T19044	Human gene signature	1.19e+03
	588	12	9.4	481	18	T19424	Human gene signature	1.19e+03
c	589	12	9.4	493	19	T19883	Human gene signature	1.19e+03
c	590	12	9.4	498	13	Q80940	Interferon-alpha-61 g	1.19e+03
c	591	12	9.4	513	19	T20793	Human gene signature	1.19e+03
	592	12	9.4	524	34	T68687	Strawberry chalcone s	1.19e+03
	593	12	9.4	547	1	Q03821	Gene encoding bovine	1.19e+03
	594	12	9.4	548	21	T11848	S. aureus strain UZG	1.19e+03
	595	12	9.4	549	30	T67647	H. pylori cytoplasmic	1.19e+03
c	596	12	9.4	549	11	Q66815	AmEPV thymidine-kinas	1.19e+03
	597	12	9.4	576	30	T68039	H. pylori cytoplasmic	1.19e+03
c	598	12	9.4	588	9	Q63918	MnSOD DNA (exon 6).	1.19e+03
	599	12	9.4	594	30	T68183	H. pylori surface or	1.19e+03
c	600	12	9.4	609	29	T48043	Yeast chromosome IX-d	1.19e+03
c	601	12	9.4	668	4	N30164	Sequence of the E. co	1.19e+03
c	602	12	9.4	668	4	N30157	Sequence of E. coli t	1.19e+03
c	603	12	9.4	668	1	Q06496	IFN-alpha 61 gene and	1.19e+03

c	604	12	9.4	668	1	N80049	Sequence of the E.col	1.19e+03
c	605	12	9.4	733	6	Q34485	Cytochrome P450 homol	1.19e+03
	606	12	9.4	743	35	T75597	Canine herpesvirus ge	1.19e+03
c	607	12	9.4	763	4	N20096	Sequence of leukocyte	1.19e+03
	608	12	9.4	771	40	V24719	H. pylori ORF 06ep103	1.19e+03
	609	12	9.4	775	3	N40220	DNA sequence includin	1.19e+03
	610	12	9.4	781	38	V06144	Viral infection gene	1.19e+03
c	611	12	9.4	813	9	Q53193	MnSOD cDNA.	1.19e+03
	612	12	9.4	819	20	Q90712	B. burgdorferi strain	1.19e+03
c	613	12	9.4	820	1	N92866	Segment of lambda 12	1.19e+03
	614	12	9.4	845	34	T66947	Chicken interleukin 1	1.19e+03
	615	12	9.4	857	2	N71048	Drosophila hsp 70 kD	1.19e+03
	616	12	9.4	861	37	T91798	Dihydropyrimidine deh	1.19e+03
c	617	12	9.4	873	21	T13724	AcNPV ORF 144, residu	1.19e+03
c	618	12	9.4	947	12	Q67862	H6/TNF-alpha sequence	1.19e+03
c	619	12	9.4	951	8	Q49779	Prosystemin.	1.19e+03
c	620	12	9.4	962	4	Q23521	Encodes resistance to	1.19e+03
	621	12	9.4	1011	14	Q89203	Sec. alcohol-dehydrog	1.19e+03
c	622	12	9.4	1029	1	Q04290	Sequence encoding bov	1.19e+03
c	623	12	9.4	1029	10	Q55186	MS-PO23 Vector.	1.19e+03
c	624	12	9.4	1032	4	N30004	Sequence of human alp	1.19e+03
c	625	12	9.4	1041	30	T67848	H. pylori cytoplasmic	1.19e+03
c	626	12	9.4	1055	17	T08769	Human DNase precursor	1.19e+03
	627	12	9.4	1071	18	T05171	Prostaglandin D recep	1.19e+03
	628	12	9.4	1080	22	T12904	Sardinian tomato yell	1.19e+03
	629	12	9.4	1080	22	T12905	Sardinian tomato yell	1.19e+03
	630	12	9.4	1086	10	Q62335	Influenza type B HA g	1.19e+03
c	631	12	9.4	1092	30	T67834	H. pylori cell envelo	1.19e+03
c	632	12	9.4	1092	30	T77514	H. pylori cell envelo	1.19e+03
c	633	12	9.4	1095	40	V18818	Human papillomavirus-	1.19e+03
	634	12	9.4	1109	1	Q03852	Sequence encoding mac	1.19e+03
	635	12	9.4	1116	35	T75679	Borrelia burgdorferi	1.19e+03
	636	12	9.4	1126	1	Q06974	Human follicle-stimul	1.19e+03
c	637	12	9.4	1139	35	T79693	BRCA2 cancer suscepti	1.19e+03
	638	12	9.4	1140	1	N81507	Encodes D-amino acid	1.19e+03
c	639	12	9.4	1152	40	V24747	H. pylori ORF hp4p134	1.19e+03
c	640	12	9.4	1161	1	N81711	Sequence of fowlpox v	1.19e+03
c	641	12	9.4	1177	20	T18809	Type I (beta-1) rat b	1.19e+03
c	642	12	9.4	1177	1	N81277	New sequence of rat p	1.19e+03
	643	12	9.4	1200	16	Q95036	Rat hippocampal neuro	1.19e+03
	644	12	9.4	1240	18	T05172	Prostaglandin D recep	1.19e+03
	645	12	9.4	1257	10	Q45289	Sequence of the repli	1.19e+03
c	646	12	9.4	1260	1	Q03250	Sequence encoding bet	1.19e+03
c	647	12	9.4	1265	9	Q53074	Human retinol binding	1.19e+03
c	648	12	9.4	1265	2	Q100027	Encodes a portion of	1.19e+03
	649	12	9.4	1272	38	T84945	Human prostate protei	1.19e+03
	650	12	9.4	1276	3	Q14632	Clone associated with	1.19e+03
	651	12	9.4	1276	21	T34391	Clone S46.	1.19e+03
c	652	12	9.4	1291	38	V03904	BVDV-II gp53 polypept	1.19e+03
	653	12	9.4	1293	16	Q68883	FeLV gp70R DNA.	1.19e+03
	654	12	9.4	1293	2	N70078	Cat recombinant leuke	1.19e+03

	655	12	9.4	1297	3	Q21645	3' coding sequence of	1.19e+03
	656	12	9.4	1327	29	T59852	Human antisecretory f	1.19e+03
	657	12	9.4	1340	34	T85980	Human neuro-D gene 3'	1.19e+03
c	658	12	9.4	1344	24	T43293	Glutaminate dehydroge	1.19e+03
c	659	12	9.4	1357	21	T11340	AFT-1 interacting pro	1.19e+03
c	660	12	9.4	1358	3	Q22162	Plasmid pGIn17 cDNA.	1.19e+03
	661	12	9.4	1361	14	Q81780	Human T-cell lymphoma	1.19e+03
c	662	12	9.4	1361	14	Q83838	B. burgdorferi S1 gen	1.19e+03
c	663	12	9.4	1406	17	Q97295	Marek's disease virus	1.19e+03
c	664	12	9.4	1407	12	Q70644	CD38.	1.19e+03
c	665	12	9.4	1407	1	Q03255	Sequence encoding bet	1.19e+03
c	666	12	9.4	1417	36	T93887	cDNA encoding a novel	1.19e+03
c	667	12	9.4	1423	14	Q85430	TOM41 gene.	1.19e+03
	668	12	9.4	1428	18	T00428	Alpha-1-antichymotryp	1.19e+03
	669	12	9.4	1428	16	T03876	Mature human wild typ	1.19e+03
	670	12	9.4	1428	13	Q75288	Alpha-1-antichymotryp	1.19e+03
c	671	12	9.4	1432	10	Q44494	Bacillus coagulans pr	1.19e+03
	672	12	9.4	1446	40	V17567	Coding sequence for G	1.19e+03
	673	12	9.4	1450	3	Q22545	KDI-I701 cDNA.	1.19e+03
	674	12	9.4	1458	40	V17566	Coding sequence for G	1.19e+03
	675	12	9.4	1473	22	T28390	Human cytochrome P450	1.19e+03
c	676	12	9.4	1479	40	V00640	Canis domesticus cDNA	1.19e+03
	677	12	9.4	1482	21	T26994	Serine threonine kina	1.19e+03
c	678	12	9.4	1482	38	T96644	Human TULP3 cDNA.	1.19e+03
	679	12	9.4	1488	40	V17565	Coding sequence for G	1.19e+03
	680	12	9.4	1500	39	T93438	Collagen binding prot	1.19e+03
	681	12	9.4	1511	39	V14508	AmEPV entomopoxvirus	1.19e+03
	682	12	9.4	1511	11	Q66798	AmEPV thymidine-kinas	1.19e+03
c	683	12	9.4	1511	39	V14508	AmEPV entomopoxvirus	1.19e+03
c	684	12	9.4	1511	5	Q28302	AmEPV tk DNA.	1.19e+03
c	685	12	9.4	1526	34	T80107	Human KF-1 gene clone	1.19e+03
c	686	12	9.4	1528	3	N40025	mRNA encoding gamma a	1.19e+03
	687	12	9.4	1540	8	Q47815	Sequence encoding 38k	1.19e+03
	688	12	9.4	1540	1	N90683	Gene encoding the 38	1.19e+03
c	689	12	9.4	1558	2	Q10896	Encodes Xenopus Bone	1.19e+03
	690	12	9.4	1559	37	T91855	DUB-1 enhancer/promot	1.19e+03
c	691	12	9.4	1586	40	V15213	Human osteogenic fusi	1.19e+03
c	692	12	9.4	1586	18	T02610	CBMP2b2 cDNA encoding	1.19e+03
c	693	12	9.4	1586	9	Q53150	Sequence encoding ost	1.19e+03
	694	12	9.4	1611	2	Q12528	Thymidylate phosphohy	1.19e+03
c	695	12	9.4	1611	4	Q26814	IGFBP6 DNA.	1.19e+03
	696	12	9.4	1618	28	T62019	Human steroidogenesis	1.19e+03
	697	12	9.4	1618	28	T62022	Human steroidogenesis	1.19e+03
	698	12	9.4	1618	28	T62021	Human steroidogenesis	1.19e+03
	699	12	9.4	1618	28	T62018	Human steroidogenesis	1.19e+03
	700	12	9.4	1618	28	T33650	Human steroidogenesis	1.19e+03
	701	12	9.4	1619	28	T62024	Human steroidogenesis	1.19e+03
c	702	12	9.4	1622	24	T37309	Aromatic acyl transfe	1.19e+03
	703	12	9.4	1640	18	T05104	Arabidopsis sp. acety	1.19e+03
c	704	12	9.4	1641	23	T34671	Nuclear inclusion bod	1.19e+03
c	705	12	9.4	1643	12	Q70643	I-19 B-lymphocyte der	1.19e+03

c	706	12	9.4	1650	11	Q65519	Human insulin-like gr	1.19e+03
	707	12	9.4	1656	21	T13716	AcNPV ORF 126, residu	1.19e+03
c	708	12	9.4	1656	33	T85271	HMG-CoA reductase deg	1.19e+03
c	709	12	9.4	1658	34	T91246	Cytochrome P450 encod	1.19e+03
	710	12	9.4	1665	6	Q34486	Cytochrome P450 homol	1.19e+03
	711	12	9.4	1665	8	Q49413	Cytochrome P450 homol	1.19e+03
	712	12	9.4	1680	3	Q14836	Human serotonin recep	1.19e+03
	713	12	9.4	1681	36	T99205	Human D5 dopamine rec	1.19e+03
c	714	12	9.4	1692	1	Q03702	Gene encoding plant e	1.19e+03
	715	12	9.4	1695	14	Q84378	Tomato yellow leaf cu	1.19e+03
	716	12	9.4	1702	20	T09025	Arabidopsis thaliana	1.19e+03
c	717	12	9.4	1761	17	T01407	H6-promoted CHV gC ge	1.19e+03
	718	12	9.4	1763	8	Q47732	Coding sequence for P	1.19e+03
	719	12	9.4	1780	17	Q99275	CM-cellulase tcl2 gen	1.19e+03
	720	12	9.4	1800	1	Q02474	Mycoplasma hyopneumon	1.19e+03
	721	12	9.4	1803	1	Q03855	Hsp70 antigen of Myco	1.19e+03
c	722	12	9.4	1811	22	T28820	Stem cell marker cDNA	1.19e+03
	723	12	9.4	1829	21	T11375	Cytochrome P450 2C8 w	1.19e+03
	724	12	9.4	1837	27	T28540	S. pyogenes detection	1.19e+03
	725	12	9.4	1856	15	Q92310	Mango class II thioes	1.19e+03
	726	12	9.4	1856	40	V15234	Class II acyl-ACP thi	1.19e+03
c	727	12	9.4	1862	5	Q29604	Opossum kidney PTH/PT	1.19e+03
c	728	12	9.4	1862	18	T15945	Opossum kidney PTH/PT	1.19e+03
c	729	12	9.4	1863	18	T15946	Opossum kidney PTH/PT	1.19e+03
	730	12	9.4	1864	2	N31405	Sequence of ANS-1 whi	1.19e+03
	731	12	9.4	1864	15	Q78892	Aspergillus nidulans	1.19e+03
	732	12	9.4	1875	28	T48669	Human EDG-2 receptor	1.19e+03
	733	12	9.4	1878	16	Q68884	FeLV gp90R DNA.	1.19e+03
	734	12	9.4	1878	2	N70079	Cat recombinant leuke	1.19e+03
	735	12	9.4	1882	13	Q80910	Plasmodium falciparum	1.19e+03
	736	12	9.4	1882	13	Q80913	Plasmodium falciparum	1.19e+03
	737	12	9.4	1888	3	N50071	Bacillus amyloliquefa	1.19e+03
c	738	12	9.4	1900	18	T05405	Schizosaccharomyces p	1.19e+03
c	739	12	9.4	1906	10	Q54694	Anther specific genom	1.19e+03
	740	12	9.4	1908	1	N91114	Npr gene encoding neu	1.19e+03
	741	12	9.4	1908	3	N50542	Sequence of the npr[B	1.19e+03
	742	12	9.4	1945	8	Q47816	Sequence encoding 43k	1.19e+03
	743	12	9.4	1949	3	Q22547	KDI-I320 cDNA.	1.19e+03
c	744	12	9.4	1951	21	Q94449	Bacterial transferrin	1.19e+03
c	745	12	9.4	1951	38	T49507	Transferrin receptor	1.19e+03
c	746	12	9.4	1954	18	T07177	Human BMP-4 cDNA clon	1.19e+03
c	747	12	9.4	1954	1	N80634	Human Bone Morphogeni	1.19e+03
c	748	12	9.4	1954	6	Q35244	Encodes human bone mo	1.19e+03
c	749	12	9.4	1954	3	Q14037	Human BMP-2B in lambd	1.19e+03
	750	12	9.4	1959	38	T93069	Streptococcus pneumon	1.19e+03
	751	12	9.4	1965	29	T41615	Adzuki bean endo-xylo	1.19e+03
c	752	12	9.4	1965	29	T41615	Adzuki bean endo-xylo	1.19e+03
	753	12	9.4	1969	1	N93087	Human basic fibroblas	1.19e+03
c	754	12	9.4	1986	33	T78596	Cytidine monophospho-	1.19e+03
c	755	12	9.4	1986	32	T48588	cDNA coding for CMP-N	1.19e+03
	756	12	9.4	1994	14	Q83235	E170 epithelial ligan	1.19e+03

	757	12	9.4	2002 2	Q12537	Pregnancy-specific be	1.19e+03
c	758	12	9.4	2016 30	T66408	Human receptor intera	1.19e+03
c	759	12	9.4	2037 10	Q71298	Probe for human haem-	1.19e+03
c	760	12	9.4	2040 5	Q31246	Chalcone synthase pro	1.19e+03
	761	12	9.4	2061 13	Q81180	B.t. toxin PS71M3 gen	1.19e+03
	762	12	9.4	2061 3	Q14670	Dipteran active toxin	1.19e+03
c	763	12	9.4	2074 35	T39790	Mouse SH3P8 gene.	1.19e+03
	764	12	9.4	2076 2	N70076	Cat recombinant leuke	1.19e+03
c	765	12	9.4	2080 33	T80098	KF-1 gene.	1.19e+03
	766	12	9.4	2089 32	T74887	Mouse neurogenic diff	1.19e+03
	767	12	9.4	2119 10	Q56757	Chitinase 2 gene.	1.19e+03
c	768	12	9.4	2119 10	Q56757	Chitinase 2 gene.	1.19e+03
c	769	12	9.4	2127 38	T96839	Intron 9 of human pRb	1.19e+03
	770	12	9.4	2160 18	T15847	Thermostable enzyme g	1.19e+03
c	771	12	9.4	2163 40	V24818	H. pylori ORF 14ep119	1.19e+03
	772	12	9.4	2169 33	T79783	Potato starch branchi	1.19e+03
c	773	12	9.4	2214 9	Q51268	Endochitinase gene CT	1.19e+03
c	774	12	9.4	2216 12	Q70072	Canine zona pellucida	1.19e+03
	775	12	9.4	2226 2	N92429	Sequence encoding amy	1.19e+03
	776	12	9.4	2243 17	T03744	Rat gax gene.	1.19e+03
	777	12	9.4	2260 23	T27731	Mutant insulin-stimul	1.19e+03
	778	12	9.4	2260 23	T27730	Insulin-stimulated pr	1.19e+03
	779	12	9.4	2307 6	Q34672	P39-alpha and P39-bet	1.19e+03
	780	12	9.4	2307 3	Q22543	KDI-353 DNA.	1.19e+03
	781	12	9.4	2310 3	Q22542	KDI-350 DNA.	1.19e+03
	782	12	9.4	2312 3	Q22541	KDI-348 DNA.	1.19e+03
	783	12	9.4	2335 3	Q22538	KDI-325 DNA.	1.19e+03
	784	12	9.4	2363 30	T58313	Oyster pearl protein,	1.19e+03
	785	12	9.4	2376 12	Q69030	Murine TrkC gene.	1.19e+03
c	786	12	9.4	2379 38	V05640	Human ribonucleotide	1.19e+03
	787	12	9.4	2404 17	Q88163	Tomato ethylene respo	1.19e+03
c	788	12	9.4	2404 17	Q88163	Tomato ethylene respo	1.19e+03
	789	12	9.4	2426 19	T04542	Bacillus stearothermo	1.19e+03
c	790	12	9.4	2430 37	V01880	Human nei-related pro	1.19e+03
	791	12	9.4	2461 38	T96852	Human pRb2/p130 tumou	1.19e+03
	792	12	9.4	2461 19	T04426	M. tuberculosis strai	1.19e+03
c	793	12	9.4	2477 24	T38689	Human CRTAM.	1.19e+03
	794	12	9.4	2484 1	Q05253	Sequence of clone 61C	1.19e+03
	795	12	9.4	2488 37	T95887	DNA for epidermal sur	1.19e+03
c	796	12	9.4	2493 16	Q91097	Human herpesvirus-6 M	1.19e+03
	797	12	9.4	2504 15	Q87446	Tribolium semaphorin	1.19e+03
c	798	12	9.4	2509 2	Q12900	Tobacco intracellular	1.19e+03
c	799	12	9.4	2520 20	T09941	Anther ear1 gene prom	1.19e+03
c	800	12	9.4	2526 14	Q89196	S. dysgalactiae mig g	1.19e+03
c	801	12	9.4	2534 1	N90104	Human non-specific cr	1.19e+03
	802	12	9.4	2549 39	V15744	Rice delta-1-pyrrolin	1.19e+03
	803	12	9.4	2557 2	Q12859	PECAM-1.	1.19e+03
	804	12	9.4	2564 16	T00880	Human cisplatin resis	1.19e+03
	805	12	9.4	2568 1	N90767	Sequence of human tra	1.19e+03
	806	12	9.4	2569 3	Q20290	Sequence encoding hum	1.19e+03
	807	12	9.4	2569 1	Q05126	Human TGF-Beta2-442 p	1.19e+03

	808	12	9.4	2575 31	T67163	Human alpha-N-acetylg	1.19e+03
c	809	12	9.4	2576 39	V16181	Maize optimised DNA s	1.19e+03
c	810	12	9.4	2576 20	T13945	Maize-optimized VIP1A	1.19e+03
c	811	12	9.4	2583 34	T79881	cDNA encoding HIV Rev	1.19e+03
c	812	12	9.4	2600 1	N90750	Protein kinase C beta	1.19e+03
c	813	12	9.4	2626 1	N81159	Human manganese super	1.19e+03
	814	12	9.4	2635 39	V15587	Human B-cell activati	1.19e+03
c	815	12	9.4	2655 39	V16174	Maize optimised DNA s	1.19e+03
c	816	12	9.4	2655 39	V16169	DNA encoding a 100 kD	1.19e+03
c	817	12	9.4	2655 20	T13954	Maize-optimized VIP1A	1.19e+03
c	818	12	9.4	2655 38	T74001	Maize optimised-Bacil	1.19e+03
c	819	12	9.4	2655 20	T13951	Maize-optimized VIP1A	1.19e+03
c	820	12	9.4	2661 27	T36580	Human DNase coding se	1.19e+03
	821	12	9.4	2662 1	N90703	Rhoptry membrane anti	1.19e+03
c	822	12	9.4	2670 15	Q87445	Drosophila semaphorin	1.19e+03
	823	12	9.4	2693 6	Q38220	NANBH virus strain HC	1.19e+03
c	824	12	9.4	2700 17	T15590	Human manganese super	1.19e+03
	825	12	9.4	2717 2	N70621	Pho8 promoter and gen	1.19e+03
	826	12	9.4	2721 3	N50114	DNA sequence encoding	1.19e+03
c	827	12	9.4	2748 1	N80665	Residues 4457 to 7204	1.19e+03
c	828	12	9.4	2763 8	Q48044	Lactobacillus lactis	1.19e+03
	829	12	9.4	2775 32	T72271	Human B cell surface	1.19e+03
	830	12	9.4	2793 21	T16372	Obesity protein codin	1.19e+03
c	831	12	9.4	2812 16	T02327	Human fibroblast acti	1.19e+03
c	832	12	9.4	2813 12	Q73464	Human Egr-2.	1.19e+03
c	833	12	9.4	2815 34	T89286	Human fibroblast acti	1.19e+03
c	834	12	9.4	2817 1	N91816	EGR2 clone.	1.19e+03
	835	12	9.4	2820 11	Q66454	Xenopus thrombospondi	1.19e+03
	836	12	9.4	2911 17	T02795	Human glutamate recep	1.19e+03
	837	12	9.4	2932 11	Q66636	Human Activin recepto	1.19e+03
	838	12	9.4	2932 16	T06031	Human ALK-3 cDNA.	1.19e+03
	839	12	9.4	2946 17	T02796	Human glutamate recep	1.19e+03
	840	12	9.4	2972 20	T03574	Rat protocadherin pc5	1.19e+03
	841	12	9.4	2976 17	T04687	Black widow spider de	1.19e+03
	842	12	9.4	3009 17	Q88160	QETR ethylene respons	1.19e+03
	843	12	9.4	3041 15	Q84693	Plasmid pGla-Mro.	1.19e+03
	844	12	9.4	3055 18	T17937	Pumpkin PP2 genomic c	1.19e+03
c	845	12	9.4	3060 2	Q13822	Human GAP b3 gene.	1.19e+03
	846	12	9.4	3083 36	T94472	Human rchd534 gene.	1.19e+03
	847	12	9.4	3083 23	T36036	rchd534 gene differen	1.19e+03
	848	12	9.4	3188 11	Q66685	Delta(4)-3-ketosteroi	1.19e+03
c	849	12	9.4	3192 17	T08222	X region antisense he	1.19e+03
	850	12	9.4	3214 2	N60714	Sequence encoding adr	1.19e+03
	851	12	9.4	3224 30	T66688	Candida albicans topo	1.19e+03
	852	12	9.4	3249 2	Q11851	Glutamate receptor 5-	1.19e+03
c	853	12	9.4	3268 22	T31275	Human STAT2 cDNA.	1.19e+03
c	854	12	9.4	3275 3	N97129	Partial sequence of m	1.19e+03
c	855	12	9.4	3340 40	V04154	Blood group antigen b	1.19e+03
c	856	12	9.4	3375 10	Q56053	DPM2 mannosyl transfe	1.19e+03
	857	12	9.4	3407 20	T12168	pJG4-5-CDK-BP full le	1.19e+03
c	858	12	9.4	3418 20	T15006	Type II (beta-2) rat	1.19e+03

c	859	12	9.4	3423 34	T80099	Human KF-1 gene.	1.19e+03
c	860	12	9.4	3435 14	Q79746	Mycoplasma pirum adhe	1.19e+03
c	861	12	9.4	3437 3	Q14937	Abelson Related Gene,	1.19e+03
	862	12	9.4	3495 38	T93104	Rat ob receptor isofo	1.19e+03
c	863	12	9.4	3602 2	Q12849	A.niger pyruvate kina	1.19e+03
c	864	12	9.4	3627 16	Q92959	Saccharomyces cerevis	1.19e+03
c	865	12	9.4	3627 12	Q70824	Protein kinase (NUF1)	1.19e+03
	866	12	9.4	3632 39	V13208	Human IL-10 receptor	1.19e+03
	867	12	9.4	3632 11	Q69214	Interleukin-10 recept	1.19e+03
	868	12	9.4	3659 17	T01469	Plasmid pRW731.15 Pvu	1.19e+03
	869	12	9.4	3659 18	T00518	TROVAC DNA fragment.	1.19e+03
	870	12	9.4	3661 20	T05902	Human IL-2 gene (p2IL	1.19e+03
	871	12	9.4	3694 2	N70622	Pho8 promoter and gen	1.19e+03
	872	12	9.4	3706 10	Q58340	Tobacco Ant32 genomic	1.19e+03
	873	12	9.4	3719 14	Q55142	Enterococcus faecalis	1.19e+03
	874	12	9.4	3738 30	T66802	Bacillus thuringiensi	1.19e+03
	875	12	9.4	3751 36	T86087	Transgenic mouse NTA1	1.19e+03
	876	12	9.4	3786 38	T89300	Human ST receptor pro	1.19e+03
	877	12	9.4	3820 1	Q03517	Plasmid pfGH1 encodin	1.19e+03
c	878	12	9.4	3856 1	Q03097	Enod2b genomic clone.	1.19e+03
	879	12	9.4	3867 15	Q51688	Bacillus thuringiensi	1.19e+03
	880	12	9.4	3867 6	Q32160	BT toxin 17b.	1.19e+03
	881	12	9.4	3867 16	Q94054	PS17b acaride-active	1.19e+03
	882	12	9.4	3878 2	N70494	Sequence encoding hum	1.19e+03
c	883	12	9.4	3889 38	T99948	Human N-acetylgalacto	1.19e+03
	884	12	9.4	3891 19	T29244	C. botulinum type A n	1.19e+03
c	885	12	9.4	3900 12	Q66950	Cell death hid cDNA.	1.19e+03
	886	12	9.4	3933 16	T03885	Human mucosal lymphoc	1.19e+03
	887	12	9.4	3934 36	T97233	Trimmed enzyme encodi	1.19e+03
	888	12	9.4	3975 1	N81157	Malaria-specific gene	1.19e+03
c	889	12	9.4	4031 39	V16189	Maize optimised DNA s	1.19e+03
c	890	12	9.4	4031 20	T13947	Maize-optimized VIP2A	1.19e+03
c	891	12	9.4	4091 1	Q04035	Stem cell leukaemia (	1.19e+03
	892	12	9.4	4165 19	T16483	cDNA encoding human E	1.19e+03
c	893	12	9.4	4199 1	Q05330	Stem cell leukaemia (	1.19e+03
	894	12	9.4	4200 35	T78912	Spinocerebellar ataxi	1.19e+03
c	895	12	9.4	4260 32	T47528	H6 promoted cytomegal	1.19e+03
c	896	12	9.4	4343 18	T00522	pMM126 coding sequenc	1.19e+03
	897	12	9.4	4346 7	Q42933	A3 maize ACCase cDNA	1.19e+03
c	898	12	9.4	4355 2	Q10947	Bovine Coronavirus E2	1.19e+03
	899	12	9.4	4410 15	Q93913	Yeast MSH1 gene.	1.19e+03
c	900	12	9.4	4458 35	T68840	Photorhabdus luminesc	1.19e+03
c	901	12	9.4	4483 1	Q03253	Sequence of clone pBS	1.19e+03
	902	12	9.4	4507 14	Q83525	SABP gene.	1.19e+03
c	903	12	9.4	4526 8	Q49780	Prosystemin.	1.19e+03
c	904	12	9.4	4550 3	Q14936	Abelson Related Gene,	1.19e+03
c	905	12	9.4	4555 17	Q94311	Tyrosine phosphatase	1.19e+03
	906	12	9.4	4594 33	T47535	CMV gB transmembrane	1.19e+03
c	907	12	9.4	4603 36	T89957	Pseudomonas aureofaci	1.19e+03
c	908	12	9.4	4604 18	T00523	pC5LCDVHAF1 coding se	1.19e+03
c	909	12	9.4	4684 29	T65017	Aspergillus oryzae al	1.19e+03

	910	12	9.4	4706 17	T10157	Vector plasmid pOnMTG	1.19e+03
	911	12	9.4	4744 17	T10155	Sockeye salmon growth	1.19e+03
	912	12	9.4	4766 10	Q70102	Malarial PfEMP3 epito	1.19e+03
	913	12	9.4	4768 33	T47531	H6 promoted cytomegal	1.19e+03
	914	12	9.4	4783 17	T05113	mSOS1 gene.	1.19e+03
	915	12	9.4	4864 36	V01455	A. thaliana inorganic	1.19e+03
c	916	12	9.4	4866 14	Q82848	P-selectin gene 5' fl	1.19e+03
c	917	12	9.4	4868 11	Q67602	Retinoblastoma protei	1.19e+03
c	918	12	9.4	4972 31	T69203	Lawsonia intracellula	1.19e+03
c	919	12	9.4	4989 15	Q85962	Mouse ST2L gene.	1.19e+03
	920	12	9.4	5008 4	Q27223	MSF precursor.	1.19e+03
	921	12	9.4	5012 24	T43175	Adzuki bean peroxidas	1.19e+03
	922	12	9.4	5042 15	Q84694	Plasmid glucoamylase	1.19e+03
c	923	12	9.4	5095 1	Q05590	Insert sequence encod	1.19e+03
	924	12	9.4	5159 9	Q51559	Keratin K6 gene.	1.19e+03
	925	12	9.4	5234 33	T47563	Cytomegalovirus gB an	1.19e+03
c	926	12	9.4	5408 15	Q95494	Human Cdn-3 DNA.	1.19e+03
	927	12	9.4	5496 12	Q73395	CviJI coding sequence	1.19e+03
	928	12	9.4	5499 5	Q31189	Alpha 6B integrin sub	1.19e+03
	929	12	9.4	5537 15	Q89554	Human cholesterol 7 a	1.19e+03
c	930	12	9.4	5775 40	V20437	Human c-K-ras oncogen	1.19e+03
c	931	12	9.4	5892 35	T79680	BRCA2 cancer suscepti	1.19e+03
	932	12	9.4	5931 15	T04821	EBV gp350/220 cDNA.	1.19e+03
c	933	12	9.4	6000 18	T06308	Arabidopsis pathogen	1.19e+03
	934	12	9.4	6124 1	Q03568	Sequence encoding the	1.19e+03
	935	12	9.4	6128 2	Q13068	CFTR 556 del A.	1.19e+03
	936	12	9.4	6129 2	Q13064	CFTR P574H.	1.19e+03
	937	12	9.4	6129 2	Q13055	CFTR I148T.	1.19e+03
	938	12	9.4	6129 2	Q13067	CFTR 129G -> C.	1.19e+03
	939	12	9.4	6129 2	Q13066	CFTR Y1092X.	1.19e+03
	940	12	9.4	6129 14	Q87795	Cystic fibrosis trans	1.19e+03
	941	12	9.4	6129 2	Q13058	CFTR Q493X.	1.19e+03
	942	12	9.4	6129 15	Q94067	DNA encoding the CFTC	1.19e+03
	943	12	9.4	6129 2	Q13054	CFTR G85E.	1.19e+03
	944	12	9.4	6146 4	Q23660	Cystic Fibrosis trans	1.19e+03
	945	12	9.4	6560 9	Q54529	Genomic region contai	1.19e+03
c	946	12	9.4	6568 39	V15949	Nevoid basal cell car	1.19e+03
	947	12	9.4	6582 38	T99451	S. carnosus nitrate r	1.19e+03
c	948	12	9.4	6675 2	Q10415	Mutant protease gene	1.19e+03
c	949	12	9.4	6675 2	Q10411	Mutant protease gene	1.19e+03
c	950	12	9.4	6675 2	Q10412	Mutant protease gene	1.19e+03
c	951	12	9.4	6675 2	Q10413	Mutant protease gene	1.19e+03
c	952	12	9.4	6675 2	Q10417	Mutant protease gene	1.19e+03
c	953	12	9.4	6711 2	Q10870	Mutant protease gene	1.19e+03
	954	12	9.4	6744 5	Q29471	Extracellular factor	1.19e+03
c	955	12	9.4	6951 1	N91477	CRl protein DNA	1.19e+03
	956	12	9.4	6971 24	T38740	Moraxella outer membr	1.19e+03
c	957	12	9.4	7360 25	T43072	Wheat acetyl-CoA carb	1.19e+03
	958	12	9.4	7462 6	Q35153	Jaagsiekte retrovirus	1.19e+03
	959	12	9.4	7475 3	Q15312	Plasmid pHPB12 encodi	1.19e+03
	960	12	9.4	7876 10	Q58707	E. coli multiple anti	1.19e+03

	961	12	9.4	8135	15	Q90134	Plasmid pBHVtkex-1::g 1.19e+03
	962	12	9.4	8392	9	Q53478	MLL gene 8.3 kb BamHI 1.19e+03
	963	12	9.4	8440	18	T13265	FeLV F6A provirus clo 1.19e+03
c	964	12	9.4	8753	2	Q13285	P.denitrificans EcoRI 1.19e+03
	965	12	9.4	8826	35	T74989	Saccharomyces cerevis 1.19e+03
c	966	12	9.4	9020	17	T07652	pSC11 FIPV N. 1.19e+03
	967	12	9.4	9060	40	V15702	Zucchini ACC synthase 1.19e+03
	968	12	9.4	9060	3	Q15132	Zucchini ACC synthase 1.19e+03
c	969	12	9.4	9096	38	T88015	Human interleukin 16 1.19e+03
	970	12	9.4	9202	2	N70608	VISNA sheep lentiviru 1.19e+03
c	971	12	9.4	9326	4	Q25051	DNA complementary to 1.19e+03
c	972	12	9.4	9402	7	Q41345	Human hepatitis C vir 1.19e+03
	973	12	9.4	9493	29	T59784	Hepatitis GB virus-A 1.19e+03
	974	12	9.4	9621	36	T94548	Mutant YLR087c gene f 1.19e+03
c	975	12	9.4	10564	22	T03854	Human fibrinogen gamm 1.19e+03
	976	12	9.4	10974	2	N60055	Sequence of VB131 clo 1.19e+03
c	977	12	9.4	11283	35	T84841	Human breast and ovar 1.19e+03
c	978	12	9.4	11385	36	T69707	Human breast cancer s 1.19e+03
c	979	12	9.4	11531	9	Q54222	BSSL/CEL Gene. 1.19e+03
	980	12	9.4	11800	11	Q67676	Acetyl CoA carboxylas 1.19e+03
	981	12	9.4	12284	20	T29591	Hog cholera virus. 1.19e+03
	982	12	9.4	12720	10	Q64211	L.lactis branched ami 1.19e+03
c	983	12	9.4	14176	34	T84564	Swinepox virus HindII 1.19e+03
	984	12	9.4	14255	40	V20477	Human ALL-1/AF-4 onco 1.19e+03
	985	12	9.4	14255	13	Q75181	ALL-1 (acute lymphocy 1.19e+03
	986	12	9.4	14255	7	Q43526	ALL-1 gene cDNA. 1.19e+03
	987	12	9.4	18177	1	N90490	DNA of human retinobl 1.19e+03
c	988	12	9.4	19124	32	T72882	Plasmodium var-7 gene 1.19e+03
	989	12	9.4	26698	20	T08126	Mouse syndecan-1 gene 1.19e+03
c	990	12	9.4	26698	20	T08126	Mouse syndecan-1 gene 1.19e+03
	991	12	9.4	26700	40	V15946	Mouse syndecan gene s 1.19e+03
c	992	12	9.4	26700	40	V15946	Mouse syndecan gene s 1.19e+03
c	993	12	9.4	26700	11	Q67902	Syndecan gene. 1.19e+03
c	994	12	9.4	29544	17	T12460	Ovine adenovirus geno 1.19e+03
	995	12	9.4	35100	40	V20441	Human c-fms oncogene. 1.19e+03
	996	12	9.4	37895	24	T34137	Gene for biosynthetic 1.19e+03
	997	12	9.4	37913	10	Q51128	Eicosapentaenoic acid 1.19e+03
c	998	12	9.4	53526	40	T94101	Human PKD1 gene. 1.19e+03
c	999	12	9.4	53577	40	T94108	Human PKD1 locus betw 1.19e+03
c	1000	12	9.4	53577	28	T18551	Human polycystic kidn 1.19e+03

ALIGNMENTS

```
RESULT    1
ID    T20320 standard; cDNA to mRNA; 128 BP.
AC    T20320;
DT    18-JUL-1996  (first entry)
DE    Human gene signature HUMGS01470.
KW    Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW    human; cloning; mapping; non-biased library; diagnosis; detection;
```



```
KW    cell typing; abnormal cell function; ss.
OS    Homo sapiens.
PN    WO9514772-A1.
PD    01-JUN-1995.
PF    11-NOV-1994; JO1916.
PR    12-NOV-1993; JP-355504.
PA    (MATS/) MATSUBARA K.
PA    (OKUB/) OKUBO K.
PI    Matsubara K,  Okubo K;
DR    WPI; 95-206931/27.
PT    Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT    for diagnosis of abnormal cell function, by preparing cDNA that
PT    reflects relative abundance of corresp. mRNA in specific human
PT    tissues
PS    Claim 1; Page 612; 2245pp; Japanese.
CC    A single-stranded DNA (or its complementary strand or the corresp.
CC    double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC    given in T19001-T26837 and which is able to hybridise to part of
CC    human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC    sequences were obtained from 3'-directed cDNA libraries prepared
CC    from various human tissues; synthesis of cDNA was initiated from the
CC    3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC    untranslated sequence is unique to a particular mRNA species, almost
CC    all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC    is constructed so as to reflect accurately the relative abundance of
CC    different mRNAs in the particular tissue from which it was derived.
CC    The appearance frequency of a given GS in a cDNA library can be
CC    determined (esp. using primers and probes derived from the GS
CC    sequences) as a means of diagnosing abnormal cell function or for
CC    recognising different cell types.
SQ    Sequence   128 BP;    41 A;    22 C;    18 G;    47 T;

  Query Match          100.0%;  Score 128;  DB 19;  Length 128;
  Best Local Similarity 100.0%;  Pred. No. 7.55e-86;
  Matches   128;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 gatcagaatttttaaagcagaattttgtcaaaagggtcatttttttgtctacccctttta 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      1 GATCAGAATTTTTAAAGCAGAATTTTGTCAAAAGGGTCATTTTTTTGTCTACCCCTTTTA 60

Db     61 cacttttcagattctcaaagtgtctcatctcaacttttaaaagaataaagaatatcttgc 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     61 CACTTTTCAGATTCTCAAAGTGTCTCATCTCAACTTTTAAAAGAATAAAGAATATCTTGC 120

Db    121 tgggcaaa 128
          ||||||||
Qy    121 TGGGCAAA 128


RESULT    2
ID    T99822 standard; DNA; 1523 BP.
```



```
AC    T99822;
DT    20-MAR-1998  (first entry)
DE    Human steroid 5alpha-reductase pseudogene SRD5AP1 coding sequence.
KW    Steroid 5alpha-reductase; enzyme; testosterone conversion; inhibitor;
KW    human; pseudogene; SRD5AP1; ss.
OS    Homo sapiens.
FH    Key             Location/Qualifiers
FT    CDS             701..1138
FT                    /*tag=  a
PN    US5679521-A.
PD    21-OCT-1997.
PF    01-JUN-1995; 457616.
PR    18-NOV-1991; US-795859.
PR    30-APR-1990; US-517661.
PR    01-JUN-1995; US-457616.
PA    (TEXA ) UNIV TEXAS SYSTEM.
PI    Andersson S, Russell DW;
DR    WPI; 97-525718/48.
DR    P-PSDB; W34450.
PT    Production of recombinant steroid 5alpha-reductase enzyme - by
PT    culturing cell containing DNA encoding the enzyme
PS    Example 8; Columns 67-70; 70pp; English.
CC    This sequence represents the DNA encoding a pseudogene of human steroid
CC    5alpha-reductase, termed SRD5AP1. The encoded enzyme can be produced by
CC    the method of the invention. The method is for producing a steroid
CC    5alpha-reductase, and comprises preparing a recombinant host cell
CC    containing a DNA segment encoding a steroid 5alpha-reductase and
CC    culturing the cell under conditions such that the steroid
CC    5alpha-reductase is produced by the cell. The steroid 5alpha-reductase
CC    produced by the method is used for identifying substances that affect the
CC    enzymatic activity of steroid 5alpha-reductase. Substances identified as
CC    inhibiting steroid 5alpha-reductase activity can be used for inhibiting
CC    the conversion of testosterone.
SQ    Sequence   1523 BP;    443 A;    360 C;    356 G;    364 T;

  Query Match           14.8%;  Score 19;  DB 37;  Length 1523;
  Best Local Similarity 100.0%;  Pred. No. 1.52e-01;
  Matches    19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    567 aaaagaataaagaatatct 585
          |||||||||||||||||||
Qy     99 AAAAGAATAAAGAATATCT 117


RESULT    3
ID    Q92782 standard; DNA; 2206 BP.
AC    Q92782;
DT    09-DEC-1995  (first entry)
DE    Human thymopoietin gene 5' end.
KW    Thymopoietin; diagnostic; therapeutic; gene therapy; ss.
OS    Homo sapiens.
```



```
FH    Key             Location/Qualifiers
FT    misc_feature    2124..2143
FT                    /*tag=  a
FT                    /label= primer
FT                    /note= "used in primer extension analysis"
FT    misc_feature    1888
FT                    /*tag=  b
FT                    /label= transcription_initiation_site
FT                    /note= "found by primer extension analysis"
FT    misc_feature    1901
FT                    /*tag=  c
FT                    /label= transcription_initiation_site
FT                    /note= "found by primer extension analysis"
FT    misc_feature    593..100
FT                    /*tag=  d
FT                    /note= "transcription factor OTF binding sequence"
FT    misc_feature    1434..1445
FT                    /*tag=  e
FT                    /note= "Sp1 H4TF-1 binding sequence"
FT    misc_feature    1668..1676
FT                    /*tag=  f
FT                    /note= "H4TF-1 binding sequence"
FT    misc_feature    1741..1750
FT                    /*tag=  g
FT                    /note= "ISRE-like binding sequence"
FT    misc_feature    1760..1766
FT                    /*tag=  h
FT                    /note= "Sp1-like binding sequence"
FT    misc_feature    1792..1797
FT                    /*tag=  i
FT                    /note= "TATA-like binding sequence"
FT    misc_feature    1889..1895
FT                    /*tag=  j
FT                    /note= "Sp1 CTF binding sequence"
PN    WO9517205-A1.
PD    29-JUN-1995.
PF    13-DEC-1994; U14356.
PR    21-DEC-1993; US-171382.
PA    (IMMU-) IMMUNOBIOLOGY RES INST INC.
PI    Culler MD,  Goldstein G,  Harris CA,  Setcavage DR;
PI    Shenbagamurthi P, Siekierka JJ,  Talle MA;
DR    WPI; 95-240474/31.
PT    Polynucleotide(s) encoding human thymopoietin proteins - used to
PT    develop prods. for diagnosis and therapy involving immune or nervous
PT    system conditions.
PS    Disclosure; Fig 12a-12c; 85pp; English.
CC    The sequence represents the 5' region of the human thymopoietin
CC    gene.  DNA encoding human thymopoietin, isolated from a cDNA
CC    library prepared from human thymus RNA, may be expressed in a
CC    host cell e.g. bacterium (preferably Escherichia coli), fungus,
CC    insect or mammalian cell for production of  recombinant thymopoietin.
```

```
CC   Thymopoietin has a regulatory effect on the mammalian immune system,
CC   and can be used for treating/modulating an immune or nervous system
CC   condition, and for the treatment of chronic infection, autoimmune
CC   disorders and certain affective psychiatric or neurological
CC   disorders.  The protein and the DNA can be used as diagnostics;
CC   the DNA can be used in gene therapy.
SQ   Sequence   2206 BP;    607 A;    481 C;    554 G;    564 T;

  Query Match           14.1%;  Score 18;  DB 15;  Length 2206;
  Best Local Similarity 100.0%;  Pred. No. 6.13e-01;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    250 agaatctgaaaagtgtaa 267
          ||||||||||||||||||
Cp     75 AGAATCTGAAAAGTGTAA 58


RESULT    4
ID   Q92776 standard; DNA; 2509 BP.
AC   Q92776;
DT   09-DEC-1995  (first entry)
DE   Human thymopoietin gene continuous fragment.
KW   Thymopoietin; diagnostic; therapeutic; gene therapy; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   exon            1888-1901..2479
FT                   /*tag=  a
FT   5'utr           1888-1901..2200
FT                   /*tag=  b
FT   intron          2480..2509
FT                   /*tag=  c
FT                   /note= "intron 1"
PN   WO9517205-A1.
PD   29-JUN-1995.
PF   13-DEC-1994; U14356.
PR   21-DEC-1993; US-171382.
PA   (IMMU-) IMMUNOBIOLOGY RES INST INC.
PI   Culler MD,  Goldstein G,  Harris CA,  Setcavage DR;
PI   Shenbagamurthi P, Siekierka JJ,  Talle MA;
DR   WPI; 95-240474/31.
PT   Polynucleotide(s) encoding human thymopoietin proteins - used to
PT   develop prods. for diagnosis and therapy involving immune or nervous
PT   system conditions.
PS   Disclosure; Fig 6a-6c; 85pp; English.
CC   The DNA is isolated from a cDNA library prepared from human thymus
CC   RNA, and may be expressed in a host cell e.g. bacterium (preferably
CC   Escherichia coli), fungus, insect or mammalian cell for production of
CC   recombinant thymopoietin.  Thymopoietin has a regulatory effect on
CC   the mammalian immune system, and can be used for treating/
CC   modulating an immune or nervous system condition, and for the
CC   treatment of chronic infection, autoimmune disorders and certain
```



```
CC   affective psychiatric or neurological disorders.  The protein and
CC   the DNA can be used as diagnostics; the DNA can be used in gene
CC   therapy.
SQ   Sequence   2509 BP;    666 A;    588 C;    655 G;    600 T;

  Query Match           14.1%;  Score 18;  DB 15;  Length 2509;
  Best Local Similarity 100.0%;  Pred. No. 6.13e-01;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    250 agaatctgaaaagtgtaa 267
          ||||||||||||||||||
Cp     75 AGAATCTGAAAAGTGTAA 58


RESULT    5
ID   Q87016 standard; cDNA; 1289 BP.
AC   Q87016;
DT   24-NOV-1995  (first entry)
DE   Clone B2 encodes functional equivalent of XET activity protein.
KW   xyloglucan endo-transglycosylase activity; tomato plant;
KW   genetically engineered plant; functional equivlent; ss.
OS   Lycopersicon esculentum.
FH   Key             Location/Qualifiers
FT   cds             232..1095
FT                   /*tag=  a
FT                   /product= immature_XET
FT                   /note= "xyloglucan endo-transglycosylase functional
FT                   equivalent"
FT   signal_peptide  232..300
FT                   /*tag=  b
FT   mat_peptide     301..1092
FT                   /*tag=  c
FT                   /note= "mature protein"
PN   WO9513384-A.
PD   18-MAY-1995.
PF   10-NOV-1994; G02467.
PR   10-NOV-1993; GB-023225.
PA   (UNIL ) UNILEVER NV.
PA   (UNIL ) UNILEVER PLC.
PI   Arrowsmith DA,  De SILVA J,  Hellyer SA,  Whiteman S;
DR   WPI; 95-194102/25.
DR   P-PSDB; R73013.
PT   A polypeptide having xyloglucan endo-transglycosylase activity  - is
PT   encoded by a nucleic acid isolated from a tomato plant, the nucleic
PT   acid being used in the prodn. of a genetically engineered plant
PS   Claim 13; Page 29-31; 61pp; English.
CC   The cDNA encodes a functional equivalent of the protein shown in
CC   R73008. R73008 has 39 % identity in 270 amino acids with nasturtium
CC   XET (Xyloglucan endo-transglycosylase) and 59 % identity in an overlap
CC   of 186 amino acids of meri-5. Clone B1 may be part of a multi-gene
CC   family in tomato.  XET activity has been implicated in cell expansion,
```



```
CC   growth of plants and fruit ripening. The nucleotide sequence may be
CC   used to produce a genetically engineered plant with a view to altering
CC   characteristics such as size, growth rate, texture and ripening
CC   (claimed).
SQ   Sequence   1289 BP;    371 A;    240 C;    271 G;    407 T;

  Query Match           13.3%;  Score 17;  DB 16;  Length 1289;
  Best Local Similarity 100.0%;  Pred. No. 2.40e+00;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1211 atattctttattctttt 1227
          |||||||||||||||||
Cp    115 ATATTCTTTATTCTTTT 99


RESULT    6
ID   T85307 standard; DNA; 4937 BP.
AC   T85307;
DT   14-APR-1998  (first entry)
DE   Arabidopsis thaliana cytochrome P450-type hydroxylase DNA.
KW   Cytochrome P450-type hydroxylase; identification; brassinosteroid;
KW   brassinosteroid inhibitor; modified plant; recombinant production;
KW   teasterone; ds.
OS   Arabidopsis thaliana.
FH   Key             Location/Qualifiers
FT   CDS             968..4773
FT                   /*tag=  a
FT                   /product= cytochrome_P450-type_hydroxylase
FT   intron          1484..1679
FT                   /*tag=  b
FT                   /number= 1
FT   intron          1830..1916
FT                   /*tag=  c
FT                   /number= 2
FT   intron          2166..3902
FT                   /*tag=  d
FT                   /number= 3
FT   intron          3990..4083
FT                   /*tag=  e
FT                   /number= 4
FT   intron          4163..4247
FT                   /*tag=  f
FT                   /number= 5
FT   intron          4355..4445
FT                   /*tag=  g
FT                   /number= 6
FT   intron          4577..4673
FT                   /*tag=  h
FT                   /number= 7
PN   WO9735986-A1.
PD   02-OCT-1997.
```



```
PF   27-MAR-1997; E01586.
PR   27-MAR-1996; US-622166.
PA   (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PI   Altmann T, Koncz C, Mathur J, Szekeres MA;
DR   WPI; 97-489649/45.
DR   P-PSDB; W27153.
PT   New isolated plant cytochrome P450-type hydroxylase gene - used to
PT   identify substances acting as brassino-steroid(s) or brassinosteroid
PT   inhibitors for the production of modified plants
PS   Claim 1; Pages 48-53; 77pp; English.
CC   The present sequence encodes Arabidopsis thaliana cytochrome
CC   P450-type hydroxylase. The hydroxylase can be used to identify
CC   brassinosteroids or brassinosteroid inhibitors, useful to produce
CC   plants with modified physiological and/or phenotypic
CC   characteristics. The modified plants may show, e.g. stimulation of
CC   growth, increased cell elongation, increased wood production,
CC   accelerated seed germination at low temperatures, an increase in
CC   dry weight, repressed anthocyanin production during growth in light
CC   and/or inhibited de-etiolation which is induced, e.g. by cytokinin,
CC   in the dark or an increase in stress tolerance. The hydroxylase or
CC   its coding sequence can also be used for the recombinant production
CC   of compounds, e.g. teasterone.
SQ   Sequence   4937 BP;    1526 A;    864 C;    870 G;    1677 T;

  Query Match           13.3%;  Score 17;  DB 37;  Length 4937;
  Best Local Similarity 100.0%;  Pred. No. 2.40e+00;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   3323 ctgaaaagtgtaaaagg 3339
          |||||||||||||||||
Cp     70 CTGAAAAGTGTAAAAGG 54


RESULT    7
ID   T03611 standard; cDNA; 125 BP.
AC   T03611;
DT   29-AUG-1996  (first entry)
DE   Protocadherin clone HUMAN-410 coding sequence.
KW   Protocadherin; pc3; pc4; pc5; human; rat; cadherin; cell adhesion; mouse;
KW   catenin; therapy; clone; ss.
OS   Homo sapiens.
PN   WO9600289-A1.
PD   04-JAN-1996.
PF   26-JUN-1995; UG8071.
PR   27-JUN-1994; US-268161.
PA   (DOHE-) DOHENY EYE INST.
PI   Suzuki S;
DR   WPI; 96-068873/07.
DR   P-PSDB; R87136.
PT   Polynucleotide(s) encoding human protocadherins pc3 and pc4 and rat
PT   pc5 - involved in cell-cell adhesion and regulation activities
```

```
PS   Example 2; Page 63; 146pp; English.
CC   T03600-T03612 represent partial cDNA clones of the human protocadherin
CC   sequence.  These sequences were isolated after screening a human brain
CC   cDNA preparation with the primers shown in T03575 and T03576.  The
CC   primers were constructed from portions of the amino acid sequences of the
CC   third and fourth extracellular domains of published cadherin sequences.
CC   The full length cDNA sequences encoding human protocadherins pc3 and pc4
CC   are represented by T03572 and T03573, respectively.  The cytoplasmic
CC   domain of cadherin interacts with the cytoskeleton through catenins and
CC   other cytoskeleton associated proteins.  The cytoplasmic domain is not
CC   present in all cadherins, but in those which possess it, it is essential
CC   for the cadherins adhesive function.  The cadherins which do not possess
CC   a cytoplasmic domain appear to function via a different method from those
CC   with a cytoplasmic domain.  The proteins encoded by these sequences are
CC   involved in cell-cell adhesion.  The proteins may have regulatory
CC   functions in the cell, as well as the cell-cell adhesive properties.
CC   Antibodies produced against the encoded sequences are useful for
CC   modulating the binding activity of these protocadherins, and can be used
CC   therapeutically.
SQ   Sequence   125 BP;    37 A;    23 C;    32 G;    33 T;

  Query Match           12.5%;  Score 16;  DB 20;  Length 125;
  Best Local Similarity 100.0%;  Pred. No. 9.07e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     72 tgtcaaaagggtcatt 87
          ||||||||||||||||
Qy     26 TGTCAAAAGGGTCATT 41


RESULT    8
ID   Q68987 standard; cDNA; 125 BP.
AC   Q68987;
DT   17-APR-1995  (first entry)
DE   Human-410 cadherin-related molecule.
KW   Cadherin; cell adhesion molecule; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             1..123
FT                   /*tag=  a
PN   WO9414960-A.
PD   07-JUL-1994.
PF   23-DEC-1993; U12588.
PR   29-DEC-1992; US-998003.
PA   (DOHE-) DOHENY EYE INST.
PI   Suzuki S;
DR   WPI; 94-293849/36.
DR   P-PSDB; R49143.
PT   Polynucleotide sequences encoding new proto:cadherins - useful
PT   for modulating natural binding and regulating activities.
PS   Example; Page 59; 114pp; English.
```



```
CC   Two regions of conserved AA sequence, one from the middle of the
CC   third cadherin extracellular subdomain (EC-3) and the other from the
CC   C-terminus of the fourth extracellular subdomain (EC-4) were
CC   identified. The corresp. degenerate oligos (Q68949, Q68950) were
CC   designed for use as PCR primers. PCR was carried out on a rat brain
CC   cDNA prepn. Two major bands of about 450 bps and 130 bps were found.
CC   The 450 bp band corresponded to the expected length between the two
CC   primer sites, but the 130 bp band could not be predicted from any
CC   of the previously identified cadherin sequences. The 450 bp and 130
CC   bp bands were extracted and sequenced. Nineteen novel partial cDNA
CC   clones were isolated. The DNA and deduced AA sequences of the
CC   clones (including sequences corresp. to the PCR primers) are given
CC   in Q68951-Q68969 and R58860-R58878. Various cDNA fragments
CC   structurally similar to the rat cDNAs were isolated from human,
CC   mouse and Xenopus brain cDNA prepns. and from Drosophila and
CC   C. elegans whole body cDNA prepns. by PCR using the above primers.
CC   The DNA and deduced AA sequences of the resulting PCR fragments
CC   (including sequences corresp. to the PCR primers) are given in
CC   Q68971, Q68972-Q68994 and R58882-R58905 and R49143. Comparison
CC   of the deduced AA sequences indicates a similarity, in particular,
CC   there are three sets of clones that appear to be cross-species
CC   homologues: RAT-218, MOUSE-322 and HUMAN-43; RAT-314, MOUSE-321 and
CC   HUMAN-11; and MOUSE-326 and HUMAN-42.
SQ   Sequence   125 BP;    37 A;    23 C;    32 G;    33 T;

  Query Match           12.5%;  Score 16;  DB 12;  Length 125;
  Best Local Similarity 100.0%;  Pred. No. 9.07e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     72 tgtcaaaagggtcatt 87
          ||||||||||||||||
Qy     26 TGTCAAAAGGGTCATT 41


RESULT    9
ID   Q48470 standard; DNA; 645 BP.
AC   Q48470;
DT   31-MAR-1994  (first entry)
DE   Bacillus subtilis PYRase pcp gene.
KW   Pyrrolidone Carboxylyl Peptidase; pcp gene; bacteria; detection; ss.
OS   Bacillus subtilis.
FH   Key             Location/Qualifiers
FT   mat_peptide     1..645
FT                   /*tag=  a
FT                   /product= PYRase
PN   FR2685334-A.
PD   25-JUN-1993.
PF   23-DEC-1991; 016059.
PR   23-DEC-1991; FR-016059.
PA   (INMR ) BIO MERIEUX.
PI   Awade A,  Cleuziat P,  Gayral J,  Robert-Baudouy J;
```



```
DR   WPI; 93-297122/38.
DR   P-PSDB; R41759.
PT   Pyrrolidone carboxylyl peptidase polypeptide(s) and their DNA -
PT   for detecting Streptococcus pyogenes and Bacillus subtilis
PS   Claim 10; Fig 5; 40pp; French.
CC   This sequence represents the coding region of the pcp gene from
CC   Bacillus subtilis. A nucleic acid probe containing a sequence
CC   from Q48470 is claimed and can be used as either a capture probe or
CC   a detection probe for B.subtilis.
SQ   Sequence   645 BP;    184 A;    166 C;    158 G;    137 T;

  Query Match           12.5%;  Score 16;  DB 8;  Length 645;
  Best Local Similarity 100.0%;  Pred. No. 9.07e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     30 tgacccttttgacaaa 45
          ||||||||||||||||
Cp     39 TGACCCTTTTGACAAA 24


RESULT   10
ID   Q50811 standard; DNA; 645 BP.
AC   Q50811;
DT   09-MAY-1994  (first entry)
DE   Bacillus subtilis pcp gene.
KW   Pyrrolidone carboxylyl peptidase; PYRase; pyrrolidonyl peptidase;
KW   exopeptidase; group A beta-haemolytic streptococci; detection; ss.
OS   Bacillus subtilis.
FH   Key             Location/Qualifiers
FT   cds             1..645
FT                   /*tag=  a
FT                   /product= PYRase
FT                   /note= "does not include termination codon"
PN   FR2686604-A.
PD   30-JUL-1993.
PF   23-DEC-1992; 015658.
PR   23-DEC-1991; FR-016059.
PA   (INMR ) BIO MERIEUX.
PI   Awade A,  Cleuziat P,  Gayral J-P,  Robert-Baudouy J;
DR   WPI; 93-338557/43.
DR   P-PSDB; R43315.
PT   New peptide(s) from bacterial pyrrolidone carboxyl:yl peptidase -
PT   for prodin. of antibodies, DNA transformed cells etc., useful e.g.
PT   for detecting specific pathogens
PS   Claim 8; Fig 5; 37pp; French.
CC   The pcp gene coding for the PYRase of Bacillus subtilis was
CC   isolated from a B.subtilis gene library. The amino acid sequence
CC   deduced from the coding region has 47.5% identity to the PYRase
CC   from S.pyogenes and more than 66% homology to it. Antibodies to
CC   the regions which are highly conserved between the PYRases of
CC   different bacteria can be used to detect any PCP-producing species.
```



```
SQ   Sequence   645 BP;    184 A;    166 C;    158 G;    137 T;

  Query Match           12.5%;  Score 16;  DB 9;  Length 645;
  Best Local Similarity 100.0%;  Pred. No. 9.07e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     30 tgacccttttgacaaa 45
          ||||||||||||||||
Cp     39 TGACCCTTTTGACAAA 24


RESULT   11
ID   V24778 standard; DNA; 1146 BP.
AC   V24778;
DT   17-JUN-1998  (first entry)
DE   H. pylori ORF 04gp11213orf5.
KW   Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW   identification; binding compound; bacteria; life cycle; activator;
KW   inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis; ds.
OS   Helicobacter pylori.
FH   Key             Location/Qualifiers
FT   CDS             1..1146
FT                   /*tag=  a
FT                   /note= "no stop codon given"
PN   WO9737044-A1.
PD   09-OCT-1997.
PF   27-MAR-1997; U05223.
PR   06-DEC-1996; US-761318.
PR   29-MAR-1996; US-625811.
PR   02-APR-1996; US-758731.
PR   25-OCT-1996; US-736905.
PR   28-OCT-1996; US-738859.
PA   (ASTR ) ASTRA AB.
PI   Alm RA,  Smith D;
DR   WPI; 97-503122/46.
DR   P-PSDB; W55269.
PT   Helicobacter pylori nucleic acid sequences and encoded
PT   polypeptide(s) - useful in vaccines to treat or prevent H. pylori
PT   infection and for diagnosis of H. pylori infection
PS   Claims 5,6; Page 239; 1145pp; English.
CC   This sequence encodes a H. pylori protein of unspecified function.
CC   The protein may be used in a vaccine to prevent or treat H. pylori
CC   infection or to identify H. pylori polypeptide binding compounds,
CC   useful as potential H. pylori life cycle activators or inhibitors. The
CC   DNA and probes derived from it may be used for the identification of
CC   H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
CC   acid sequences complementary to the DNA act as antisense sequences and
CC   can be used to prevent the translation of H. pylori mRNA. Antibodies
CC   against the protein can be used in immunoassays to evaluate the abundance
CC   and distribution of H. pylori-specific antigens. The genomic sequence of
CC   H. pylori (ATCC 55679) was determined from overlapping contigs generated
```

```
CC   by mechanically shearing the bacterial DNA. The sequences were analysed
CC   for ORF of at least 180 nucleotides, and the predicted coding regions
CC   defined by computer evaluation. To identify likely H. pylori antigens for
CC   vaccine development, the amino acid sequences predicted from various ORF
CC   were analysed for significant homology to other known or exported
CC   membrane proteins. Having identified and determined the sequences of
CC   interest, particular regions can be isolated from H. pylori by PCR
CC   amplification for recombinant polypeptide production, e.g. in E. coli
CC   hosts.
SQ   Sequence   1146 BP;    442 A;    247 C;    204 G;    253 T;

  Query Match          12.5%;  Score 16;  DB 40;  Length 1146;
  Best Local Similarity 100.0%;  Pred. No. 9.07e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    679 gaaaagtgtaaaaggg 694
          ||||||||||||||||
Cp     68 GAAAAGTGTAAAAGGG 53


RESULT   12
ID   Q94401 standard; cDNA to mRNA; 168 BP.
AC   Q94401;
DT   23-JAN-1996  (first entry)
DE   Tyrosine kinase-like protein coding sequence.
KW   Tyrosine kinase-like protein; human; osteoblast cell line MG63;
KW   detection; ds.
OS   Homo sapiens.
PN   J07115973-A.
PD   09-MAY-1995.
PF   25-OCT-1993; 266468.
PR   25-OCT-1993; JP-266468.
PA   (ASAH ) ASAHI KASEI KOGYO KK.
DR   WPI; 95-202834/27.
DR   P-PSDB; R77777.
PT   New tyrosine kinase-like polypeptide and its DNA - isolated from
PT   human osteoblast MG63 cells and useful for specific detection of an
PT   osteoblast
PS   Claim 1; Page 5; 5pp; Japanese.
CC   This sequence encodes a tyrosine kinase-like protein.  This sequence
CC   was isolated from human osteoblast cell line MG63 after culture in
CC   alpha-MEM.  The tyrosine kinase-like protein may be used in the
CC   specific detection of osteoblasts.
SQ   Sequence   168 BP;    45 A;    35 C;    48 G;    40 T;

  Query Match          11.7%;  Score 15;  DB 15;  Length 168;
  Best Local Similarity 100.0%;  Pred. No. 3.30e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    138 tatcttgctgggcaa 152
          ||||||||||||||| 
```



```
Qy    113 TATCTTGCTGGGCAA 127


RESULT   13
ID   T22939 standard; cDNA to mRNA; 276 BP.
AC   T22939;
DT   24-OCT-1996  (first entry)
DE   Human gene signature HUMGS04664.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 1250; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   276 BP;    90 A;    38 C;    43 G;    103 T;

  Query Match          11.7%;  Score 15;  DB 22;  Length 276;
  Best Local Similarity 88.2%;  Pred. No. 3.30e+01;
  Matches   15;  Conservative    0;  Mismatches   2;  Indels   0;  Gaps   0;

Db    107 atattctttnntctttt 123
          |||||||||  ||||||
Cp    115 ATATTCTTTATTCTTTT 99
```



```
RESULT   14
ID   T20572 standard; cDNA to mRNA; 301 BP.
AC   T20572;
DT   01-AUG-1996  (first entry)
DE   Human gene signature HUMGS01787.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 674; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   301 BP;    101 A;    48 C;    54 G;    97 T;

  Query Match          11.7%;  Score 15;  DB 19;  Length 301;
  Best Local Similarity 93.8%;  Pred. No. 3.30e+01;
  Matches   15;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db    228 ttnattcttttaaaag 243
          || ||||||||||||||
Cp    109 TTTATTCTTTTAAAAG 94


RESULT   15
ID   Q59533 standard; cDNA; 378 BP.
AC   Q59533;
```



```
DT   16-MAR-1994  (first entry)
DE   Human brain Expressed Sequence Tag EST02715.
KW   Gene transcription product; genetic markers; tagging; in vivo;
KW   transcription; mapping; locations; chromosomes; chromosomal; ss.
OS   Homo sapiens.
PN   WO9316178-A.
PD   19-AUG-1993.
PF   12-FEB-1993; U01294.
PR   12-FEB-1992; US-837195.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
PI   Adams MD, Moreno RF, Venter CJ;
DR   WPI; 93-272882/34.
PT   Enriched oligonucleotides and corresp. sequences - used as
PT   markers for human genes transcribed in-vivo, facilitate tagging
PT   of most human genes
PS   Example 4; Page 190; 500pp; English.
CC   The Expressed Sequence Tag was isolated from a human brain cDNA
CC   library as part of a large set of ESTs which can be used as markers
CC   for human genes transcribed in vivo. They can be used to facilitate
CC   tagging of most human genes, for mapping locations of expressed genes
CC   on chromosomes, for individual or forensic identification, for mapping
CC   locations of disease-associated genes, for identification of tissue
CC   type, and for prepn. of antisense sequences, probes and constructs.
CC   EST02715 has a 'poor' coding probability as evaluated using the
CC   coding-region prediction program CRM. See also Q59041-Q61440.
SQ   Sequence   378 BP;    116 A;    70 C;    82 G;    108 T;

  Query Match          11.7%;  Score 15;  DB 8;  Length 378;
  Best Local Similarity 100.0%;  Pred. No. 3.30e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    127 tcagaatttttaaag 141
          |||||||||||||||
Qy      3 TCAGAATTTTTAAAG 17


RESULT   16
ID   Q60061 standard; DNA; 401 BP.
AC   Q60061;
DT   16-MAR-1994  (first entry)
DE   Human brain Expressed Sequence Tag EST02044.
KW   Gene transcription product; genetic markers; tagging; in vivo;
KW   transcription; mapping; locations; chromosomes; chromosomal; ss.
OS   Homo sapiens.
PN   WO9316178-A.
PD   19-AUG-1993.
PF   12-FEB-1993; U01294.
PR   12-FEB-1992; US-837195.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
PI   Adams MD, Moreno RF, Venter CJ;
DR   WPI; 93-272882/34.
```

```
PT   Enriched oligonucleotides and corresp. sequences - used as
PT   markers for human genes transcribed in-vivo, facilitate tagging
PT   of most human genes
PS   Example 4; Page 275; 500pp; English.
CC   The Expressed Sequence Tag was isolated from a human brain cDNA
CC   library as part of a large set of ESTs which can be used as markers
CC   for human genes transcribed in vivo. They can be used to facilitate
CC   tagging of most human genes, for mapping locations of expressed genes
CC   on chromosomes, for individual or forensic identification, for mapping
CC   locations of disease-associated genes, for identification of tissue
CC   type, and for prepn. of antisense sequences, probes and constructs.
CC   EST02044 has a "poor" coding probability as evaluated using the
CC   coding-region prediction program CRM. See also Q59041-Q61440.
SQ   Sequence   401 BP;    132 A;    61 C;    78 G;    127 T;

  Query Match           11.7%;  Score 15;  DB 8;  Length 401;
  Best Local Similarity 100.0%;  Pred. No. 3.30e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    160 acaaaatttctgcttt 174
          ||||||||||||||||
Cp     28 ACAAAATTCTGCTTT 14


RESULT   17
ID   T43489 standard; DNA; 449 BP.
AC   T43489;
DT   11-AUG-1997  (first entry)
DE   ATM gene exon 57.
KW   ATM gene; ataxia-telangiectasia; progressive genetic disorder; antibody;
KW   central nervous system; immune system; chromosomal instability; therapy;
KW   cancer predisposition; radiation sensitivity; cell cycle abnormality;
KW   multi-system disease; autosomal recessive; cerebellar ataxia; cerebellum;
KW   general motor dysfunction; Purkinje cell; oculocutaneous telangiectasia;
KW   blood vessel; bulbar conjunctiva; facial skin; A-T; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   intron          1..202
FT                   /*tag= a
FT                   /number= 56
FT   exon            203..343
FT                   /*tag= b
FT                   /number= 57
FT   intron          344..449
FT                   /*tag= c
FT                   /number= 57
PN   WO9636691-A1.
PD   21-NOV-1996.
PF   16-MAY-1996; U07025.
PR   16-MAY-1995; US-441822.
PR   08-APR-1996; US-629001.
```



```
PA   (KOHN/) KOHN K I.
PA   (UYRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.
PI   Shiloh Y;
DR   WPI; 97-012070/01.
PT   New isolated ataxia-telangiectasia gene - used to develop prods. for
PT   the study, diagnosis and treatment of ataxia-telangiectasia.
PS   Claim 1; Page 71; 153pp; English.
CC   T43444-T43496 represent exons of the ATM gene of the invention.
CC   Ataxia-telangiectasia (A-T) is caused by mutations, insertions, or
CC   deletions in the coding region of the ATM gene. A-T is a progressive
CC   genetic disorder affecting the central nervous and immune systems. A-T
CC   involves chromosomal instability, cancer predisposition, radiation
CC   sensitivity, and cell cycle abnormalities. A-T is a multi-system disease
CC   inherited in an autosomal recessive manner. Cerebellar ataxia that
CC   gradually develops into general motor dysfunction is the first clinical
CC   hallmark of A-T, and results from progressive loss of Purkinje cells in
CC   the cerebellum. Oculocutaneous telangiectasia (dilation of blood vessels)
CC   develops in the bulbar conjunctiva and facial skin, and is later
CC   accompanied by graying of the hair and atrophic changes in the skin. The
CC   co-occurrence of cerebellar ataxia and telangiectases in the conjunctivae
CC   and occasionally on the facial skin (the second early hallmark of A-T)
CC   usually establishes the differential diagnosis of A-T from other
CC   cerebellar ataxias. The ATM gene, can be used in methods for detecting
CC   carriers of a defective gene that causes A-T. The gene can also be used
CC   to generate antibodies. The antibodies and methods can be used in the
CC   study, diagnosis and therapy of A-T.
SQ   Sequence   449 BP;    128 A;    70 C;    93 G;    158 T;

  Query Match           11.7%;  Score 15;  DB 31;  Length 449;
  Best Local Similarity 100.0%;  Pred. No. 3.30e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    248 tttaaagcagaattt 262
          |||||||||||||||
Qy     11 TTTAAAGCAGAATTT 25


RESULT   18
ID   Q20588 standard; DNA; 650 BP.
AC   Q20588;
DT   15-APR-1992  (first entry)
DE   Sequence from potato patatin 5' untranslated regions.
KW   Enzyme; cyclodextrin glycosyltransferase; plant vector;
KW   starch degradation; ss.
OS   Solanum tuberosum Russet Burbank.
PN   WO9119808-A.
PD   26-DEC-1991.
PF   11-JUN-1991; U04116.
PR   11-JUN-1990; US-536392.
PA   (CALG-) CALGENE INC.
PI   Stalker DM, Shewmaker CK, Oakes JV;
```



```
DR   WPI; 92-024425/03.
PT   Gene for cyclodextrin glycosyltransferase - used for transforming
PT   plants for prodn. of starch cyclodextrin cpds., while minimising
PT   degradation
PS   Example; Fig 3; 68pp; English.
CC   A binary vector, pCGN2160b, was constructed contg. the patatin-5'
CC   region for Solanum tuberosum var Kennebec, DNA encoding a transit
CC   peptide from soybean RuBisCo SSU protein, 48 bp of DNA encoding 16
CC   AAs of mature RuBisCo SSU protein from pea, the cyclodextrin
CC   glycosyltransferase (CGT) coding region from K. pneumoneae and the
CC   nos-3'-region.A.tumefaciens strain 2760 was transformed with
CC   pCGN2160b and the transformants were co-cultivated with Solanum
CC   tuberosum var. Russet Burbank tuber discs to produce transformed
CC   plants.  The transformed plants were grown and  cyclodextrin (CD)
CC   was isolated from tubers.
SQ   Sequence   650 BP;    248 A;    97 C;    116 G;    189 T;

  Query Match           11.7%;  Score 15;  DB 3;  Length 650;
  Best Local Similarity 100.0%;  Pred. No. 3.30e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     59 tttattcttttaaaa 73
          |||||||||||||||
Cp    109 TTTATTCTTTTAAAA 95


RESULT   19
ID   Q20587 standard; DNA; 703 BP.
AC   Q20587;
DT   15-APR-1992  (first entry)
DE   Sequence from potato patatin 5' untranslated regions.
KW   Enzyme; cyclodextrin glycosyltransferase; plant vector;
KW   starch degradation; ss.
OS   Solanum tuberosum Maris Piper.
PN   WO9119808-A.
PD   26-DEC-1991.
PF   11-JUN-1991; U04116.
PR   11-JUN-1990; US-536392.
PA   (CALG-) CALGENE INC.
PI   Stalker DM, Shewmaker CK, Oakes JV;
DR   WPI; 92-024425/03.
PT   Gene for cyclodextrin glycosyltransferase - used for transforming
PT   plants for prodn. of starch cyclodextrin cpds., while minimising
PT   degradation
PS   Example; Fig 2; 68pp; English.
CC   A binary vector, pCGN2160b, was constructed contg. the patatin-5'
CC   region for Solanum tuberosum var Kennebec, DNA encoding a transit
CC   peptide from soybean RuBisCo SSU protein, 48 bp of DNA encoding 16
CC   AAs of mature RuBisCo SSU protein from pea, the cyclodextrin
CC   glycosyltransferase (CGT) coding region from K. pneumoneae and the
CC   nos-3'-region.A.tumefaciens strain 2760 was transformed with
```



```
CC   pCGN2160b and the transformants were co-cultivated with Solanum
CC   tuberosum var. Russet Burbank tuber discs to produce transformed
CC   plants.  The transformed plants were grown and  cyclodextrin (CD)
CC   was isolated from tubers.
SQ   Sequence   703 BP;    271 A;    112 C;    126 G;    194 T;

  Query Match           11.7%;  Score 15;  DB 3;  Length 703;
  Best Local Similarity 100.0%;  Pred. No. 3.30e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     53 tttattcttttaaaa 67
          |||||||||||||||
Cp    109 TTTATTCTTTTAAAA 95


RESULT   20
ID   Q20586 standard; DNA; 718 BP.
AC   Q20586;
DT   15-APR-1992  (first entry)
DE   Sequence from potato patatin 5' untranslated regions.
KW   Enzyme; cyclodextrin glycosyltransferase; plant vector;
KW   starch degradation; ss.
OS   Solanum tuberosum Kennebec variety.
PN   WO9119808-A.
PD   26-DEC-1991.
PF   11-JUN-1991; U04116.
PR   11-JUN-1990; US-536392.
PA   (CALG-) CALGENE INC.
PI   Stalker DM, Shewmaker CK, Oakes JV;
DR   WPI; 92-024425/03.
PT   Gene for cyclodextrin glycosyltransferase - used for transforming
PT   plants for prodn. of starch cyclodextrin cpds., while minimising
PT   degradation
PS   Example; Fig 2; 68pp; English.
CC   A binary vector, pCGN2160b, was constructed contg. the patatin-5'
CC   region for Solanum tuberosum var Kennebec, DNA encoding a transit
CC   peptide from soybean RuBisCo SSU protein, 48 bp of DNA encoding 16
CC   AAs of mature RuBisCo SSU protein from pea, the cyclodextrin
CC   glycosyltransferase (CGT) coding region from K. pneumoneae and the
CC   nos-3'-region.A.tumefaciens strain 2760 was transformed with
CC   pCGN2160b and the transformants were co-cultivated with Solanum
CC   tuberosum var. Russet Burbank tuber discs to produce transformed
CC   plants.  The transformed plants were grown and  cyclodextrin (CD)
CC   was isolated from tubers.
SQ   Sequence   718 BP;    273 A;    115 C;    132 G;    198 T;

  Query Match           11.7%;  Score 15;  DB 3;  Length 718;
  Best Local Similarity 100.0%;  Pred. No. 3.30e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     59 tttattcttttaaaa 73
```

```
              ||||||||||||||||
Cp    109 TTTATTCTTTTAAAA 95


RESULT   21
ID   N82022 standard; DNA; 767 BP.
AC   N82022;
DT   22-OCT-1990  (first entry)
DE   Gene encoding 25 kD surface antigen , Pfs25.
KW   25 kD surface antigen; Pfs25; vaccine; malaria; zygotes; ookinites;
KW   mosquito; transmission-blocking immunity; sexual development; ss.
OS   Plasmodium falciparum.
FH   Key             Location/Qualifiers
FT   signal_peptide  81..122
FT                   /*tag=  a
FT   mat_peptide     123..725
FT                   /*tag=  b
FT                   /label=Pfs25
PN   US7188918-A.
PD   08-NOV-1988.
PF   02-MAY-1988; 188918.
PR   02-MAY-1988; US-188918.
PA   (USSH) US Dept of Health and Human.
PI   Miller LH, Kaslow DC;
DR   WPI; 88-368450/51.
DR   P-PSDB; P82052.
PT   New gene for surface antigen of plasmodium falciparum - in sexual
PT   development stages - useful in transmission-blocking malaria vaccine.
PS   Disclosure; p; English.
CC   The gene encodes the 25 kD surface protein of zygotes and ookinates
CC   of P. falciparum.  The cloned gene can be used to express an antigen
CC   which is useful for a malarial vaccine designed to induce transmission
CC   blocking immunity to Pfs25 and other sexual stage antigens.
SQ   Sequence   767 BP;   308 A;   76 C;   122 G;   261 T;

  Query Match           11.7%;  Score 15;  DB 1;  Length 767;
  Best Local Similarity 100.0%;  Pred. No. 3.30e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     60 tttattcttttaaaa 74
          ||||||||||||||||
Cp    109 TTTATTCTTTTAAAA 95


RESULT   22
ID   Q43503 standard; DNA; 767 BP.
AC   Q43503;
DT   29-SEP-1993  (first entry)
DE   Sequence encoding the sexual stage surface protein Pfs25
KW   Malaria; sexual stage surface protein; Pfs25; recombinant protein;
KW   vaccine; ss.
```



```
OS   Plasmodium falciparum.
FH   Key             Location/Qualifiers
FT   cds             75..725
FT                   /*tag=  a
FT   cds             84..125
FT                   /*tag=  b
FT                   /product= hydrophobic region
FT   cds             681..725
FT                   /*tag=  c
FT                   /product= hydrophobic region
PN   US5217898-A.
PD   08-JUN-1993.
PF   02-MAY-1988; 188918.
PR   02-MAY-1988; US-188918.
PR   22-FEB-1991; US-658845.
PR   08-MAY-1991; US-697275.
PA   (CHIR ) CHIRON CORP.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
PI   Barr PJ,  Kaslow DC;
DR   WPI; 93-196250/24.
DR   P-PSDB; R38049.
PT   Antigen for producing transmission blocking antibodies against
PT   malaria - obtd. by expressing recombinant PPS25 Plasmodium
PT   falciparum protein in yeast cells
PS   Disclosure; columns 9-10; 10pp; English.
CC   Pfs25 contains a putative signal sequence at the amino terminus and
CC   a short hydrophobic anchor at the carboxyl terminus. These regions
CC   are identified by the hydrophobic regions noted in FT. The inventors
CC   have engineered, by chemical synthesis and mutagenesis, a gene
CC   expressible in yeast cells which produces an analogue of Pfs25,
CC   designated Pfs25-B. Absent from the yeast-expressed protein was the
CC   amino-terminal secretory sequence and the hydrophobic carboxyl
CC   terminal region. The N-linked glycosylation sites were also removed.
CC   Pfs25-B was shown to react with conformation dependent MAbs despite
CC   its conformational changes. In addition it elicited transmission-
CC   blocking activity in mice and monkeys. The DNA segment encoding
CC   Pfs25-B is operably linked to a vector, e.g., pBS24-Pfs25-B.
SQ   Sequence   767 BP;   308 A;   76 C;   122 G;   261 T;

  Query Match           11.7%;  Score 15;  DB 7;  Length 767;
  Best Local Similarity 100.0%;  Pred. No. 3.30e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     60 tttattcttttaaaa 74
          ||||||||||||||||
Cp    109 TTTATTCTTTTAAAA 95


RESULT   23
ID   T13643 standard; DNA; 861 BP.
AC   T13643;
```



```
DT   06-SEP-1996  (first entry)
DE   AcNPV ORF 27, residues 22600-23460.
KW   Autographa californica nuclear polyhedrosis virus clone 6;
KW   disruption; non-essential gene; heterologous protein production;
KW   expression vector; baculovirus; ss.
OS   Autographa californica nuclear polyhedrosis virus clone 6.
FH   Key             Location/Qualifiers
FT   cds             1..861
FT                   /*tag=  a
FT                   /number= ORF 27
FT                   /product= inhibitor of apoptosis protein (IAP1)
FT                   /note= "corresponds to AcNPV nucleotides
FT                   22600-23460"
PN   WO9601320-A2.
PD   18-JAN-1996.
PF   30-JUN-1995; IB0578.
PR   04-JUL-1994; GB-013420.
PA   (NATU-) NATURAL ENVIRONMENT RES COUNCIL.
PI   Ayres M,  Bishop D,  Possee R;
DR   WPI; 96-087670/09.
DR   GENBANK; L22858.
PT   Autographa californica nuclear polyhedrosis virus complete genome
PT   sequence  - useful in the prodn. of vectors for enhanced
PT   heterologous protein expression, such as interleukin(s),
PT   interferon(s) and neurotoxin(s).
PS   Claim 4; Page 90-186; 122pp; English.
CC   T13636-731 show open reading frames 13, 20, 22-30, 32, 38, 41-46, 50-60,
CC   62-63, 66, 68-79, 81-87, 91-92, 96-98, 101-103, 106-127, 129-130, 140-
CC   146, 148-150, 152 and 154 from a total of 154 ORFs identified in the
CC   Autographa californica nuclear polyhedrosis virus (AcNPV) clone 6. Each
CC   gene is numbered according to its position in the virus genome beginning
CC   at the left end of the linear map, and irrespective of its orientation.
CC   The direction of transcription is relative to that of the polyhedrin
CC   gene. Of the 154 ORFs identified it was found that some of the ORFs (ORFs
CC   27, 30, 32, 71, 86, 123, 126 and 127) are dispensable for virus
CC   replication in cell culture or insect larvae.  These genes can be deleted
CC   from the genome to: (a) provide additional sites for inserting single or
CC   multiple copies of foreign genes; and (b) to reduce the size of the virus
CC   genome. The present sequence ORF 27 is on the same strand as the
CC   polyhedrin gene. ORF 27 is referred to as ORF 13 in literature by other
CC   authors.  The gene encodes a potential apoptosis inhibitory protein
CC   (IAP1) contg. at least 1 potential zinc finger.
SQ   Sequence   861 BP;   248 A;   175 C;   202 G;   236 T;

  Query Match           11.7%;  Score 15;  DB 21;  Length 861;
  Best Local Similarity 100.0%;  Pred. No. 3.30e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    615 agaattttgtcaaaa 629
          ||||||||||||||||
Qy     19 AGAATTTTGTCAAAA 33
```



```
RESULT   24
ID   Q78152 standard; DNA; 1047 BP.
AC   Q78152;
DT   06-JUN-1995  (first entry)
DE   Tuber specific class 1 patatin promoter.
KW   Potato; tuber; promoter; patatin; barley; endosperm; enzyme;
KW   ADP glucose pyrophosphorylase; AGP; starch; transgenic plant; ss.
OS   Solanum tuberosum.
PN   WO9424292-A.
PD   27-OCT-1994.
PF   07-APR-1994; E01082.
PR   08-APR-1993; GB-007408.
PA   (DANI-) DANISCO AS.
PI   Kleczkowski L,  Marcussen J,  Okkels F,  Olsen O,  Poulsen P;
PI   Villand P;
DR   WPI; 94-341871/42.
PT   Transgenic starch producing organism - comprises exogenous ADP
PT   glucose pyrophosphorylase enzyme DNA sequence
PS   Claim 18; Fig 17; 88pp; English.
CC   Q78150 is a 1.8 kb cDNA fragment encoding the small subunit ADP-
CC   glucose pyrophosphorylase from barley endosperm. For the construction of
CC   plasmid pPPS1, a 1.8 kb EcoRI-HindIII cDNA fragment encoding the barley
CC   endosperm AGP small subunit (beps) was cloned i n the BamHI site of
CC   plasmid pPATA1. Plasmid pPATA 1 is a deriv. of a plasmid pUC19 and
CC   has tuber specific patatin class I promoter (Q78152), a polylinker
CC   cloning region, and a 35S terminator. From the resulting plasmid
CC   the 3.1 kb KpnI fragment contg. the patatin promoter, the beps cDNA,
CC   and the 35S terminator was isolated and inserted in the KpnI site of
CC   the plant transformation vector pBKL4 to yield plasmid pPPS1, which
CC   is claimed. The patating promoter (Q78152) is a tuber specific
CC   promoter from potato (Mignery et al. 1988, Gene 62:27-44).
SQ   Sequence   1047 BP;   375 A;   162 C;   172 G;   338 T;

  Query Match           11.7%;  Score 15;  DB 13;  Length 1047;
  Best Local Similarity 100.0%;  Pred. No. 3.30e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    395 tttattcttttaaaa 409
          ||||||||||||||||
Cp    109 TTTATTCTTTTAAAA 95


RESULT   25
ID   T76777 standard; DNA; 1390 BP.
AC   T76777;
DT   15-SEP-1997  (first entry)
DE   Staphylococcus aureus exfoliative toxin A (ETA) gene.
KW   Asymmetric polymerase chain reaction; nucleic acid amplification;
KW   PCR; detection; assay; exfoliative toxin A; ETA; skin lesion;
```

```
KW   competitive primer; capture probe; ds.
OS   Staphylococcus aureus.
FH   Key             Location/Qualifiers
FT   primer_bind     complement (165..184)
FT                   /*tag=  a
FT                   /bound_moeity= ETA-A2
FT                   /note= "See T76778"
FT   primer_bind     417..436
FT                   /*tag=  b
FT                   /bound_moeity= ETA-B
FT                   /note= "See T76779"
FT   primer_bind     321..360
FT                   /*tag=  c
FT                   /bound_moeity= capture_probe
FT                   /note= "See T76781 and T76782"
FT   primer_bind     389..410
FT                   /*tag=  d
FT                   /bound_moeity= label_probe
FT                   /note= "See T76783"
PN   US5627054-A.
PD   06-MAY-1997.
PF   05-APR-1996; 628417.
PR   05-APR-1996; US-628417.
PA   (USSA ) US SEC OF ARMY.
PI   Gillespie D;
DR   WPI; 97-271311/24.
PT   Quantitative nucleic acid amplification - by competitor primer
PT   asymmetric polymerase chain reaction
PS   Example 1; Columns 7-10; 9pp; English.
CC   In a specific example of a novel process for amplifying an amount
CC   (known or unknown) of a double-stranded nucleic acid segment to produce
CC   single-stranded nucleic acid in an amount that is proportional to the
CC   starting amount of the nucleic acid, this Staphylococcus aureus
CC   exfoliative toxin A (ETA) gene was used as the DNA template. The
CC   region comprising nucleotides 165-436 was amplified by symmetric,
CC   asymmetric or competitor primer asymmetric PCR using the primers
CC   ETA-A2 and ETA-B (see T76778 and T76779). For asymmetric PCR, the
CC   amount of primer ETA-B was reduced and for competitor primer
CC   asymmetric PCR a competitor primer ETA-B2 (see T76780) was added
CC   with upstream primer ETA-A2 after the initial cycling reaction. PCR
CC   products containing ETA-specific sequences were detected
CC   radioactively by a capture system which employed a bifunctional
CC   capture probe ETA-CP (see T76781 and T76782). ETA-CP was designed
CC   to capture the amplified sense strand onto capture membranes
CC   through hybridisation between the first 40 nucleotides of ETA-CP
CC   and nucleotides 321-360 of the ETA gene and through hybridisation
CC   of the poly(dA) tail on ETA-CP with poly(dT) tails on the capture
CC   membranes. A radioactively labelled "label probe" (see T76783),
CC   complementary to nucleotides 389-410 of the ETA gene was used to
CC   detect the amplicons. Results showed that hybridisation of the
CC   capture probe and label probe to the denatured symmetric PCR
```



```
CC   product was much less efficient than hybridisation to the
CC   single-stranded PCR products of the asymmetric and competitor
CC   primer asymmetric reactions.
SQ   Sequence   1390 BP;    556 A;    161 C;    248 G;    425 T;

  Query Match           11.7%;  Score 15;  DB 31;  Length 1390;
  Best Local Similarity 100.0%;  Pred. No. 3.30e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1061 caaagtgtctcatct 1075
          |||||||||||||||
Qy     76 CAAAGTGTCTCATCT 90


RESULT   26
ID   Q03256 standard; DNA; 1431 BP.
AC   Q03256;
DT   25-JUL-1990  (first entry)
DE   Sequence encoding beta-1,3-glucanase derived from cDNA clone pBSGluc39.3.
KW   Beta-1,3-glucanase; papermaking; polysaccharide hydrolase;
KW   ds.
OS   Nicotiana sp.
FH   Key             Location/Qualifiers
FT   cds             31..1140
FT                   /*tag=  a
PN   EP-353191-A.
PD   31-JAN-1990.
PF   20-JUL-1989; 810555.
PR   29-JUL-1988; US-353312.
PA   (CIBA) Ciba Geigy AG.
PI   Shinshi H, Wenzler H, Hofsteenge J, Ryals J, Sperisen C;
DR   WPI; 90-031717/05.
DR   P-PSDB; R03056.
PT   Recombinant DNA encoding beta-1,3-glucanase polypeptide(s) -
PT   useful in plant protection and biomass conversion pref. isolated
PT   from tobacco.
PS   Disclosure; Fig 11; 41pp; English.
CC   Beta-1,3-glucanase is involved in plant defence against pathogens and as
CC   it hydrolyses polyssaccharides, useful in biomass conversion or paper
CC   making. The clones allow for transfer of the gene to other species and
CC   control of gene expression.
SQ   Sequence   1431 BP;    416 A;    284 C;    292 G;    439 T;

  Query Match           11.7%;  Score 15;  DB 1;  Length 1431;
  Best Local Similarity 100.0%;  Pred. No. 3.30e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1141 tgagatgagacactt 1155
          |||||||||||||||
Cp     92 TGAGATGAGACACTT 78
```



```
RESULT   27
ID   T68767 standard; DNA; 1657 BP.
AC   T68767;
DT   19-AUG-1997  (first entry)
DE   ATM mutant 497del7514 coding sequence.
KW   ATM gene; ataxia-telangiectasia; progressive genetic disorder; antibody;
KW   central nervous system; immune system; chromosomal instability; therapy;
KW   cancer predisposition; radiation sensitivity; cell cycle abnormality;
KW   multi-system disease; autosomal recessive; cerebellar ataxia; cerebellum;
KW   general motor dysfunction; Purkinje cell; oculocutaneous telangiectasia;
KW   blood vessel; bulbar conjunctiva; facial skin; A-T; mutein; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             1..546
FT                   /*tag=  a
FT   misc_difference 496..497
FT                   /*tag=  b
FT                   /note= "site of 7514 nucleotide deletion"
PN   WO9636691-A1.
PD   21-NOV-1996.
PF   16-MAY-1996; U07025.
PR   16-MAY-1995; US-441822.
PR   08-APR-1996; US-629001.
PA   (KOHN/) KOHN K I.
PA   (UYRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.
PI   Shiloh Y;
DR   WPI; 97-012070/01.
DR   P-PSDB; W19684.
PT   New isolated ataxia-telangiectasia gene - used to develop prods. for
PT   the study, diagnosis and treatment of ataxia-telangiectasia.
PS   Claim 5; Page -; 153pp; English.
CC   T68731-T68785 represent mutations of the ATM gene of the invention (see
CC   T43497 for wild type sequence). The ATM gene of the invention, is located
CC   in the human genome at chromosome segment 11q22-23. Ataxia-telangiectasia
CC   (A-T) is caused by mutations, insertions, or deletions (such as
CC   represented by this sequence) in the coding region of the ATM gene. A-T
CC   is a progressive genetic disorder affecting the central nervous and
CC   immune systems. A-T involves chromosomal instability, cancer
CC   predisposition, radiation sensitivity, and cell cycle abnormalities. A-T
CC   is a multi-system disease inherited in an autosomal recessive manner. The
CC   wild type gene, can be used in methods for detecting carriers of a
CC   defective gene that causes A-T. The gene can also be used to generate
CC   antibodies. The methods and antibodies can be used in the study,
CC   diagnosis and therapy of A-T.
SQ   Sequence   1657 BP;    545 A;    289 C;    376 G;    447 T;

  Query Match           11.7%;  Score 15;  DB 31;  Length 1657;
  Best Local Similarity 100.0%;  Pred. No. 3.30e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```



```
Db    542 tttaaagcagaaattt 556
          |||||||||||||||
Qy     11 TTTAAAGCAGAATTT 25


RESULT   28
ID   Q20243 standard; DNA; 1807 BP.
AC   Q20243;
DT   25-MAR-1992  (first entry)
DE   Mycoplasma-specific DNA fragment.
KW   Detection; ss.
OS   Mycoplasma hyopneumoniae.
PN   J03254700-A.
PD   13-NOV-1991.
PF   06-MAR-1990; 052687.
PR   06-MAR-1990; JP-052687.
PA   (TAKA-) TAKARA SHUZO KK.
DR   WPI; 92-002469/01.
PT   Highly sensitive detection of mycoplasma - by detecting region of
PT   restriction enzyme map and amplifying DNA sequence
PS   Disclosure; Fig 3; 10pp; Japanese.
CC   The sequence is a 1807bp fragment of M. hyopneumoniae and was used
CC   to test a kit for the detection of pig mycoplasma. The DNA is first
CC   amplified by PCR and then the probe is used to detect the amplified
CC   DNA. This method is highly sensitive. See also Q20243-Q20245 and
CC   Q21024-Q21027.
SQ   Sequence   1807 BP;    698 A;    205 C;    254 G;    648 T;

  Query Match           11.7%;  Score 15;  DB 3;  Length 1807;
  Best Local Similarity 100.0%;  Pred. No. 3.30e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    645 ttttaaaagaataaa 659
          |||||||||||||||
Qy     95 TTTTAAAAGAATAAA 109


RESULT   29
ID   N93087 standard; DNA; 1969 BP.
AC   N93087;
DT   25-JUN-1990  (first entry)
DE   Human basic fibroblast growth factor.
KW   Basic fibroblast growth factor; pUC9-TSF11; pUC9delH3-pTSF-3; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   mat_peptide     970..1434
FT                   /*tag=  a
FT                   /product=human basic fibroblast growth factor
PN   EP-298723-A.
PD   11-JAN-1989.
PF   06-JUL-1988; 306158.
```

```
PR   07-JUL-1987; US-070797.
PA   (BIOT-) Biotechn Res Assoc.
PI   Fiddes JC, Abraham JA, Protter A;
DR   WPI; 89-009785/02.
DR   P-PSDB; P94038.
PT   Recombinant DNA encoding new fibroblast growth factor
PT   analogues - useful eg for accelerating wound healing and
PT   to control neovascularisation.
PS   Disclosure; Fig 1-2 and 1-2; 44pp; English.
CC   The sequence may be mutated to encode an analogue, bFGF-C78/96S,
CC   which has reduced affinity for heparin. One or more positively-charged
CC   AAs in the heparin-binding domain (AAs 128-138) are replaced by neutral
CC   or negatively-charged residues. A recombinant vector (pUC9-TSF11 or
CC   pUC9delH3-pTSF-3) contg. the mutated DNA may be used to transform host
CC   cells for prodn. of the analogue.
CC   See also N93088-89.
SQ   Sequence   1969 BP;   458 A;   481 C;   572 G;   458 T;

  Query Match          11.7%;  Score 15;  DB 1;  Length 1969;
  Best Local Similarity 100.0%;  Pred. No. 3.30e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    233 cttttaaaagttgag 247
          ||||||||||||||| 
Cp    103 CTTTTAAAAGTTGAG 89


RESULT   30
ID   Q64158 standard; DNA; 2128 BP.
AC   Q64158;
DT   03-FEB-1995  (first entry)
DE   Partial coding sequence of tyrosine kinase receptor protein.
KW   Tyrosine kinase; receptor; proto-oncogene; trk; detection;
KW   diagnosis; antibody; treatment; tumour; antisense; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             1..1953
FT                   /*tag= a
FT                   /product= Partial sequence of protein tyrosine
FT                   kinase.
PN   DE4239817-A.
PD   01-JUN-1994.
PF   26-NOV-1992; 239817.
PR   26-NOV-1992; DE-239817.
PA   (CHEM-) CHEMOTHERAPEUTISCHES FORSCHUNG.
PI   Holtrich U,  Ruebsamen-waigmann H,  Strebhardt K;
DR   WPI; 94-184380/23.
DR   P-PSDB; R54089.
PT   New protein tyrosine kinase and related nucleic acid - vectors,
PT   transformed cells, etc., useful for diagnosis and treatment of
PT   tumours
```



```
PS   Claim 4; Page 8; 9pp; German.
CC   The gene is related to the trk proto-oncogene.  Antibodies against
CC   the encoded polypeptide are useful for diagnosis and for the
CC   treatment of tumours.  The antibodies may also be radiolabelled or
CC   coupled to a cytotoxin for destruction of cancer cells.  Antisense
CC   nucleic acid can be used to inhibit gene expression.
SQ   Sequence   2128 BP;   532 A;   556 C;   504 G;   536 T;

  Query Match          11.7%;  Score 15;  DB 11;  Length 2128;
  Best Local Similarity 100.0%;  Pred. No. 3.30e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1659 tatcttgctgggcaa 1673
          ||||||||||||||| 
Qy    113 TATCTTGCTGGGCAA 127


RESULT   31
ID   Q56758 standard; DNA; 2255 BP.
AC   Q56758;
DT   12-OCT-1994  (first entry)
DE   Chitinase derivative gene #1.
KW   Chitinase; derivative; beta-1,4-glycoside bond; chitin;
KW   beta-N-acetylhexosaminidase activity; yeast; ss.
OS   Rhizopus niveus.
FH   Key             Location/Qualifiers
FT   cds             138..1750
FT                   /*tag= a
FT                   /product= Chitinase
FT   exon            138..242
FT                   /*tag= b
FT   intron          243..305
FT                   /*tag= c
FT   exon            306..498
FT                   /*tag= d
FT   intron          499..566
FT                   /*tag= e
FT   exon            567..1750
FT                   /*tag= f
PN   J06046849-A.
PD   22-FEB-1994.
PF   28-JUL-1992; 201427.
PR   28-JUL-1992; JP-201427.
PA   (KAGO ) KAGOME KK.
DR   WPI; 94-097015/12.
DR   P-PSDB; R48669.
PT   New chitinase enzyme and coding sequence - cleaves beta-1,4-glycoside
PT   bond of chitin but has no beta-N-acetylhexosaminidase activity
PS   Claim 11; Page 21-23; 66pp; Japanese.
CC   The sequences given in Q56756-61 encode chitinase derivatives which
CC   cleave the beta-1,4-glycoside bond of chitin but have substantially
```



```
CC   no beta-N-acetylhexosaminidase activity.  These sequences may be
CC   introduced into yeast and cultured for the production of the
CC   chitinase proteins.
SQ   Sequence   2255 BP;   614 A;   448 C;   446 G;   747 T;

  Query Match          11.7%;  Score 15;  DB 10;  Length 2255;
  Best Local Similarity 100.0%;  Pred. No. 3.30e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    808 aaagggtcatttttt 822
          ||||||||||||||| 
Qy     31 AAAGGGTCATTTTTT 45


RESULT   32
ID   V04155 standard; DNA; 2781 BP.
AC   V04155;
DT   22-JUN-1998  (first entry)
DE   Blood group antigen binding (Bab) adhesin babb gene.
KW   Blood group antigen binding adhesin; babb gene; infection;
KW   gastritis; acid peptic disease; therapy; diagnosis; vaccine;
KW   immunisation; ss.
OS   Helicobacter pylori strain CCUG 17875.
FH   Key             Location/Qualifiers
FT   CDS             409..2532
FT                   /*tag= a
FT   sig_peptide     409..462
FT                   /*tag= b
FT   mat_peptide     463..2529
FT                   /*tag= c
PN   WO9747646-A1.
PD   18-DEC-1997.
PF   10-JUN-1997; SE1009.
PR   19-MAR-1997; SE-001014.
PR   10-JUN-1996; SE-002287.
PA   (BORE/) BOREN T.
PI   Arnqvist A, Boren T, Hammarstrom L, Ilver D, Normark S,
PI   Marstroem L;
DR   WPI; 98-052240/05.
DR   P-PSDB; W41523.
PT   Helicobacter pylori blood group antigen binding adhesin protein -
PT   that binds fucosylated blood group antigen, useful to diagnose,
PT   prevent and treat H. pylori infection
PS   Claim 13; Page 34-36; 51pp; English.
CC   This DNA sequence comprises the Helicobacter pylori CCUG 17875
CC   babb gene that codes for a blood group antigen binding (Bab)
CC   adhesin (see W41523).  The Babb adhesin was isolated using a novel
CC   technique, receptor activity directed affinity tagging (retagging).
CC   The N-terminal 20 amino acids (see W41521) of the isolated protein
CC   was used to design degenerate primers that were used to isolate
CC   the babb gene.  A baba gene (see V04154) was also isolated.  The
```



```
CC   bab genes can be used in the recombinant production of Bab
CC   adhesins.  The Bab adhesins and DNA can be used in the
CC   diagnosis, treatment and/or prophylaxis of H. pylori induced
CC   infections such as gastritis and acid peptic disease, e.g. by
CC   active immunisation.
SQ   Sequence   2781 BP;   903 A;   585 C;   544 G;   749 T;

  Query Match          11.7%;  Score 15;  DB 40;  Length 2781;
  Best Local Similarity 100.0%;  Pred. No. 3.30e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    360 ttattcttttaaaag 374
          ||||||||||||||| 
Cp    108 TTATTCTTTTAAAAG 94


RESULT   33
ID   Q92521 standard; cDNA to mRNA; 3157 BP.
AC   Q92521;
DT   26-NOV-1995  (first entry)
DE   Human colonic adenocarcinoma kinase 2 (CCK-2) cDNA.
KW   Mammary carcinoma kinase 10; MCK-10; transmembrane receptor; CCK-2;
KW   receptor tyrosine kinase; colonic adenocarcinoma kinase 2; cancer; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             370..2937
FT                   /*tag= a
PN   WO9514088-A.
PD   26-MAY-1995.
PF   16-NOV-1994; E03797.
PR   16-NOV-1993; US-153397.
PA   (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PI   Alves PHZ,  Ullrich A;
DR   WPI; 95-224054/29.
DR   P-PSDB; R75503.
PT   New nucleic acid encoding MCK-10 receptor tyrosine kinase - and
PT   derived vectors, transformed cells, proteins and antibodies useful
PT   for diagnosis and treatment of proliferative disease, esp. cancer,
PT   and for screening modulators
PS   Disclosure; Page 56-57; 115pp; English.
CC   A member of the mammary carcinomal kinase 10 (MCK-10) receptor
CC   tyrosine kinase family was identified using a PCR (with two degenerate
CC   oligo primer pools based on conserved sequences of the kinase domain of
CC   receptor tyrosine kinases) and cDNA prepd. from colonic
CC   adenocarcinoma RNA. The nt sequence of the novel receptor,
CC   designated CCK-2, is given in Q92521 and the deduced AA sequence in
CC   R75503. Analysis of the CCK-2 nt and AA sequence indicated
CC   significant homology with MCK-10 throughout the extracellular,
CC   transmembrane and intracellular regions. The regions of homology
CC   extend into the N-terminus consensus sequence for the discoidin I
CC   like family of proteins.
```

```
SQ   Sequence   3157 BP;    790 A;    802 C;    759 G;    806 T;

  Query Match           11.7%;  Score 15;  DB 15;  Length 3157;
  Best Local Similarity 100.0%;  Pred. No. 3.30e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2643 tatcttgctgggcaa 2657
        |||||||||||||||
Qy    113 TATCTTGCTGGGCAA 127


RESULT   34
ID   T93784 standard; cDNA; 3157 BP.
AC   T93784;
DT   16-FEB-1998  (first entry)
DE   CCK-2, a human mammary carcinoma kinase LO MCK-10) family member, cDNA.
KW   Mammary carcinoma kinase; MCK-10; CCK-2; receptor tyrosine kinase;
KW   proliferative disease; cancer; MCK-10 activity; aberrant expression; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             370..2934
FT                   /*tag=  a
PN   US5677144-A.
PD   14-OCT-1997.
PF   08-NOV-1994; 336343.
PR   16-NOV-1993; US-153397.
PA   (ALVE/) ALVES F H E.
PA   (ULLR/) ULLRICH A.
PI   Alves FHE,  Ullrich A;
DR   WPI; 97-511869/47.
DR   P-PSDB; W34671.
PT   Truncated receptor tyrosine kinase CCK-2 - and nucleic acid coding
PT   for it, useful for cancer diagnosis
PS   Claim 9; Fig 3; 70pp; English.
CC   The present sequence represents the cDNA sequence of human CCK-2, a
CC   member of the mammary carcinoma kinase 10 (MCK-10, T93785) family of
CC   receptor tyrosine kinases. Expression of CCK-2 is associated with
CC   proliferative diseases such as cancer. The CCK-2 gene was identified by
CC   PCR and a cDNA prepared from colonic adenocarcinoma RNA. CCK-2 is
CC   expressed in a wide variety of cancer cell lines and tumour tissue. The
CC   CCK-2 nucleic acids can be used for diagnostic purposes to detect
CC   aberrant expression of CCK-2 genes. Engineered cell lines, containing
CC   recombinant vectors with the present sequence, are useful for producing
CC   infectious retroviral particles. The cell lines may also be used to
CC   evaluate and screen drugs involved in CCK-2 activation and regulation.
SQ   Sequence   3157 BP;    790 A;    802 C;    759 G;    806 T;

  Query Match           11.7%;  Score 15;  DB 35;  Length 3157;
  Best Local Similarity 100.0%;  Pred. No. 3.30e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```



```
Db   2643 tatcttgctgggcaa 2657
        |||||||||||||||
Qy    113 TATCTTGCTGGGCAA 127


RESULT   35
ID   Q92523 standard; cDNA to mRNA; 3157 BP.
AC   Q92523;
DT   26-NOV-1995  (first entry)
DE   Human colonic adenocarcinoma kinase 2 (CCK-2) cDNA.
KW   Mammary carcinoma kinase 10; MCK-10; transmembrane receptor; CCK-2;
KW   receptor tyrosine kinase; colonic adenocarcinoma kinase 2; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             370..2937
FT                   /*tag=  a
PN   WO9514089-A.
PD   26-MAY-1995.
PF   16-NOV-1994; E03799.
PR   16-NOV-1993; US-153397.
PA   (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PI   Alves FHE,  Ullrich A;
DR   WPI; 95-224055/29.
DR   P-PSDB; R75505.
PT   New nucleic acid encoding CCK-2 receptor tyrosine kinase - and
PT   derived vectors, transformed cells, proteins and antibodies, useful
PT   for diagnosis and treatment of proliferative and nervous system
PT   diseases and for screening modulators
PS   Disclosure; Page 73-74; 115pp; English.
CC   A member of the mammary carcinomal kinase 10 (MCK-10) receptor
CC   tyrosine kinase family was identified using a PCR (with two
CC   degenerate oligo primer pools based on conserved sequences of the
CC   kinase domains of receptor tyrosine kinases) and cDNA prepd. from
CC   colonic adenocarcinoma RNA. The nt sequence of the novel receptor,
CC   designated CCK-2, is given in Q92523 and the deduced AA sequence in
CC   R75505. Analysis of CCK-2 nt and AA sequences indicated significant
CC   homology with MCK-10 throughout the extracellular, transmembrane
CC   and intracellular regions. The regions of homology extend into the
CC   N-terminus consensus seuence fothe discoidin I like family of
CC   proteins. CCK-2 was predominantly found in all stromal cells
CC   whereas MCK-10 expression was stronglyu confined to neoplastic
CC   cells themselves. Between the two RTKs, the juxtamembrane region
CC   is the region of most extensive sequence divergence.
SQ   Sequence   3157 BP;    790 A;    802 C;    759 G;    806 T;

  Query Match           11.7%;  Score 15;  DB 15;  Length 3157;
  Best Local Similarity 100.0%;  Pred. No. 3.30e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2643 tatcttgctgggcaa 2657
        |||||||||||||||
```



```
Qy    113 TATCTTGCTGGGCAA 127


RESULT   36
ID   Q27254 standard; DNA; 3372 BP.
AC   Q27254;
DT   02-FEB-1993  (first entry)
DE   Growth Factor Receptor Bound protein GRB-1.
KW   Tyrosine phophorylation; epidermal growth factor receptor; EGFR;
KW   src homology domain; SH2; SH3; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             43..2217
FT                   /*tag=  a
FT                   /product= GRB-1
PN   WO9213001-A.
PD   06-AUG-1992.
PF   17-JAN-1992; U00434.
PR   18-JAN-1991; US-643237.
PA   (UYNY ) UNIV NEW YORK STATE.
PI   Margolis BL, Schlessinger J, Skolnik EY;
DR   WPI; 92-284605/34.
DR   P-PSDB; R26060.
PT   Probe from tyrosine-phosphorylated portion of receptor tyrosine
PT   kinase - used for detection of proteins capable of binding to
PT   receptors, useful for e.g. identifying susceptibility to cancer
PT   and diabetes
PS   Claim 21; Fig 4; 86pp; English.
CC   The GRB-1 coding sequence was isolated from human brain stem lambda
CC   gt11 expression library by screening with tyrosine phosphorylated
CC   C-terminal tail of the EGF Receptor. An initial clone contained a
CC   900bp insert encoding an SH2 domain. Probes were derived from the
CC   partial clone which were then used to isolate clones containing
CC   insert sequences located 5' or 3' from the initial clone. The
CC   full-length coding sequence was obtained. It encodes a protein,
CC   designated GRB-1, which contains two SH2 domains and one SH3 domain.
CC   The protein is capable of binding to tyrosine-phosphorylated domains
CC   of receptor tyrosine kinases.
CC   See also Q27255.
SQ   Sequence   3372 BP;    1112 A;    661 C;    713 G;    886 T;

  Query Match           11.7%;  Score 15;  DB 4;  Length 3372;
  Best Local Similarity 100.0%;  Pred. No. 3.30e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    856 cttttcagattctca 870
          |||||||||||||||
Qy     63 CTTTTCAGATTCTCA 77


RESULT   37
```



```
ID   T03198 standard; DNA; 3372 BP.
AC   T03198;
DT   16-MAY-1996  (first entry)
DE   Human GRB-1 genomic DNA aequence.
KW   GRB-1; growth factor receptor bound; tyrosine kinase; regulation;
KW   cell growth; cellular metabolism; screening; signal transduction;
KW   cancer; diabetes; CORT technique; cloning of receptor targets; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             43..2217
FT                   /*tag=  a
FT                   /product= GRB-1
FT                   /note= "growth factor receptor bound-1"
PN   WO9524426-A1.
PD   14-SEP-1995.
PF   13-MAR-1995; U03385.
PR   11-MAR-1994; US-208887.
PA   (UYNY ) UNIV NEW YORK STATE.
PI   Margolis BL,  Schlessinger J,  Skolnik EY;
DR   WPI; 95-328235/42.
DR   P-PSDB; R85784.
PT   DNA encoding tyrosine kinase-binding proteins - used to screen
PT   agents capable of modulating cell growth or cellular metabolism
PS   Example 3;  Page 125-127;  215pp;  English.
CC   Using a new cloning technique, CORT (cloning of receptor targets)
CC   several new tyrosine kinase (TK) binding proteins were isolated. Growth
CC   factor receptor bound proteins GRB-1,GRB-2, GRB-3, GRB-4, GRB-7 and
CC   GRB-10 were isolated using this method. This sequence encodes GRB-1.
CC   The proteins bind to a tyrosine-phosphorylated domain of a eukaryotic
CC   TK. GRB proteins can be used for screening agents which are capable
CC   of modulating cell growth that occurs via signal transduction through
CC   TKs. Such agents can be used to prevent or inhibit cell growth or to
CC   counteract tumour development. GRB proteins are also useful for
CC   identifying susceptibility to diseases associated with alterations in
CC   cellular metabolism mediated by TK pathways e.g. cancer and diabetes.
SQ   Sequence   3372 BP;    1110 A;    661 C;    716 G;    885 T;

  Query Match           11.7%;  Score 15;  DB 19;  Length 3372;
  Best Local Similarity 100.0%;  Pred. No. 3.30e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    856 cttttcagattctca 870
          |||||||||||||||
Qy     63 CTTTTCAGATTCTCA 77


RESULT   38
ID   N50094 standard; DNA; 3852 BP.
AC   N50094;
DT   10-OCT-1991  (first entry)
DE   DNA encoding Factor-VIIIC derivative.
```

```
KW   Factor-VIIIC; ds; blood-clotting.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   mat_peptide     1..3852
FT                   /*tag= a
FT                   /label= human Factor-VIIIC derivative
PN   EP-150735-A.
PD   07-AUG-1985.
PF   11-JAN-1985; 100223.
PR   12-JAN-1984; US-570062.
PR   26-OCT-1984; US-664919.
PA   (CHIR-) CHIRON CORP.
PA   (NORD-) NORDISK GENTOFTE.
PI   Kuo G, Masiarz F, Truett M, Valenzuela P, Rasmussen M,
PI   Pavaloro J;
DR   WPI; 85-191681/32.
PT   Prodn. of human Factor VIII C or its precursors or sub-units - by
PT   using recombinant DNA techniques with Factor VIII C gene
PT   expressed in host.
PS   Disclosure; Appendix B; 58pp; English.
CC   The sequence encodes a derivative of human Factor-VIIIC, which is
CC   useful as an anticoagulant for the treatment of haemophilia.
SQ   Sequence   3852 BP;    1334 A;    795 C;    778 G;    945 T;

  Query Match           11.7%;  Score 15;  DB 3;  Length 3852;
  Best Local Similarity 100.0%;  Pred. No. 3.30e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1853 agcagaattttgtca 1867
          |||||||||||||||
Qy     16 AGCAGAATTTTGTCA 30


RESULT   39
ID   N90844 standard;  DNA; 4698 BP.
AC   N90844;
DT   24-JAN-1990  (first entry)
DE   Tobacco basic beta-1,3 glucanase gene.
KW   Tobacco basic beta-1,3 glucanase; pBSGluc39.3; herbicide resistance;
KW   transgenic plants.
OS   Nicotiana tabaccum cv. Xanthi.
PN   EP-332104-A.
PD   13-SEP-1989.
PF   06-MAR-1989; 103888.
PR   08-MAR-1988; US-165667.
PA   (CIBA) Ciba Geigy AG.
PI   Ryals J, Montoya A, Harms C, Duesing J, Sperisen C, Meins F,
PI   Payne G;
DR   WPI; 89-265342/37.
PT   Chemical induction of cloned genes in plants - DNA sequence and vectors
PT   and processes for inducible expression leading to herbicide and insect
```



```
PT   resistance.
PS   Claim 80; page 75-78; 117pp; English.
CC   Sequence is the tobacco basic beta-1,3-glucanase gene which is contained
CC   in clone pBSGluc39.3. It is a chemically inducible gene used to regulate
CC   transcription of an associated DNA sequence in plant tissue.
CC   See also N90367, N90368,  N90841-N90846, and P91384-P91386.
SQ   Sequence   4698 BP;    1508 A;    857 C;    785 G;    1546 T;    2 Others;

  Query Match           11.7%;  Score 15;  DB 1;  Length 4698;
  Best Local Similarity 100.0%;  Pred. No. 3.30e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   3504 tgagatgagacactt 3518
          |||||||||||||||
Cp     92 TGAGATGAGACACTT 78


RESULT   40
ID   Q03254 standard; DNA; 4699 BP.
AC   Q03254;
DT   25-JUL-1990  (first entry)
DE   Sequence of clone BSGluc39.3 encoding beta-1,3-glucanase from genomic
DE   sequence.
KW   Beta-1,3-glucanase; papermaking; polysaccharide hydrolase;
KW   ds.
OS   Nicotiana sp.
PN   EP-353191-A.
PD   31-JAN-1990.
PF   20-JUL-1989; 810555.
PR   29-JUL-1988; US-353312.
PA   (CIBA) Ciba Geigy AG.
PI   Shinshi H, Wenzler H, Hofsteenge J, Ryals J, Sperisen C;
DR   WPI; 90-031717/05.
PT   Recombinant DNA encoding beta-1,3-glucanase polypeptide(s) -
PT   useful in plant protection and biomass conversion pref. isolated
PT   from tobacco.
PS   Disclosure; Fig 9; 41pp; English.
CC   Beta-1,3-glucanase is involved in plant defence against pathogens and as
CC   it hydrolyses polyssaccharides, useful in biomass conversion or paper
CC   making. The clones allow for transfer of the gene to other species and
CC   control of gene expression.
SQ   Sequence   4699 BP;    1508 A;    857 C;    785 G;    1547 T;    2 Others;

  Query Match           11.7%;  Score 15;  DB 1;  Length 4699;
  Best Local Similarity 100.0%;  Pred. No. 3.30e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   3505 tgagatgagacactt 3519
          |||||||||||||||
Cp     92 TGAGATGAGACACTT 78
```



```
RESULT   41
ID   T45423 standard; cDNA; 5912 BP.
AC   T45423;
DT   30-MAR-1998  (first entry)
DE   Partial ATM gene transcript, clone 7-9.
KW   Ataxia-telangiectasia; A-T; mutated; ATM; 11q22-23; signal transduction;
KW   DNA damage; cell cycle control; screening; gene therapy; catalytic;
KW   phosphatidylinositol-3 kinase; PI-3; cancer predisposition; carrier;
KW   untranslated region; UTR; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   5'UTR           1..259
FT                   /*tag= a
FT   CDS             260..5386
FT                   /*tag= b
FT                   /product= PI_3-kinase_homologue
FT   3'UTR           5387..5912
FT                   /*tag= c
PN   WO9636695-A1.
PD   21-NOV-1996.
PF   16-MAY-1996; U07040.
PR   28-JUL-1995; US-508836.
PR   16-MAY-1995; US-441822.
PR   21-JUN-1995; US-493092.
PA   (UYRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI   Collins FS, Shiloh Y, Tagle DA;
DR   WPI; 97-012074/01.
PT   New gene ATM implicated in ataxia-telangiectasia and related protein
PT   - useful in screening methods, partic. for identifying disease
PT   carriers.
PS   Example 2; Page 68-71; 127pp; English.
CC   This is a 5.9 kb cDNA clone, designated 7-9, obtained from a fibroblast
CC   cDNA library. This sequence represents only part of the ATM gene
CC   transcript. Mutations in the ATM gene cause ataxia-telangiectasia (A-T),
CC   a progressive genetic disorder affecting the central nervous and immune
CC   systems. The ATM gene, located at chromosome 11q22-23, is probably
CC   involved in a novel signal transduction system that links DNA damage
CC   surveillance to cell cycle control. The ATM gene product (W07655) has a
CC   highly conserved C-terminal region showing high sequence homology to the
CC   catalytic domain of phosphatidylinositol-3 kinases. A-T mutations affect
CC   a variety of tissues and lead to cancer predisposition. Identification of
CC   A-T carriers, by analysis at nucleic acid or protein levels, allows
CC   better supervision and treatment of such subjects who are at increased
CC   risk of developing cancer and are particularly sensitive to radiation.
CC   The transgenic animals and transformed cells are useful as models of the
CC   human disease. Also viral vectors expressing the ATM protein can be used
CC   in gene therapy of A-T.
SQ   Sequence   5912 BP;    1885 A;    1085 C;    1281 G;    1661 T;
```



```
  Query Match           11.7%;  Score 15;  DB 37;  Length 5912;
  Best Local Similarity 100.0%;  Pred. No. 3.30e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   4271 tttaaagcagaattt 4285
          |||||||||||||||
Qy     11 TTTAAAGCAGAATTT 25


RESULT   42
ID   T43519 standard; cDNA; 5912 BP.
AC   T43519;
DT   13-AUG-1997  (first entry)
DE   ATM gene clone 7-9.
KW   ATM gene; ataxia-telangiectasia; progressive genetic disorder; antibody;
KW   central nervous system; immune system; chromosomal instability; therapy;
KW   cancer predisposition; radiation sensitivity; cell cycle abnormality;
KW   multi-system disease; autosomal recessive; cerebellar ataxia; cerebellum;
KW   general motor dysfunction; Purkinje cell; oculocutaneous telangiectasia;
KW   blood vessel; bulbar conjunctiva; facial skin; A-T; ss.
OS   Homo sapiens.
PN   WO9636691-A1.
PD   21-NOV-1996.
PF   16-MAY-1996; U07025.
PR   16-MAY-1995; US-441822.
PR   08-APR-1996; US-629001.
PA   (KOHN/) KOHN K I.
PA   (UYRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.
PI   Shiloh Y;
DR   WPI; 97-012070/01.
PT   New isolated ataxia-telangiectasia gene - used to develop prods. for
PT   the study, diagnosis and treatment of ataxia-telangiectasia.
PS   Example 2; Page 83-86; 153pp; English.
CC   This sequence represents clone 7-9 of the ATM gene, isolated from a
CC   fibroblast cDNA library. The full length wild type ATM open reading frame
CC   sequence is shown in T43497. T43444-T43496 represent exons of the ATM
CC   gene of the invention, which is located in the genome at chromosome
CC   segment 11q22-23. Ataxia-telangiectasia (A-T) is caused by mutations,
CC   insertions, or deletions in the coding region of the ATM gene. A-T is a
CC   progressive genetic disorder affecting the central nervous and immune
CC   systems. A-T involves chromosomal instability, cancer predisposition,
CC   radiation sensitivity, and cell cycle abnormalities. A-T is a
CC   multi-system disease inherited in an autosomal recessive manner.
CC   Cerebellar ataxia that gradually develops into general motor dysfunction
CC   is the first clinical hallmark of A-T, and results from progressive loss
CC   of Purkinje cells in the cerebellum. Oculocutaneous telangiectasia
CC   (dilation of blood vessels) develops in the bulbar conjunctiva and facial
CC   skin, and is later accompanied by graying of the hair and atrophic
CC   changes in the skin. The co-occurrence of cerebellar ataxia and
CC   telangiectases in the conjunctivae and occasionally on the facial skin
CC   (the second early hallmark of A-T) usually establishes the differential
```

```
CC   diagnosis of A-T from other cerebellar ataxias. The gene, can be used in
CC   methods for detecting carriers of a defective gene that causes A-T. The
CC   gene can also be used to generate antibodies. The methods and antibodies
CC   can be used in the study, diagnosis and therapy of A-T.
SQ   Sequence   5912 BP;    1885 A;    1085 C;    1281 G;    1661 T;

  Query Match           11.7%;  Score 15;  DB 31;  Length 5912;
  Best Local Similarity 100.0%;  Pred. No. 3.30e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   4271 tttaaagcagaattt 4285
          |||||||||||||||
Qy     11 TTTAAAGCAGAATTT 25


RESULT   43
ID   T03571 standard; cDNA; 6300 BP.
AC   T03571;
DT   13-MAY-1996  (first entry)
DE   Factor-VIII cDNA.
KW   Factor-VIII; blood clotting; haemophilia A; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds              7..6300
FT                    /*tag= a
FT   signal_peptide   7..63
FT                    /*tag= b
FT   mat_peptide      64..6300
FT                    /*tag= c
FT                    /note= "the sequence is incomplete since
FT                    it encodes only amino acids 1-2079
FT                    of mature Factor-VIII"
PN   EP-690126-A1.
PD   03-JAN-1996.
PF   13-JUN-1988; 110468.
PR   12-JUN-1987; EP-201121.
PA   (IMMO ) IMMUNO AG.
PI   Pannekoek H,  Van Leen RW,  Van Ooyen AJJ,  Verbeet MP;
DR   WPI; 96-051261/06.
DR   P-PSDB; R86863.
PT   New deletion mutant of factor VIII - useful for treating factor VIII
PT   deficiency, esp. haemophilia A
PS   Example 1; Fig 1; 31pp; English.
CC   A human liver cDNA sequence (T03571) in expression vector pCLB89
CC   encodes the Factor-VIII precursor (see R86863).  The cDNA was
CC   subjected to loop-out mutagenesis and used in the construction of
CC   vectors utilised in the transformation of e.g. COS, CHO and C127
CC   mammalian cells, or Bacillus or Kluyveromyces microbial cells,
CC   for prodn. of a Factor-VIII derivative useful in the treatment
CC   of haemophilia A.  The Factor-VIII derivative has the formula
CC   NR-LR-CR, where NR is amino acids 1-712 of Factor-VIII, LR is a
```



```
CC   short peptide (R86864), and CR is amino amino acids 1638-2332 of
CC   Factor-VIII.
SQ   Sequence   6300 BP;    2026 A;    1340 C;    1323 G;    1611 T;

  Query Match           11.7%;  Score 15;  DB 18;  Length 6300;
  Best Local Similarity 100.0%;  Pred. No. 3.30e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   3971 agcagaattttgtca 3985
          |||||||||||||||
Qy     16 AGCAGAATTTTGTCA 30


RESULT   44
ID   T79728 standard; DNA; 6763 BP.
AC   T79728;
DT   11-DEC-1997  (first entry)
DE   Beta-lactoglobulin cloning vector pMAD6.
KW   Protein C; transgenic animal; sheep; rabbit; cattle; goat; milk;
KW   blood clotting; anticoagulant; beta-lactoglobulin; pMAD6; vector;
KW   ss.
OS   Chimeric Ovis aries;
OS   Chimeric synthetic.
PN   WO9720043-A1.
PD   05-JUN-1997.
PF   26-NOV-1996; U18866.
PR   13-JUN-1996; US-019692.
PR   30-NOV-1995; US-565074.
PA   (PPLT-) PPL THERAPEUTICS.
PA   (SYMO ) SYMOGENETICS INC.
PI   Cottingham I,  Foster DC,  Garner I,  Prunkard DE;
PI   Sprecher CA,  Temperley SM;
DR   WPI; 97-310599/28.
PT   Production of protein C in transgenic animal - useful for high
PT   quantity protein C production with therapeutic value
PS   Example 1; Page 80-85; 99pp; English.
CC   This DNA sequence comprises plasmid pMAD6, a pUC18-based vector
CC   containing ovine beta-lactoglobulin gene sequences (see T79725).
CC   Nucleic acids encoding two-chain cleavage site-modified human
CC   protein C (see T79723-24) can be cloned into the vector so that
CC   the protein C gene is under the control of the beta-lactoglobulin
CC   promoter.  This construct can be used in claimed methods for the
CC   production of human protein C in the milk of transgenic animals.
CC   Modification of the protein C two-chain cleavage site improves the
CC   maturation of single-chain protein C to two-chain protein C.
SQ   Sequence   6763 BP;    1473 A;    1817 C;    1709 G;    1764 T;

  Query Match           11.7%;  Score 15;  DB 34;  Length 6763;
  Best Local Similarity 100.0%;  Pred. No. 3.30e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```



```
Db    947 gaattttgtcaaaag 961
          |||||||||||||||
Qy     20 GAATTTTGTCAAAAG 34


RESULT   45
ID   T51357 standard; DNA; 7053 BP.
AC   T51357;
DT   17-NOV-1997  (first entry)
DE   Factor VIII:C coding sequence.
KW   Factor VIII:C; analogue; glycoprotein; blood coagulation cascade;
KW   fibrinogen; fibrin clot; haemostasis; haemophilia A; bleeding diathesis;
KW   plasma protease; thrombin; immunogen; antibody; haemophiliac; therapy;
KW   proteolytic cleavage; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   signal_peptide   1..57
FT                    /*tag= a
FT   mat_peptide      58..7053
FT                    /*tag= b
FT                    /product= mature native Factor VIII:C
FT                    /note= "no stop codon given"
PN   WO9703195-A1.
PD   30-JAN-1997.
PF   09-JUL-1996; U11444.
PR   11-JUL-1995; US-001025.
PA   (CHIR ) CHIRON CORP.
PI   Cohen FE,  Hung DT,  Innis M;
DR   WPI; 97-119050/11.
DR   P-PSDB; W11329.
PT   Factor VIII:C analog modified adjacent to a non-activating Arg
PT   residue - used in the treatment of haemophiliacs, by improvement of
PT   haemostasis
PS   Disclosure; Fig 1; 90pp; English.
CC   This sequence represents the coding sequence for native Factor VIII:C.
CC   Factor VIII:C is a large glycoprotein that participates in the blood
CC   coagulation cascade that ultimately converts soluble fibrinogen to
CC   insoluble fibrin clot, effecting haemostasis. A deficiency in Factor
CC   VIII:C is responsible for haemophilia A, which is an X-chromosome-linked
CC   inherited bleeding diathesis. Factor VIII:C is activated by plasma
CC   proteases, such as thrombin. During activation the mature polypeptide is
CC   cleaved to generate heavy and light chain fragments that are further
CC   cleaved. This sequence is mutated, using mutagenic primers, to produce
CC   the active Factor VIII:C analogues of the invention (such as W11330). The
CC   analogues comprise a native Factor VIII:C polypeptide modified at a site
CC   adjacent to a non-activating Arg residue so that a Arg-Pro or Pro-Arg
CC   dipeptide is created. Complexes of two or more of the analogues, nucleic
CC   acids and vectors encoding them may be used alone or in conjunction with
CC   each other, for the prevention or treatment of active Factor VIII:C
CC   deficiency in a mammal. The analogues may be used as immunogens to raise
CC   antibodies, and in the treatment of haemophiliacs, by improvement of
```



```
CC   haemostasis. The analogues are resistant to proteolytic cleavage and
CC   display increased plasma half-life. They may be administered at lower
CC   dosages and by different modes of administration.
SQ   Sequence   7053 BP;    2232 A;    1518 C;    1489 G;    1814 T;

  Query Match           11.7%;  Score 15;  DB 33;  Length 7053;
  Best Local Similarity 100.0%;  Pred. No. 3.30e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   3965 agcagaattttgtca 3979
          |||||||||||||||
Qy     16 AGCAGAATTTTGTCA 30


RESULT   46
ID   Q66615 standard; cDNA; 7056 BP.
AC   Q66615;
DT   10-NOV-1994  (first entry)
DE   Sequence of human factor VIII.
KW   Factor VIII; haemostasis; haemophilia A; clotting cascade;
KW   fibrinogen; fibrin; thrombin; proteolytic enzyme; co-factor; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds              1..7053
FT                    /*tag= a
PN   WO9411503-A.
PD   26-MAY-1994.
PF   01-OCT-1993; U09438.
PR   13-NOV-1992; US-976086.
PR   14-SEP-1993; US-121202.
PA   (GEMY ) GENETICS INST INC.
PI   Kaufman RJ,  Pittman D,  Rehemtulla A,  Wozney JM;
DR   WPI; 94-183504/22.
DR   P-PSDB; R55352.
PT   Nucleic acid encoding porcine factor VIII - used to obtain
PT   porcine and human-porcine chimeric factor VIII for treating
PT   haemophilia
PS   Disclosure; Page 27-36; 61pp; English.
CC   Prepn. of human factor VIII cDNA has been set forth in detail, e.g.,
CC   US Patent No. 4,757,006 issued July 12, 1988 and in Toole et al.,
CC   Nature 312:312(1984). A recombinant clone contg. the nucleotide
CC   sequence in Q66615, designated as pSP64-VIII, is on deposit at the
CC   ATTC under Accession No. ATCC 39812. Chimeric forms of factor VIII
CC   include those where various domains of the human factor VIII have
CC   been replaced, in whole or in part, by analogous porcine factor VIII
CC   domains, and include, chimeric forms where the A1 and/or A2 domains,
CC   in whole or in part, of the human factor VIII sequence have been
CC   replaced, in whole or in part, by the A1 and/or the A2 domains of
CC   porcine factor VIII. Specifically provided are chimeric factor VIII
CC   sequences comprising the A1, A2, A3, B, C1 and C2 human domains  as
CC   set forth in Q66615, where the A1 and/or A2 domains, as well as
```

```
CC   other segments, such as the regions corresp. to the AA numbers
CC   336-372, 336-740, 372-740, 700-740 and combinations of these
CC   regions have been replaced in whole or in part with porcine factor
CC   VIII sequences as set forth in Q66616 and R55353.
SQ   Sequence   7056 BP;    2232 A;    1517 C;    1491 G;    1816 T;

  Query Match           11.7%;  Score 15;  DB 10;  Length 7056;
  Best Local Similarity 100.0%;  Pred. No. 3.30e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   3965 agcagaattttgtca 3979
          ||||||||||||||||
Qy     16 AGCAGAATTTTGTCA 30


RESULT   47
ID   N81543 standard; DNA; 7059 BP.
AC   N81543;
DT   04-DEC-1990  (first entry)
DE   Human Factor VIII-C analog having exon 4 deleted.
KW   Human Factor VIII-C analog; exon deletion; coagulation disorders;
KW   haemophilia; ss.
PN   EP-265778-A.
PD   04-MAY-1988.
PF   14-OCT-1987; 115043.
PR   15-OCT-1986; US-919153.
PA   (RORE) Rorer Int Overseas.
PI   Sarver N, Drohan W.;
DR   WPI; 88-120930/18.
PT   Human Factor VIII-C analogue free of other proteins - produced from
PT   recombinant DNA and used in the treatment of coagulation disorders.
PS   Claim 2; Page 13; 42pp; English.
CC   The genetically engineered analogue can provide a dependable and
CC   readily available therapeutic agent to be used in the treatment of
CC   haemophilia and coagulation disorders in humans or animals.
CC   See also N81543-45.
SQ   Sequence   7059 BP;    2207 A;    1543 C;    1483 G;    1826 T;

  Query Match           11.7%;  Score 15;  DB 1;  Length 7059;
  Best Local Similarity 100.0%;  Pred. No. 3.30e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   3752 agcagaattttgtca 3766
          ||||||||||||||||
Qy     16 AGCAGAATTTTGTCA 30


RESULT   48
ID   N50375 standard; DNA; 7440 BP.
AC   N50375;
DT   23-NOV-1991  (first entry)
```



```
DE   DNA sequence encoding human antihaemophilic factor.
KW   Human antihaemophilic factor; haemophilia; as DNA; factor VIIIC.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             159..7214
FT                   /*tag=  a
PN   WO8501961-A.
PD   09-MAY-1985.
PF   12-OCT-1984; U01641.
PR   28-OCT-1983; US-546650.
PR   24-OCT-1984; US-644036.
PA   (GENE-) Genetics Institute.
PI   Toole JJ Jr;
DR   WPI; 85-122479/20.
DR   P-PSDB; P50319.
PT   Prodn. of human and porcine factor VIIIC - by using recombinant DNA
PT   techniques for their cellular prodn.
PS   Claim 36; page 53; 71pp; English.
CC   The DNA encodes human antihaemophilic factor (AHF). Human AHF is
CC   useful for the treatment of haemophilia.
SQ   Sequence   7440 BP;    2318 A;    1689 C;    1497 G;    1936 T;

  Query Match           11.7%;  Score 15;  DB 3;  Length 7440;
  Best Local Similarity 100.0%;  Pred. No. 3.30e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   4123 agcagaattttgtca 4137
          ||||||||||||||||
Qy     16 AGCAGAATTTTGTCA 30


RESULT   49
ID   N60689 standard; DNA; 7440 BP.
AC   N60689;
DT   26-JUN-1991  (first entry)
DE   Sequence encoding human factor VIII C.
KW   Procoagulant; haemophilia A; therapy; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             159..7214
FT                   /*tag=  a
PN   WO8606101-A.
PD   23-OCT-1986.
PF   11-APR-1986; U00774.
PR   12-APR-1985; US-725350.
PR   11-APR-1986; US-010085.
PR   11-APR-1986; WO-U00774.
PA   (GENE-) GENETICS INST INC.
PA   (TOOL/) TOOLE J J.
PI   Toole JJ;
DR   WPI; 86-291668/44.
```



```
DR   P-PSDB; P60741.
PT   Protein having pro-coagulant activity - obtd. from truncated
PT   version of DNA sequence which expresses complete human factor
PT   VIII-C
PS   Claim 5; Table 1,Pages 3-9; 30pp; English.
CC   The procoagulant polypeptides of the invention are truncated
CC   versions of human factor VIII C, having a portion of AAs 760-1708
CC   (see P60741) removed. Specifically claimed polypeptides have the
CC   following deletions: AAs 1001-1581; AAs 779-1658; and AAs 779-1693;
CC   and the DNA encoding them. The procoagulant proteins lessen the
CC   likelihood of developing antibodies to the procoagulants and
CC   facilitate the treatment of patients in need of procoagulant therapy
CC   e.g. Haemophilia A treatment.
SQ   Sequence   7440 BP;    2318 A;    1624 C;    1560 G;    1936 T;

  Query Match           11.7%;  Score 15;  DB 2;  Length 7440;
  Best Local Similarity 100.0%;  Pred. No. 3.30e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   4123 agcagaattttgtca 4137
          ||||||||||||||||
Qy     16 AGCAGAATTTTGTCA 30


RESULT   50
ID   N81096 standard; cDNA; 8241 BP.
AC   N81096;
DT   29-OCT-1990  (first entry)
DE   cDNA sequence encoding human Factor VIII signal sequence and mature
DE   protein on plasmid pCLB89
KW   Haemophilia A; bloodclotting; treatment; diagnosis; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             7..63
FT                   /*tag=  a
FT                   /product=Signal peptide
FT   mat_peptide     64..7062
FT                   /*tag=  b
PN   EP-294910-A.
PD   14-DEC-1988.
PF   13-JUN-1988; 201209.
PR   12-JUN-1987; EP-201121.
PR   3-JUN-1988; EP-201209.
PA   (KONN) Gist-Brocades NV.
PI   van Ooyen AJJ, Pannekoek H, Verbeet MP, van Leen RW;
DR   WPI; 88-355361/50.
DR   P-PSDB; P80659.
PT   Proteins having Factor VIII activity -
PT   comprising deletion mutant proteins of Factor VIII in which
PT   central region has been deleted
PS   Example; Fig 1; 37pp; English.
```



```
CC   The isolation of Factor VIII mRNA from human liver, and the preparation,
CC   purification and identification of cDNA and its assembly in the plasmid
CC   pEP121 resulting in plasmid pCLB89 have been described in patent
CC   application EP 0253455. The Factor VIII polypeptides of the invention
CC   include deletion mutant proteins of Factor VIII in which the central
CC   region or "B domain" as well as a portion of the 92 kD region has been
CC   deleted. The polypeptides have enhanced Factor VIII activity and/or
CC   decreased immunogenicity and can be used for the treatment of
CC   Haemophilia A. They can also be used to prepare antibodies. The Abs.
CC   can be used to determine the concn. of Factor VIII in a body fluid.
SQ   Sequence   8241 BP;    2625 A;    1759 C;    1703 G;    2154 T;

  Query Match           11.7%;  Score 15;  DB 1;  Length 8241;
  Best Local Similarity 100.0%;  Pred. No. 3.30e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   3971 agcagaattttgtca 3985
          ||||||||||||||||
Qy     16 AGCAGAATTTTGTCA 30


Search completed: Tue Dec 22 21:07:10 1998
Job time : 145 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Tue Dec 22 20:50:41 1998; MasPar time 222.11 Seconds
1270.032 Million cell updates/sec

Tabular output not generated.

Title: >US-08-530-112A-1320
Description: (1-128) from US08530112A.seq
Perfect Score: 128
N.A. Sequence: 1 GATCAGAATTTTTAAAGCAG.....AGAATATCTTGCTGGGCAAA 128
Comp: CTAGTCTTAAAAATTTCGTC.....TCTTATAGAACGACCCGTTT

Scoring table: TABLE jmetric
Gap 60

Nmatch STD : Dbase 0; Query 0

Searched: 567134 seqs, 1101898692 bases x 2

Post-processing: Minimum Match 0%
Listing first 1000 summaries

Database: embl55
1:em_ba 2:em_fun 3:em_htg 4:em_hum1 5:em_hum2 6:em_in
7:em_om 8:em_or 9:em_ov 10:em_pat 11:em_ph 12:em_pl
13:em_ro 14:em_vi
Database: genbankl07
15:gb_ba1 16:gb_ba2 17:gb_htg 18:gb_in 19:gb_om 20:gb_ov
21:gb_pat 22:gb_ph 23:gb_pl1 24:gb_pl2 25:gb_pr1
26:gb_pr2 27:gb_pr3 28:gb_ro 29:gb_st 30:gb_sts 31:gb_sy
32:gb_un 33:gb_vi

Statistics: Mean 7.655; Variance 3.175; scale 2.411

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	Query Match	Length	DB	ID	Description	Pred. No.
1		20	15.6	196475	17	AC005386	*** SEQUENCING IN PROG	3.65e-01
2		19	14.8	1523	21	I70241	Sequence 25 from paten	1.71e+00
3		19	14.8	1523	21	I12338	Sequence 25 from paten	1.71e+00
4		19	14.8	1645	25	HUMSRD5AP	H.sapiens steroid redu	1.71e+00
5	c	19	14.8	36599	17	CEW04A8	*** SEQUENCING IN PROG	1.71e+00
6		19	14.8	72245	26	HS93C23	Human DNA sequence fro	1.71e+00
7	c	19	14.8	91286	18	CEY102A5C	Caenorhabditis elegans	1.71e+00
8		19	14.8	116215	26	AC004130	Homo sapiens BAC clone	1.71e+00
9		19	14.8	253516	17	CEY6B3	Caenorhabditis elegans	1.71e+00
10	c	18	14.1	552	25	HSPA20F7	H.sapiens flow-sorted	7.65e+00
11		18	14.1	2286	26	U92952	Trachypithecus francoi	7.65e+00
12	c	18	14.1	2370	28	RSU08214	Rattus sp. DNA binding	7.65e+00
13	c	18	14.1	2509	26	HSTMPO1	Human thymopoietin (TM	7.65e+00
14	c	18	14.1	4377	24	ATU87834	Arabidopsis thaliana z	7.65e+00
15		18	14.1	21494	18	CEY17D7A	Caenorhabditis elegans	7.65e+00
16	c	18	14.1	32784	18	CEF08G5	Caenorhabditis elegans	7.65e+00
17		18	14.1	40753	17	AC005715	*** SEQUENCING IN PROG	7.65e+00
18		18	14.1	45022	18	CEC34D1	Caenorhabditis elegans	7.65e+00
19	c	18	14.1	69060	26	AC005196	Homo sapiens chromosom	7.65e+00
20	c	18	14.1	71153	26	HUAE000662	Homo sapiens T-cell re	7.65e+00
21	c	18	14.1	97630	26	HUMTCRADCV	Human Tcr-C-delta gene	7.65e+00
22		18	14.1	112690	17	HS39B17	Human DNA sequence ***	7.65e+00
23		18	14.1	148583	17	CEY17D7	Caenorhabditis elegans	7.65e+00
24	c	18	14.1	161824	17	AC005061	*** SEQUENCING IN PROG	7.65e+00
25		18	14.1	174045	25	HS107M3	Human DNA sequence fro	7.65e+00
26		18	14.1	179793	17	AC004925	*** SEQUENCING IN PROG	7.65e+00
27		18	14.1	197273	26	AC002461	Human BAC clone RG204I	7.65e+00
28		18	14.1	236157	17	AC005703	*** SEQUENCING IN PROG	7.65e+00
29	c	17	13.3	1279	23	LETXETB2	L.esculentum tXET-B2 m	3.26e+01
30	c	17	13.3	1289	21	A44631	Sequence 7 from Patent	3.26e+01
31		17	13.3	1794	18	DMFUGENE	D.melanogaster fused g	3.26e+01
32	c	17	13.3	2908	18	MIU09180	Meloidogyne incognita	3.26e+01
33	c	17	13.3	4365	26	AB014563	Homo sapiens mRNA for	3.26e+01
34	c	17	13.3	4937	23	ATCYP450D	A.thaliana gene cytoch	3.26e+01
35		17	13.3	5025	18	DMFUSED	D.melanogaster fused g	3.26e+01
36		17	13.3	5025	18	DROFUSEDST	Drosophila melanogaste	3.26e+01
37		17	13.3	10118	15	BSUEROP	B.subtilis urease oper	3.26e+01
38		17	13.3	23952	18	CEF29D10	Caenorhabditis elegans	3.26e+01
39	c	17	13.3	23960	18	CEC54C8	Caenorhabditis elegans	3.26e+01
40		17	13.3	25350	18	CEF14F11	Caenorhabditis elegans	3.26e+01
41		17	13.3	27102	18	CEF22B8	Caenorhabditis elegans	3.26e+01
42		17	13.3	38574	18	CELF56B3	Caenorhabditis elegans	3.26e+01

43	c	17	13.3	40475	18	CELD1009	Caenorhabditis elegans	3.26e+01
44	c	17	13.3	40697	26	AF039905	Homo sapiens cosmid Q9	3.26e+01
45	c	17	13.3	45034	26	AF020803	Homo sapiens cosmid Q3	3.26e+01
46	c	17	13.3	45911	24	AB017070	Arabidopsis thaliana g	3.26e+01
47		17	13.3	47554	18	CELD1007	Caenorhabditis elegans	3.26e+01
48	c	17	13.3	70319	26	AC002453	Human BAC clone GS440B	3.26e+01
49		17	13.3	84132	26	AC005348	Homo sapiens chromosom	3.26e+01
50		17	13.3	87835	23	AB005237	Arabidopsis thaliana g	3.26e+01
51		17	13.3	87869	26	HS286B10	Homo sapiens DNA seque	3.26e+01
52	c	17	13.3	89603	27	HSU63834	Human KIT protein and	3.26e+01
53		17	13.3	94319	26	HS341I10	Human DNA sequence fro	3.26e+01
54		17	13.3	94386	26	HS452M16	Human DNA sequence ***	3.26e+01
55	c	17	13.3	100000	26	AP000043	Homo sapiens genomic D	3.26e+01
56	c	17	13.3	106988	26	AC001226	Genomic sequence from	3.26e+01
57	c	17	13.3	119453	26	AC004006	Human PAC clone DJ0740	3.26e+01
58	c	17	13.3	135448	17	AC004900	*** SEQUENCING IN PROG	3.26e+01
59		17	13.3	140679	17	HS246H3	Human DNA sequence ***	3.26e+01
60		17	13.3	144676	26	HS390013	Human DNA sequence fro	3.26e+01
61		17	13.3	148883	25	HS358H7	Human DNA sequence fro	3.26e+01
62		17	13.3	177328	17	HS462D8	Human DNA sequence ***	3.26e+01
63	c	17	13.3	179544	26	AF015720	Homo sapiens chromosom	3.26e+01
64	c	17	13.3	206214	17	AC005013	*** SEQUENCING IN PROG	3.26e+01
65		17	13.3	211468	17	CEY1B3	Caenorhabditis elegans	3.26e+01
66	c	17	13.3	212610	15	BSUB0019	Bacillus subtilis comp	3.26e+01
67	c	17	13.3	215807	17	AC005296	*** SEQUENCING IN PROG	3.26e+01
68	c	17	13.3	323000	26	HS229041	Homo sapiens 959 kb co	3.26e+01
69		16	12.5	125	21	I51029	Sequence 74 from paten	1.31e+02
70		16	12.5	125	21	I63651	Sequence 74 from paten	1.31e+02
71		16	12.5	125	21	I79948	Sequence 74 from paten	1.31e+02
72	c	16	12.5	388	26	AF021994	Homo sapiens ID:CLL121	1.31e+02
73	c	16	12.5	576	15	RPZ82522	R.prowazekii genomic D	1.31e+02
74	c	16	12.5	645	21	I25539	Sequence 2 from patent	1.31e+02
75	c	16	12.5	645	21	A25847	B.subtilis pcp gene.	1.31e+02
76		16	12.5	886	16	PAUB9892	Pseudomonas aeruginosa	1.31e+02
77	c	16	12.5	919	27	AF069970	Nasalis larvatus 12S r	1.31e+02
78		16	12.5	1149	20	U71242	Desmognathus wrighti 1	1.31e+02
79	c	16	12.5	1269	15	PAU16318	Pseudomonas aeruginosa	1.31e+02
80	c	16	12.5	1334	15	BSPCPG	B.subtilis pcp gene fo	1.31e+02
81		16	12.5	1490	15	MC060	M.capricolum DNA for C	1.31e+02
82		16	12.5	1811	19	OCMT2P	Rabbit MT-2 processed	1.31e+02
83		16	12.5	1811	19	RABMT2PS	Oryctolagus cuniculus	1.31e+02
84		16	12.5	1877	27	HSU33453	Human protease nexin-1	1.31e+02
85	c	16	12.5	1976	28	MMU56909	Mus musculus multiple	1.31e+02
86		16	12.5	2268	18	CEVT22E6R	Caenorhabditis elegans	1.31e+02
87		16	12.5	2322	15	BLU35855	Bacillus licheniformis	1.31e+02
88	c	16	12.5	2410	28	AF045254	Mus musculus clone 7a	1.31e+02
89	c	16	12.5	3071	15	LLLACZ	Lactococcus lactis lac	1.31e+02
90		16	12.5	3116	26	AC002216	Homo sapiens (subclone	1.31e+02
91	c	16	12.5	3173	28	AF045253	Mus musculus clone 4b	1.31e+02
92		16	12.5	3305	18	PTND7	P.tetraurelia ND7 gene	1.31e+02
93		16	12.5	3450	28	MUSGLYTNS	M.musculus alpha-1,3-g	1.31e+02

94	c	16	12.5	3546	23	SCYOL161C	S.cerevisiae chromosom	1.31e+02
95	c	16	12.5	3626	26	HS2NRAMP8	Homo sapiens natural r	1.31e+02
96	c	16	12.5	3745	28	AF045252	Mus musculus clone 2b	1.31e+02
97	c	16	12.5	4103	26	AB004857	Homo sapiens mRNA for	1.31e+02
98		16	12.5	4451	15	LLU60828	Lactococcus lactis gal	1.31e+02
99	c	16	12.5	4803	20	GGGTLYSO	G.gallus gene for goos	1.31e+02
100	c	16	12.5	4803	20	GG325GEN	G.gallus mRNA (clone 3	1.31e+02
101	c	16	12.5	5369	23	SCYOL160W	S.cerevisiae chromosom	1.31e+02
102	c	16	12.5	6566	26	AF043105	Homo sapiens glutathio	1.31e+02
103		16	12.5	6793	15	SCU28972	Spiroplasma citri orfa	1.31e+02
104	c	16	12.5	7172	16	AF082008	Lactococcus lactis NCD	1.31e+02
105		16	12.5	8213	25	HUMSFRS	Homo sapiens splicing	1.31e+02
106		16	12.5	8699	18	PFSC03018	Plasmodium falciparum	1.31e+02
107		16	12.5	9778	15	U67550	Methanococcus jannasch	1.31e+02
108	c	16	12.5	10130	23	SCYLL040C	S.cerevisiae chromosom	1.31e+02
109	c	16	12.5	10640	15	MPAE000029	Mycoplasma pneumoniae	1.31e+02
110		16	12.5	12215	23	AF001317	Saccharomyces cerevisi	1.31e+02
111		16	12.5	12451	18	CEZC302	Caenorhabditis elegans	1.31e+02
112		16	12.5	12622	15	MPAE000032	Mycoplasma pneumoniae	1.31e+02
113		16	12.5	14756	25	HSHFATPRO	H.sapiens mRNA for hFa	1.31e+02
114	c	16	12.5	14993	18	CEF57G9	Caenorhabditis elegans	1.31e+02
115	c	16	12.5	15284	15	U67462	Methanococcus jannasch	1.31e+02
116		16	12.5	16823	15	AE000793	Borrelia burgdorferi p	1.31e+02
117		16	12.5	16933	15	BBU43414	Borrelia burgdorferi 1	1.31e+02
118	c	16	12.5	17070	15	AE001153	Borrelia burgdorferi (	1.31e+02
119		16	12.5	19194	18	CELC29G2	Caenorhabditis elegans	1.31e+02
120		16	12.5	21548	15	BACYCB20	Bacillus subtilis DNA	1.31e+02
121		16	12.5	27207	18	PFSC03017	Plasmodium falciparum	1.31e+02
122	c	16	12.5	29355	18	CEH13N06	Caenorhabditis elegans	1.31e+02
123	c	16	12.5	30140	18	CEF09C8	Caenorhabditis elegans	1.31e+02
124		16	12.5	30213	18	CELB0302	Caenorhabditis elegans	1.31e+02
125	c	16	12.5	31906	18	CELEK813	Caenorhabditis elegans	1.31e+02
126		16	12.5	32266	18	CER07E5	Caenorhabditis elegans	1.31e+02
127	c	16	12.5	32412	18	CEF53B6	Caenorhabditis elegans	1.31e+02
128	c	16	12.5	33038	18	CELB0261	Caenorhabditis elegans	1.31e+02
129	c	16	12.5	34671	18	CELF01G12	Caenorhabditis elegans	1.31e+02
130		16	12.5	35000	15	AB000617	Bacillus subtilis geno	1.31e+02
131		16	12.5	36029	26	HUMU138C3	Homo sapiens cosmid U1	1.31e+02
132		16	12.5	36262	18	CELF28B3	Caenorhabditis elegans	1.31e+02
133		16	12.5	37121	18	CELT19B4	Caenorhabditis elegans	1.31e+02
134		16	12.5	37760	26	AC004174	Homo sapiens clone UWG	1.31e+02
135	c	16	12.5	37768	18	CELF53B3	Caenorhabditis elegans	1.31e+02
136		16	12.5	38162	26	AC004095	Human Cosmid g1248a143	1.31e+02
137		16	12.5	38897	18	CELF41C3	Caenorhabditis elegans	1.31e+02
138		16	12.5	40596	18	CELC36C5	Caenorhabditis elegans	1.31e+02
139	c	16	12.5	42730	18	CEC23R4	Caenorhabditis elegans	1.31e+02
140		16	12.5	43675	18	CELF10G7	Caenorhabditis elegans	1.31e+02
141	c	16	12.5	44072	18	DMC121E7	Drosophila melanogaste	1.31e+02
142	c	16	12.5	44175	26	AC000049	Homo sapiens Chromosom	1.31e+02
143		16	12.5	54632	24	AB015472	Arabidopsis thaliana g	1.31e+02
144		16	12.5	57421	18	AC004113	Drosophila melanogaste	1.31e+02

	145	16	12.5	73350	18	AC002493	Drosophila melanogaste 1.31e+02
c	146	16	12.5	76819	26	AC004043	Homo sapiens chromosom 1.31e+02
c	147	16	12.5	79227	26	HS148E22	Human DNA sequence fro 1.31e+02
	148	16	12.5	82073	26	HS172K10	Homo sapiens DNA seque 1.31e+02
	149	16	12.5	84544	23	AB005239	Arabidopsis thaliana g 1.31e+02
	150	16	12.5	93250	26	HS447B16	Human DNA sequence fro 1.31e+02
	151	16	12.5	94320	26	AC002564	Human BAC clone RG164L 1.31e+02
c	152	16	12.5	94665	17	AC004979	*** SEQUENCING IN PROG 1.31e+02
	153	16	12.5	101733	17	AC005518	*** SEQUENCING IN PROG 1.31e+02
	154	16	12.5	103566	26	AC003087	Human BAC clone RG119H 1.31e+02
	155	16	12.5	106346	26	AC004045	Homo sapiens chromosom 1.31e+02
c	156	16	12.5	109926	27	AC004055	Homo sapiens chromosom 1.31e+02
	157	16	12.5	110000	17	CEY105C5	Caenorhabditis elegans 1.31e+02
	158	16	12.5	110000	17	CEY38E10	Caenorhabditis elegans 1.31e+02
c	159	16	12.5	110384	25	HSAC000117	Human BAC clone RG062A 1.31e+02
	160	16	12.5	114144	23	ATU78721	Arabidopsis thaliana c 1.31e+02
	161	16	12.5	114847	27	HS257I20	Homo sapiens DNA seque 1.31e+02
	162	16	12.5	117132	17	AC004954	*** SEQUENCING IN PROG 1.31e+02
	163	16	12.5	118591	17	AC005024	*** SEQUENCING IN PROG 1.31e+02
	164	16	12.5	120832	17	HS836E8	Human DNA sequence *** 1.31e+02
	165	16	12.5	129079	17	AC004955	*** SEQUENCING IN PROG 1.31e+02
	166	16	12.5	130690	15	D90901	Synechocystis sp. PCC6 1.31e+02
	167	16	12.5	132783	26	AC000378	Human Chromosome 11 pa 1.31e+02
c	168	16	12.5	132790	26	HS167A14	Human DNA sequence fro 1.31e+02
c	169	16	12.5	132826	17	AC005627	*** SEQUENCING IN PROG 1.31e+02
c	170	16	12.5	133191	26	AC004027	Human PAC clone DJ0791 1.31e+02
c	171	16	12.5	133237	27	AC005550	Homo sapiens PAC clone 1.31e+02
	172	16	12.5	133925	26	AC004562	Homo sapiens chromosom 1.31e+02
	173	16	12.5	134929	27	AC004655	Homo sapiens Xp22-140- 1.31e+02
c	174	16	12.5	135959	17	AC005383	*** SEQUENCING IN PROG 1.31e+02
	175	16	12.5	137191	17	AC005476	*** SEQUENCING IN PROG 1.31e+02
	176	16	12.5	137947	27	HS45I4	Human DNA sequence *** 1.31e+02
	177	16	12.5	139958	27	HUAC002519	Human Chromosome 16 BA 1.31e+02
c	178	16	12.5	143751	26	AC004150	Homo sapiens BAC clone 1.31e+02
c	179	16	12.5	145576	27	AC003665	Homo sapiens chromosom 1.31e+02
c	180	16	12.5	149554	25	HSAC000111	Human BAC clone 068P20 1.31e+02
	181	16	12.5	150147	27	AC002543	Homo sapiens BAC clone 1.31e+02
	182	16	12.5	153568	26	AC002981	Homo sapiens Xp22 BAC 1.31e+02
c	183	16	12.5	154166	17	AC005479	*** SEQUENCING IN PROG 1.31e+02
c	184	16	12.5	156399	26	AC002456	Human BAC clone RG013L 1.31e+02
c	185	16	12.5	157872	17	HS272L16	Human DNA sequence *** 1.31e+02
c	186	16	12.5	160850	17	AC004957	*** SEQUENCING IN PROG 1.31e+02
c	187	16	12.5	160899	17	AC004869	*** SEQUENCING IN PROG 1.31e+02
c	188	16	12.5	161475	26	HS164C20	Homo sapiens DNA seque 1.31e+02
	189	16	12.5	162750	17	AC003061	*** SEQUENCING IN PROG 1.31e+02
c	190	16	12.5	164675	26	AC004384	Human Chromosome X clo 1.31e+02
c	191	16	12.5	165612	17	HS545L17	Human DNA sequence *** 1.31e+02
	192	16	12.5	166126	26	AC004142	Homo sapiens BAC clone 1.31e+02
	193	16	12.5	167575	17	HS531H16	Human DNA sequence *** 1.31e+02
	194	16	12.5	168238	17	AC005014	*** SEQUENCING IN PROG 1.31e+02
	195	16	12.5	170136	26	AC003082	Human BAC clone RG011J 1.31e+02

c	196	16	12.5	170658	26	AC004072	Human Chromosome X clo 1.31e+02
c	197	16	12.5	172502	17	AC005039	*** SEQUENCING IN PROG 1.31e+02
c	198	16	12.5	173494	17	HS662N3	Human DNA sequence *** 1.31e+02
	199	16	12.5	173767	17	AC005076	*** SEQUENCING IN PROG 1.31e+02
	200	16	12.5	176681	17	AC005696	*** SEQUENCING IN PROG 1.31e+02
c	201	16	12.5	176781	17	AC002989	*** SEQUENCING IN PROG 1.31e+02
	202	16	12.5	177311	26	AC004385	Human Chromosome X clo 1.31e+02
c	203	16	12.5	178698	17	HS67K17	Human DNA sequence *** 1.31e+02
c	204	16	12.5	180175	17	AC005105	*** SEQUENCING IN PROG 1.31e+02
c	205	16	12.5	181588	26	AC002487	Human BAC clone RG369K 1.31e+02
	206	16	12.5	187051	17	AC004688	*** SEQUENCING IN PROG 1.31e+02
	207	16	12.5	188362	27	AC005019	Homo sapiens BAC clone 1.31e+02
	208	16	12.5	191233	17	AC005409	*** SEQUENCING IN PROG 1.31e+02
c	209	16	12.5	196473	17	HS473B4	Human DNA sequence *** 1.31e+02
c	210	16	12.5	198456	17	AC003115	*** SEQUENCING IN PROG 1.31e+02
	211	16	12.5	217873	26	HUAC004382	Homo sapiens Chromosom 1.31e+02
	212	16	12.5	221160	15	BSUB0002	Bacillus subtilis comp 1.31e+02
c	213	16	12.5	235134	17	AC004849	*** SEQUENCING IN PROG 1.31e+02
c	214	16	12.5	310126	17	CEY48B6	Caenorhabditis elegans 1.31e+02
	215	16	12.5	326569	17	AC004912	*** SEQUENCING IN PROG 1.31e+02
	216	15	11.7	168	10	E09097	cDNA encoding tyrosine 4.98e+02
c	217	15	11.7	234	30	HSU31714	Human WI-616 primer de 4.98e+02
c	218	15	11.7	260	24	AF072647	Lycopersicon esculentu 4.98e+02
	219	15	11.7	293	28	RATRGLEC	Rat 45S precursor RNA 4.98e+02
	220	15	11.7	301	30	HS234VC1	H. sapiens (D5S495) DN 4.98e+02
	221	15	11.7	305	30	G25090	human STS EST253100. 4.98e+02
	222	15	11.7	305	29	CEXMTRRSSA	Cellana tramoserica mi 4.98e+02
c	223	15	11.7	316	30	G09389	human STS CHLC.GATA26A 4.98e+02
c	224	15	11.7	324	30	G02624	human STS WI-616. 4.98e+02
	225	15	11.7	338	15	BVU28074	Bartonella vinsonii ci 4.98e+02
c	226	15	11.7	338	23	YSCTRP1A	S.cerevisiae n-(5'-pho 4.98e+02
	227	15	11.7	340	25	HUMFVIII08	Human clotting factor 4.98e+02
	228	15	11.7	341	25	HUMFVIII09	Human clotting factor 4.98e+02
c	229	15	11.7	400	30	G14376	human STS SHGC-9045 cl 4.98e+02
c	230	15	11.7	400	30	G16458	human STS SHGC-13742 c 4.98e+02
	231	15	11.7	403	30	G41785	c1D5/CA Human Chromoso 4.98e+02
	232	15	11.7	403	30	G11919	human STS 1D5/CA clone 4.98e+02
	233	15	11.7	405	30	HS338WB5	H. sapiens (D15495) DN 4.98e+02
	234	15	11.7	413	30	G23918	human STS WI-15800. 4.98e+02
	235	15	11.7	424	18	DIRRG5SQ	D.immitis 5S rRNA gene 4.98e+02
c	236	15	11.7	438	26	AF048B48	Homo sapiens erythroid 4.98e+02
	237	15	11.7	447	30	G16216	human STS SHGC-3101 cl 4.98e+02
c	238	15	11.7	481	15	AF054517	Synechococcus PCC7002 4.98e+02
	239	15	11.7	491	30	G29557	human STS SHGC-33900, 4.98e+02
	240	15	11.7	520	23	ATSNRNA	A.thaliana At1 gene fo 4.98e+02
	241	15	11.7	543	13	AB010142	Mus musculus gene for 4.98e+02
c	242	15	11.7	604	25	HUMEV121	Human EV12 protein gen 4.98e+02
c	243	15	11.7	675	28	RNU53403	Rattus norvegicus chro 4.98e+02
c	244	15	11.7	686	26	D63345	Human DNA for selenopr 4.98e+02
	245	15	11.7	695	26	AG000528	Homo sapiens genomic D 4.98e+02
	246	15	11.7	695	18	AF010241	Caenorhabditis elegans 4.98e+02

c	247	15	11.7	698	20	AF053828	Melospiza melodia mito 4.98e+02
c	248	15	11.7	699	20	AF053831	Melospiza melodia mito 4.98e+02
c	249	15	11.7	700	20	AF053858	Melospiza melodia mito 4.98e+02
c	250	15	11.7	700	20	AF053841	Melospiza melodia mito 4.98e+02
c	251	15	11.7	700	20	AF053857	Melospiza melodia mito 4.98e+02
c	252	15	11.7	700	20	AF053866	Melospiza melodia mito 4.98e+02
c	253	15	11.7	700	20	AF053863	Melospiza melodia mito 4.98e+02
c	254	15	11.7	700	20	AF053871	Melospiza melodia mito 4.98e+02
c	255	15	11.7	700	20	AF053877	Melospiza melodia mito 4.98e+02
c	256	15	11.7	700	20	AF053864	Melospiza melodia mito 4.98e+02
c	257	15	11.7	700	20	AF053856	Melospiza melodia mito 4.98e+02
c	258	15	11.7	700	20	AF053833	Melospiza melodia mito 4.98e+02
c	259	15	11.7	700	20	AF053862	Melospiza melodia mito 4.98e+02
c	260	15	11.7	700	20	AF053836	Melospiza melodia mito 4.98e+02
c	261	15	11.7	700	20	AF053854	Melospiza melodia mito 4.98e+02
c	262	15	11.7	700	20	AF053849	Melospiza melodia mito 4.98e+02
c	263	15	11.7	700	20	AF053832	Melospiza melodia mito 4.98e+02
c	264	15	11.7	700	20	AF053852	Melospiza melodia mito 4.98e+02
c	265	15	11.7	700	20	AF053881	Melospiza melodia mito 4.98e+02
c	266	15	11.7	700	20	AF053853	Melospiza melodia mito 4.98e+02
c	267	15	11.7	700	20	AF053861	Melospiza melodia mito 4.98e+02
c	268	15	11.7	700	20	AF053869	Melospiza melodia mito 4.98e+02
c	269	15	11.7	700	20	AF053879	Melospiza melodia mito 4.98e+02
c	270	15	11.7	700	20	AF053880	Melospiza melodia mito 4.98e+02
c	271	15	11.7	700	20	AF053850	Melospiza melodia mito 4.98e+02
c	272	15	11.7	700	20	AF053874	Melospiza melodia mito 4.98e+02
c	273	15	11.7	700	20	AF053875	Melospiza melodia mito 4.98e+02
c	274	15	11.7	700	20	AF053878	Melospiza melodia mito 4.98e+02
c	275	15	11.7	700	20	AF053872	Melospiza melodia mito 4.98e+02
c	276	15	11.7	700	20	AF053851	Melospiza melodia mito 4.98e+02
c	277	15	11.7	700	20	AF053876	Melospiza melodia mito 4.98e+02
c	278	15	11.7	700	20	AF053839	Melospiza melodia mito 4.98e+02
c	279	15	11.7	700	20	AF053882	Melospiza melodia mito 4.98e+02
c	280	15	11.7	700	20	AF053865	Melospiza melodia mito 4.98e+02
c	281	15	11.7	700	20	AF053846	Melospiza melodia mito 4.98e+02
c	282	15	11.7	700	20	AF053859	Melospiza melodia mito 4.98e+02
c	283	15	11.7	700	20	AF053829	Melospiza melodia mito 4.98e+02
c	284	15	11.7	700	20	AF053855	Melospiza melodia mito 4.98e+02
c	285	15	11.7	700	20	AF053860	Melospiza melodia mito 4.98e+02
c	286	15	11.7	700	20	AF053842	Melospiza melodia mito 4.98e+02
c	287	15	11.7	700	20	AF053848	Melospiza melodia mito 4.98e+02
c	288	15	11.7	700	20	AF053834	Melospiza melodia mito 4.98e+02
c	289	15	11.7	700	20	AF053843	Melospiza melodia mito 4.98e+02
c	290	15	11.7	700	20	AF053835	Melospiza melodia mito 4.98e+02
c	291	15	11.7	700	20	AF053847	Melospiza melodia mito 4.98e+02
c	292	15	11.7	700	20	AF053868	Melospiza melodia mito 4.98e+02
c	293	15	11.7	700	20	AF053837	Melospiza melodia mito 4.98e+02
c	294	15	11.7	700	20	AF053844	Melospiza melodia mito 4.98e+02
c	295	15	11.7	700	20	AF053845	Melospiza melodia mito 4.98e+02
c	296	15	11.7	700	20	AF053867	Melospiza melodia mito 4.98e+02
c	297	15	11.7	700	20	AF053838	Melospiza melodia mito 4.98e+02

c	298	15	11.7	700	20	AF053840	Melospiza melodia mito 4.98e+02
c	299	15	11.7	700	20	AF053873	Melospiza melodia mito 4.98e+02
c	300	15	11.7	700	20	AF053870	Melospiza melodia mito 4.98e+02
	301	15	11.7	706	26	AG000529	Homo sapiens genomic D 4.98e+02
	302	15	11.7	706	26	AG000530	Homo sapiens genomic D 4.98e+02
c	303	15	11.7	717	26	AG000498	Homo sapiens genomic D 4.98e+02
c	304	15	11.7	717	26	AG000499	Homo sapiens genomic D 4.98e+02
	305	15	11.7	719	26	AG000527	Homo sapiens genomic D 4.98e+02
	306	15	11.7	732	26	AG000346	Homo sapiens genomic D 4.98e+02
c	307	15	11.7	767	18	PFS25	Plasmodium falciparum 4.98e+02
c	308	15	11.7	778	30	G41973	SHGC-69008 Human Homo 4.98e+02
	309	15	11.7	792	25	HSPST1CA	H.sapiens Pst 1 fragme 4.98e+02
c	310	15	11.7	880	24	AB010877	Nicotiana tabacum mRNA 4.98e+02
	311	15	11.7	885	26	AF053684	Lemur catta NADH dehyd 4.98e+02
	312	15	11.7	895	25	LEMMTTGH	Lemur catta mitochondr 4.98e+02
	313	15	11.7	910	28	MM1UBR2	Mus musculus ubiquitin 4.98e+02
c	314	15	11.7	953	18	LSB14GLCT	L.stagnalis mRNA homol 4.98e+02
c	315	15	11.7	960	15	BVCSBAKER	B.vinsonii gltA gene (4.98e+02
	316	15	11.7	1000	25	HSUATM49	Human ataxia-telangiec 4.98e+02
c	317	15	11.7	1011	23	CMCAB	C.moewusii cab mRNA fo 4.98e+02
c	318	15	11.7	1047	21	A40257	Sequence 7 from Patent 4.98e+02
	319	15	11.7	1057	27	HSD72518	Human destrin-2 pseudo 4.98e+02
	320	15	11.7	1105	28	RNRN45	Rat 45S rDNA gene tran 4.98e+02
	321	15	11.7	1105	28	RATRGL7	rat 45s precursor rna 4.98e+02
	322	15	11.7	1130	15	SPTRJ5251	S.pneumoniae right jun 4.98e+02
c	323	15	11.7	1138	23	POTPATB	Potato patatin class I 4.98e+02
c	324	15	11.7	1139	23	POTPATC	Potato patatin class I 4.98e+02
c	325	15	11.7	1140	28	TIU35171	Trinomys iheringi ligh 4.98e+02
	326	15	11.7	1161	20	U71222	Desmognathus fuscus 12 4.98e+02
	327	15	11.7	1161	20	U71237	Desmognathus aeneus 12 4.98e+02
	328	15	11.7	1162	20	U71241	Desmognathus fuscus 12 4.98e+02
	329	15	11.7	1162	20	U71223	Desmognathus ochrophae 4.98e+02
	330	15	11.7	1162	20	U71230	Desmognathus imitator 4.98e+02
	331	15	11.7	1164	20	U71234	Desmognathus conanti 1 4.98e+02
	332	15	11.7	1164	20	U71235	Desmognathus conanti 1 4.98e+02
	333	15	11.7	1164	20	U71227	Desmognathus ochrophae 4.98e+02
	334	15	11.7	1167	18	PFACSPK	Plasmodium reichenowi 4.98e+02
c	335	15	11.7	1194	26	AF052509	Homo sapiens esterase 4.98e+02
c	336	15	11.7	1198	25	HSFGFG1	Human basic fibroblast 4.98e+02
	337	15	11.7	1251	15	STAETAA	S.aureus eta gene enco 4.98e+02
	338	15	11.7	1311	15	BORFLII	Borrelia burgdorferi f 4.98e+02
	339	15	11.7	1328	25	HUMMNMP	Human major nuclear ma 4.98e+02
c	340	15	11.7	1335	18	DDU15210	Dictyostelium discoide 4.98e+02
	341	15	11.7	1377	20	XLIGVH6	X.laevis Vh VI gene fo 4.98e+02
	342	15	11.7	1378	28	MUSAGS	Mus musculus alpha-D-g 4.98e+02
	343	15	11.7	1390	21	I48974	Sequence 1 from patent 4.98e+02
	344	15	11.7	1391	15	STAETA	S.aureus exfoliative t 4.98e+02
c	345	15	11.7	1391	23	STPPGT16	S.tuberosum promoter r 4.98e+02
c	346	15	11.7	1407	23	STPB24	S.tuberosum promoter r 4.98e+02
c	347	15	11.7	1439	15	ECCYBB	E. coli cybB gene for 4.98e+02
c	348	15	11.7	1453	28	RATHES1A	Rat mRNA for transcrip 4.98e+02

	349	15	11.7	1455	15	STASETA	Staphylococcus aureus 4.98e+02
	350	15	11.7	1558	23	SCYOR238W	S.cerevisiae chromosom 4.98e+02
	351	15	11.7	1585	28	MMU50716	Mus musculus alpha-gal 4.98e+02
	352	15	11.7	1590	28	MMCTCA1	M.musculus mRNA for cy 4.98e+02
c	353	15	11.7	1701	25	S81809	basic fibroblast growt 4.98e+02
	354	15	11.7	1737	23	AF049098	Trifolium repens beta- 4.98e+02
c	355	15	11.7	1770	23	STPATGENE	S.tuberosum class I pa 4.98e+02
	356	15	11.7	1807	15	MYCREQSEQ	M.hyopneumoniae genome 4.98e+02
	357	15	11.7	1836	28	RATPHOLA01	Rat phospholamban gene 4.98e+02
c	358	15	11.7	1845	23	AB003516	Panax ginseng mRNA for 4.98e+02
	359	15	11.7	1890	26	AF050199	Homo sapiens putative 4.98e+02
	360	15	11.7	1903	23	SCU03914	Saccharomyces cerevisi 4.98e+02
c	361	15	11.7	1905	15	CONCYTSYN	C.burnetii citrate syn 4.98e+02
c	362	15	11.7	1915	28	MMU71191	Mus musculus insulin-l 4.98e+02
	363	15	11.7	1927	28	MMU53219	Mus musculus GTPase IG 4.98e+02
c	364	15	11.7	1969	21	I09137	Sequence 1 from Patent 4.98e+02
c	365	15	11.7	1969	21	I08089	Sequence 2 from Patent 4.98e+02
c	366	15	11.7	2011	16	AF000365	Borrelia burgdorferi o 4.98e+02
	367	15	11.7	2015	33	FLCHAGY	Influenza C/Yamagata/1 4.98e+02
	368	15	11.7	2015	33	FLCHAAA	Influenza C/Ann Arbor/ 4.98e+02
	369	15	11.7	2126	23	SCYOR237W	S.cerevisiae chromosom 4.98e+02
	370	15	11.7	2184	23	CAU53216	Candida albicans 3-deo 4.98e+02
	371	15	11.7	2220	23	SCTIP1	S.cerevisiae TIP1 gene 4.98e+02
	372	15	11.7	2255	23	RCHCHI1	R. niveus chi1 gene fo 4.98e+02
	373	15	11.7	2255	21	E06738	Rhizopus niveus chitin 4.98e+02
	374	15	11.7	2285	26	U92957	Pygathrix bieti NADH d 4.98e+02
	375	15	11.7	2285	26	U92958	Pygathrix bieti NADH d 4.98e+02
	376	15	11.7	2285	26	U92951	Pygathrix bieti NADH d 4.98e+02
	377	15	11.7	2285	26	U92960	Pygathrix bieti NADH d 4.98e+02
	378	15	11.7	2285	26	U92956	Pygathrix bieti NADH d 4.98e+02
	379	15	11.7	2285	26	U92959	Pygathrix bieti NADH d 4.98e+02
	380	15	11.7	2286	26	U92961	Pygathrix roxellana NA 4.98e+02
	381	15	11.7	2286	26	U92955	Trachypithecus francoi 4.98e+02
	382	15	11.7	2286	26	U92963	Pygathrix roxellana NA 4.98e+02
	383	15	11.7	2286	26	U92954	Trachypithecus francoi 4.98e+02
	384	15	11.7	2286	26	U92962	Pygathrix roxellana NA 4.98e+02
	385	15	11.7	2286	26	U92953	Trachypithecus francoi 4.98e+02
	386	15	11.7	2287	26	U92966	Trachypithecus leucoce 4.98e+02
	387	15	11.7	2287	26	U92969	Trachypithecus phayrei 4.98e+02
	388	15	11.7	2287	26	U92964	Pygathrix nemaeus NADH 4.98e+02
	389	15	11.7	2287	26	U92967	Trachypithecus leucoce 4.98e+02
	390	15	11.7	2287	26	U92965	Pygathrix nemaeus NADH 4.98e+02
	391	15	11.7	2287	26	U92968	Trachypithecus leucoce 4.98e+02
	392	15	11.7	2288	26	U92970	Trachypithecus phayrei 4.98e+02
	393	15	11.7	2301	28	MMJ002198	Mus musculus domestici 4.98e+02
	394	15	11.7	2400	16	LLJ002203	Lactococcus lactis lac 4.98e+02
	395	15	11.7	2474	18	DMU13014	Drosophila melanogaste 4.98e+02
	396	15	11.7	2500	23	YSCYSP3G	Yeast subtilisin-like 4.98e+02
	397	15	11.7	2536	24	AB010394	Yarrowia lipolytica ge 4.98e+02
	398	15	11.7	2540	23	SCYOR003W	S.cerevisiae chromosom 4.98e+02
c	399	15	11.7	2572	25	HSIRLBEX4	H.sapiens IRLB gene (e 4.98e+02

	400	15	11.7	2638	23	SC13OLP	S.commersonii (pOSML13 4.98e+02
	401	15	11.7	2763	23	SCYGL145W	S.cerevisiae chromosom 4.98e+02
	402	15	11.7	2801	28	MMU34071	Mus musculus alpha-gal 4.98e+02
c	403	15	11.7	2808	20	NVRETINBP	N.viridescens mRNA for 4.98e+02
c	404	15	11.7	2825	28	MUSHES1	Mouse helix-loop-helix 4.98e+02
c	405	15	11.7	2901	25	HSIRLB5	H.sapiens IRLB gene -(e 4.98e+02
c	406	15	11.7	2961	25	HUMCOL4A3X	Human alpha-3 type IV 4.98e+02
	407	15	11.7	3013	28	RATFABP	Rat intestinal fatty a 4.98e+02
	408	15	11.7	3044	23	SPCEN2B	S.pombe cen2 central r 4.98e+02
c	409	15	11.7	3050	23	SCU09499	Saccharomyces cerevisi 4.98e+02
	410	15	11.7	3096	25	HSRPTK	H.sapiens mRNA for rec 4.98e+02
	411	15	11.7	3157	21	I68739	Sequence 3 from patent 4.98e+02
c	412	15	11.7	3157	21	I68740	Sequence 5 from patent 4.98e+02
	413	15	11.7	3171	18	CETC57RAN	C.elegans (mut-2 mutat 4.98e+02
	414	15	11.7	3171	18	CEU12433	Caenorhabditis elegans 4.98e+02
	415	15	11.7	3219	18	EHU89655	Entamoeba histolytica 4.98e+02
	416	15	11.7	3259	28	MMTYRO10	M.musculus (C57Bl/6) T 4.98e+02
	417	15	11.7	3321	28	AF056079	Mus musculus amyloid p 4.98e+02
	418	15	11.7	3372	21	I40040	Sequence 1 from patent 4.98e+02
	419	15	11.7	3372	25	HUMP13KIN	Human P13-kinase assoc 4.98e+02
	420	15	11.7	3372	21	I13158	Sequence 1 from patent 4.98e+02
	421	15	11.7	3372	21	I69429	Sequence 1 from patent 4.98e+02
c	422	15	11.7	3372	21	I40046	Sequence 48 from paten 4.98e+02
	423	15	11.7	3547	28	AB001509	Mesocricetus auratus g 4.98e+02
	424	15	11.7	3602	15	LLU44843	Lactococcus lactis pla 4.98e+02
c	425	15	11.7	3635	23	SCPGM2	S.cerevisiae gene for 4.98e+02
	426	15	11.7	3638	16	LLCAJ3960	Lactococcus lactis cre 4.98e+02
	427	15	11.7	3653	27	AF009356	Homo sapiens regulator 4.98e+02
	428	15	11.7	3743	16	AF013595	Lactococcus lactis pla 4.98e+02
	429	15	11.7	3852	21	I04400	Sequence 35 from Paten 4.98e+02
	430	15	11.7	3852	21	E00422	Human genomic DNA codi 4.98e+02
	431	15	11.7	3866	25	HUM21DC5E	Homo sapiens (subclone 4.98e+02
	432	15	11.7	3892	20	CHKLYH	Chicken lysyl hydroxyl 4.98e+02
c	433	15	11.7	3950	23	SCYBR203W	S.cerevisiae chromosom 4.98e+02
c	434	15	11.7	3994	27	HSA9948	Homo sapiens mfap2 gen 4.98e+02
c	435	15	11.7	4029	23	POTPATA	Potato patatin class I 4.98e+02
	436	15	11.7	4074	25	HUMFAC13	H.sapiens factor VIII 4.98e+02
c	437	15	11.7	4080	23	SCYJL108C	S.cerevisiae chromosom 4.98e+02
c	438	15	11.7	4081	23	SCYDL215C	S.cerevisiae chromosom 4.98e+02
	439	15	11.7	4107	28	MMU20238	Mus musculus GTPase-ac 4.98e+02
c	440	15	11.7	4171	18	CELM116	Caenorhabditis elegans 4.98e+02
	441	15	11.7	4183	18	ONAMY71	Ostrinia nubilalis ost 4.98e+02
c	442	15	11.7	4204	18	PFU21939	Plasmodium falciparum 4.98e+02
	443	15	11.7	4236	19	OARLGPR	O.aries DNA for BLG ge 4.98e+02
c	444	15	11.7	4267	31	PRS304	Yeast integrative vect 4.98e+02
c	445	15	11.7	4274	31	PRS404	Yeast integrative vect 4.98e+02
	446	15	11.7	4294	23	SCYDR032C	S.cerevisiae chromosom 4.98e+02
	447	15	11.7	4353	26	AF009205	Homo sapiens clone L5 4.98e+02
c	448	15	11.7	4699	21	I56911	Sequence 6 from patent 4.98e+02
c	449	15	11.7	4699	21	I38436	Sequence 6 from patent 4.98e+02
c	450	15	11.7	4699	21	I75104	Sequence 6 from patent 4.98e+02

c	451	15	11.7	4699	21	I59777	Sequence 6 from patent 4.98e+02
c	452	15	11.7	4699	21	I07376	Sequence 28 from Paten 4.98e+02
c	453	15	11.7	4699	23	TOBGLB13B	Nicotiana tabacum gluc 4.98e+02
c	454	15	11.7	4783	31	PRS314	Yeast centromere vecto 4.98e+02
c	455	15	11.7	4788	31	PRS414	Yeast centromere vecto 4.98e+02
	456	15	11.7	4838	23	SCGDH2A	S.cerevisiae GDH2 gene 4.98e+02
c	457	15	11.7	4893	26	HSXLRPGN1	H.sapiens RP3 gene (XL 4.98e+02
c	458	15	11.7	4958	23	SCYBR202W	S.cerevisiae chromosom 4.98e+02
	459	15	11.7	4967	27	HSU73595	Human prohormone conve 4.98e+02
	460	15	11.7	5041	23	SCYDR033W	S.cerevisiae chromosom 4.98e+02
	461	15	11.7	5061	26	HSPHOSI3K	H.sapiens mRNA for pho 4.98e+02
c	462	15	11.7	5257	23	AB001569	Carrot DNA for transpo 4.98e+02
	463	15	11.7	5275	31	CVU37456	Yeast CUP1 expression/ 4.98e+02
c	464	15	11.7	5345	23	BOSRK3	B.oleracea SRK3 gene. 4.98e+02
c	465	15	11.7	5386	15	MYCVPP11	Mycoplasma hyopneumoni 4.98e+02
c	466	15	11.7	5616	31	PRS424	Yeast episomal vector 4.98e+02
	467	15	11.7	5637	28	RATCAMKI	Rat calmodulin-depende 4.98e+02
c	468	15	11.7	5746	23	SCU14730	Saccharomyces cerevisi 4.98e+02
c	469	15	11.7	5781	31	CVU37457	Yeast CUP1 expression- 4.98e+02
	470	15	11.7	5912	25	HSU26455	Human phosphatidylinos 4.98e+02
	471	15	11.7	5954	20	XLACTA	Xenopus laevis sarcome 4.98e+02
c	472	15	11.7	5999	21	I08188	Sequence 3 from Patent 4.98e+02
c	473	15	11.7	6094	23	STPATG1	Potato (tetraploid var 4.98e+02
	474	15	11.7	6099	18	TBANTAT	Trypanosoma brucei Ant 4.98e+02
	475	15	11.7	6487	26	HSA2MGLB1	H.sapiens gene for alp 4.98e+02
c	476	15	11.7	6525	31	AF063848	Cloning vector pESC-TR 4.98e+02
	477	15	11.7	6624	31	CVU37458	Yeast CUP1 expression- 4.98e+02
c	478	15	11.7	6779	31	AF008932	DNA-binding vector pOD 4.98e+02
c	479	15	11.7	6827	31	CVU57443	Cloning vector pODB8, 4.98e+02
c	480	15	11.7	6970	18	CED1043	Caenorhabditis elegans 4.98e+02
c	481	15	11.7	6973	18	CELP09D1	Caenorhabditis elegans 4.98e+02
	482	15	11.7	6999	21	I77105	Sequence 1 from patent 4.98e+02
	483	15	11.7	7056	21	I27063	Sequence 1 from patent 4.98e+02
c	484	15	11.7	7190	31	CVU46855	Cloning vector pAS1-te 4.98e+02
	485	15	11.7	7272	21	I05404	Sequence 3 from Patent 4.98e+02
c	486	15	11.7	7345	25	HUMGOS8PP	Human helix-loop-helix 4.98e+02
	487	15	11.7	7440	21	I08345	Sequence 2 from Patent 4.98e+02
	488	15	11.7	7440	21	I08457	Sequence 1 from Patent 4.98e+02
c	489	15	11.7	7445	26	HSU79410	Homo sapiens thrombosp 4.98e+02
	490	15	11.7	7614	33	PAP41CG	Human papillomavirus t 4.98e+02
c	491	15	11.7	8033	18	PPINMP	Paramecium primaurelia 4.98e+02
	492	15	11.7	8241	21	A05328	Synthetic Factor VIII 4.98e+02
	493	15	11.7	8241	21	A07042	Artificial mRNA for fa 4.98e+02
c	494	15	11.7	8392	31	CVU30497	Cloning vector pAS2-1, 4.98e+02
c	495	15	11.7	8414	31	CVU30496	Cloning vector pAS2, c 4.98e+02
	496	15	11.7	8467	26	AB002292	Human mRNA for KIAA029 4.98e+02
	497	15	11.7	8831	21	E00527	Complete cDNA encoding 4.98e+02
	498	15	11.7	8856	23	SOU69142	Spinacia oleracea beta 4.98e+02
	499	15	11.7	8967	25	HSFVIIIR	Human mRNA for factor 4.98e+02
	500	15	11.7	8967	21	I71409	Sequence 1 from patent 4.98e+02
	501	15	11.7	9009	21	I63424	Sequence 1 from patent 4.98e+02

	502	15	11.7	9009	25	HUMFVIII	Human coagulation fact 4.98e+02
	503	15	11.7	9009	21	I31901	Sequence 3 from patent 4.98e+02
	504	15	11.7	9029	25	HUMFVIIIC	Human coagulation fact 4.98e+02
	505	15	11.7	9163	15	ECAFAF	E.coli afaF, afaA, afa 4.98e+02
	506	15	11.7	9314	18	CELW06B4	Caenorhabditis elegans 4.98e+02
	507	15	11.7	9385	27	HSU33841	Human ataxia telangiec 4.98e+02
c	508	15	11.7	9705	23	CHLC27	C.ellipsoidea rbcL, rp 4.98e+02
c	509	15	11.7	9993	16	AE001285	Chlamydia trachomatis 4.98e+02
c	510	15	11.7	10288	21	A58524	Sequence 28 from Paten 4.98e+02
	511	15	11.7	10511	15	U67520	Methanococcus jannasch 4.98e+02
c	512	15	11.7	10797	20	PGNCP37	Columba livia annexin 4.98e+02
	513	15	11.7	10807	21	I47708	Sequence 7 from patent 4.98e+02
c	514	15	11.7	10980	15	AE001035	Archaeoglobus fulgidus 4.98e+02
c	515	15	11.7	11189	16	AE001339	Chlamydia trachomatis 4.98e+02
	516	15	11.7	11447	15	HPAE000599	Helicobacter pylori se 4.98e+02
c	517	15	11.7	11630	15	ECAE000239	Escherichia coli K-12 4.98e+02
	518	15	11.7	11759	15	HPAE000616	Helicobacter pylori se 4.98e+02
	519	15	11.7	12215	15	U67560	Methanococcus jannasch 4.98e+02
c	520	15	11.7	13072	15	U67473	Methanococcus jannasch 4.98e+02
c	521	15	11.7	13414	21	A58523	Sequence 27 from Paten 4.98e+02
c	522	15	11.7	13474	23	AB011484	Arabidopsis thaliana g 4.98e+02
c	523	15	11.7	13651	18	CELD2007	C. elegans cosmid D200 4.98e+02
	524	15	11.7	14102	15	AE001117	Borrelia burgdorferi (4.98e+02
	525	15	11.7	14382	28	MMU50715	Mus musculus alpha-gal 4.98e+02
	526	15	11.7	14613	33	NPHCUECDYS	Autographa californica 4.98e+02
	527	15	11.7	15155	21	I04391	Sequence 26 from Paten 4.98e+02
	528	15	11.7	15495	15	U39694	Mycoplasma genitalium 4.98e+02
c	529	15	11.7	15858	15	D90782	E.coli genomic DNA, Ko 4.98e+02
c	530	15	11.7	16342	16	BMAJ758	Bacillus megaterium 16 4.98e+02
	531	15	11.7	16481	25	HUMXIST	Homo sapiens X (inacti 4.98e+02
c	532	15	11.7	16635	15	D90781	E.coli genomic DNA, Ko 4.98e+02
	533	15	11.7	17925	28	MMFHIT2	Mus musculus fragile h 4.98e+02
	534	15	11.7	18032	15	EFU09422	Enterococcus faecalis 4.98e+02
c	535	15	11.7	20129	18	CEKO2D3	Caenorhabditis elegans 4.98e+02
c	536	15	11.7	22278	18	CET05E12	Caenorhabditis elegans 4.98e+02
c	537	15	11.7	22484	25	HSU29953	Human pigment epithel1 4.98e+02
	538	15	11.7	22803	23	SCVII228	S.cerevisiae 22.8kb ge 4.98e+02
c	539	15	11.7	23196	20	FRU63926	Fugu rubripes platelet 4.98e+02
	540	15	11.7	23379	18	CELT08E11	Caenorhabditis elegans 4.98e+02
c	541	15	11.7	24049	18	CELZK354	Caenorhabditis elegans 4.98e+02
c	542	15	11.7	24232	18	CEC03E10	Caenorhabditis elegans 4.98e+02
	543	15	11.7	24384	18	CELR03H4	Caenorhabditis elegans 4.98e+02
	544	15	11.7	24628	23	SC9673	S.cerevisiae chromosom 4.98e+02
	545	15	11.7	24749	18	CET28B11	Caenorhabditis elegans 4.98e+02
c	546	15	11.7	25402	18	CELT28H11	Caenorhabditis elegans 4.98e+02
c	547	15	11.7	25433	18	CEK08F4	Caenorhabditis elegans 4.98e+02
	548	15	11.7	25601	18	CELB0252	Caenorhabditis elegans 4.98e+02
	549	15	11.7	25987	18	CEE01G6	Caenorhabditis elegans 4.98e+02
c	550	15	11.7	26076	18	CEF14D7	Caenorhabditis elegans 4.98e+02
c	551	15	11.7	26542	18	CEM18	Caenorhabditis elegans 4.98e+02
c	552	15	11.7	27254	18	CER07B5	Caenorhabditis elegans 4.98e+02

c	553	15	11.7	27394	18	CEF32B6	Caenorhabditis elegans	4.98e+02
c	554	15	11.7	27428	18	CEF59B10	Caenorhabditis elegans	4.98e+02
c	555	15	11.7	27874	20	MMU71085	Mus musculus insulin-l	4.98e+02
	556	15	11.7	28684	18	CEC44H9	Caenorhabditis elegans	4.98e+02
c	557	15	11.7	28736	18	CET06G6	Caenorhabditis elegans	4.98e+02
	558	15	11.7	29039	18	CEF56P3	Caenorhabditis elegans	4.98e+02
	559	15	11.7	29350	18	CEC14B1	Caenorhabditis elegans	4.98e+02
c	560	15	11.7	29792	23	SPAC7D4	S.pombe chromosome I c	4.98e+02
	561	15	11.7	29848	25	HUMU39B3	Human cosmid U39B3, co	4.98e+02
	562	15	11.7	30469	24	SCU43491	Saccharomyces cerevisi	4.98e+02
	563	15	11.7	30848	18	AC004307	Drosophila melanogaste	4.98e+02
	564	15	11.7	31531	18	CEZC434	Caenorhabditis elegans	4.98e+02
	565	15	11.7	31536	18	CEE04D5	Caenorhabditis elegans	4.98e+02
c	566	15	11.7	31754	22	AF009630	Bacteriophage bIL170,	4.98e+02
	567	15	11.7	32009	18	CET06D8	Caenorhabditis elegans	4.98e+02
	568	15	11.7	33013	18	DMU87286	Drosophila melanogaste	4.98e+02
	569	15	11.7	33016	18	CELF58G4	Caenorhabditis elegans	4.98e+02
c	570	15	11.7	33038	18	CEY6G8	Caenorhabditis elegans	4.98e+02
c	571	15	11.7	33921	18	CEW09D10	Caenorhabditis elegans	4.98e+02
	572	15	11.7	33950	18	CEF37B12	Caenorhabditis elegans	4.98e+02
	573	15	11.7	34024	18	CEC18E9	Caenorhabditis elegans	4.98e+02
c	574	15	11.7	34193	18	CEC38C10	Caenorhabditis elegans	4.98e+02
	575	15	11.7	34263	18	CET02E9	Caenorhabditis elegans	4.98e+02
	576	15	11.7	35227	18	CELK03A1	Caenorhabditis elegans	4.98e+02
c	577	15	11.7	35253	18	CELC24A8	Caenorhabditis elegans	4.98e+02
	578	15	11.7	35303	18	CELC44B11	Caenorhabditis elegans	4.98e+02
c	579	15	11.7	35369	24	SPCC663	S.pombe chromosome III	4.98e+02
c	580	15	11.7	35435	17	CEH06P14	Caenorhabditis elegans	4.98e+02
c	581	15	11.7	35500	18	CEC54C6	Caenorhabditis elegans	4.98e+02
c	582	15	11.7	35557	18	CEY106G6D	Caenorhabditis elegans	4.98e+02
	583	15	11.7	35728	18	CELK08B5	Caenorhabditis elegans	4.98e+02
	584	15	11.7	35811	23	SC9718	S.cerevisiae chromosom	4.98e+02
c	585	15	11.7	35888	18	CET19B10	Caenorhabditis elegans	4.98e+02
c	586	15	11.7	35964	18	CEB0365	Caenorhabditis elegans	4.98e+02
	587	15	11.7	36052	18	CELZK652	C. elegans cosmid ZK65	4.98e+02
	588	15	11.7	36189	18	CELC08G9	Caenorhabditis elegans	4.98e+02
c	589	15	11.7	36238	26	D87015	Human (lambda) DNA for	4.98e+02
	590	15	11.7	36254	28	AC003995	Mouse Cosmid ma66a09d	4.98e+02
	591	15	11.7	36298	18	CELF25B4	Caenorhabditis elegans	4.98e+02
	592	15	11.7	36306	18	CEM142	Caenorhabditis elegans	4.98e+02
	593	15	11.7	36309	18	CEF45G2	Caenorhabditis elegans	4.98e+02
c	594	15	11.7	36493	23	SPBC2G2	S.pombe chromosome II	4.98e+02
	595	15	11.7	36495	24	AB015473	Arabidopsis thaliana g	4.98e+02
c	596	15	11.7	36510	26	AC004397	Homo sapiens Cosmid gl	4.98e+02
	597	15	11.7	36558	25	HSU81834	Human cosmid LL12NC01-	4.98e+02
c	598	15	11.7	36730	18	CELF33H12	Caenorhabditis elegans	4.98e+02
c	599	15	11.7	36733	24	SPBC3B9	S.pombe chromosome II	4.98e+02
	600	15	11.7	36987	26	U73639	Human Chromosome 11 Co	4.98e+02
	601	15	11.7	36997	18	CEZK512	Caenorhabditis elegans	4.98e+02
c	602	15	11.7	37013	24	SPAC3A12	S.pombe chromosome I c	4.98e+02
c	603	15	11.7	37018	17	AC005647	*** SEQUENCING IN PROG	4.98e+02

c	604	15	11.7	37136	26	HSSRL9A13	Human DNA sequence fro	4.98e+02
c	605	15	11.7	37404	23	SC8119	S.cerevisiae chromosom	4.98e+02
c	606	15	11.7	37476	18	CELP53B1	Caenorhabditis elegans	4.98e+02
c	607	15	11.7	37549	23	SCXLTORF5	S.cerevisiae DNA from	4.98e+02
c	608	15	11.7	37640	18	CEZK1073	Caenorhabditis elegans	4.98e+02
c	609	15	11.7	37689	18	CEF18C12	Caenorhabditis elegans	4.98e+02
c	610	15	11.7	37711	18	CEZK896	Caenorhabditis elegans	4.98e+02
	611	15	11.7	38000	18	CEC3998I0	Caenorhabditis elegans	4.98e+02
	612	15	11.7	38031	27	AC005764	Homo sapiens chromosom	4.98e+02
	613	15	11.7	38336	18	CELT09B4	Caenorhabditis elegans	4.98e+02
c	614	15	11.7	38383	18	CEF35H8	Caenorhabditis elegans	4.98e+02
c	615	15	11.7	38499	18	CELF02E8	Caenorhabditis elegans	4.98e+02
c	616	15	11.7	38505	18	CEZK930	Caenorhabditis elegans	4.98e+02
c	617	15	11.7	38525	18	CELB0414	Caenorhabditis elegans	4.98e+02
	618	15	11.7	38640	18	CELF44A2	Caenorhabditis elegans	4.98e+02
	619	15	11.7	38731	18	CELF32A5	Caenorhabditis elegans	4.98e+02
	620	15	11.7	38754	26	AC005173	Homo sapiens clone UWG	4.98e+02
c	621	15	11.7	38811	17	HSE116C6	Human DNA sequence fro	4.98e+02
c	622	15	11.7	38919	18	DMC133E12	Drosophila melanogaste	4.98e+02
c	623	15	11.7	39090	23	CS39KBCIV	S.cerevisiae 39kb DNA	4.98e+02
c	624	15	11.7	39111	18	CEF54D5	Caenorhabditis elegans	4.98e+02
	625	15	11.7	39172	26	AC005264	Homo sapiens chromosom	4.98e+02
	626	15	11.7	39333	18	CEB0564	Caenorhabditis elegans	4.98e+02
	627	15	11.7	39721	18	CELC23H3	Caenorhabditis elegans	4.98e+02
c	628	15	11.7	39805	18	CELZK430	Caenorhabditis elegans	4.98e+02
	629	15	11.7	39852	18	CELF32A6	Caenorhabditis elegans	4.98e+02
c	630	15	11.7	39916	18	CELC31H1	Caenorhabditis elegans	4.98e+02
c	631	15	11.7	39931	18	CELC54E4	Caenorhabditis elegans	4.98e+02
	632	15	11.7	39933	18	CEW01A6	Caenorhabditis elegans	4.98e+02
c	633	15	11.7	39960	25	HSU162C4	Human DNA sequence fro	4.98e+02
c	634	15	11.7	39977	18	CET08D2	Caenorhabditis elegans	4.98e+02
c	635	15	11.7	39989	18	CEF14F8	Caenorhabditis elegans	4.98e+02
	636	15	11.7	40265	18	CEM28	Caenorhabditis elegans	4.98e+02
c	637	15	11.7	40302	18	CEC11G6	Caenorhabditis elegans	4.98e+02
	638	15	11.7	40450	18	CEF49E12	Caenorhabditis elegans	4.98e+02
c	639	15	11.7	40685	26	AC002052	Homo sapiens Chromosom	4.98e+02
	640	15	11.7	40915	24	SPAC4G9	S.pombe chromosome I c	4.98e+02
	641	15	11.7	40921	18	CELB0432	Caenorhabditis elegans	4.98e+02
	642	15	11.7	40997	18	CELR05P9	Caenorhabditis elegans	4.98e+02
	643	15	11.7	41009	17	CEY94A7	Caenorhabditis elegans	4.98e+02
c	644	15	11.7	41100	24	SPBC2D10	S.pombe chromosome II	4.98e+02
c	645	15	11.7	41152	18	CET08G3	Caenorhabditis elegans	4.98e+02
	646	15	11.7	41322	26	AC004076	Homo sapiens chromosom	4.98e+02
c	647	15	11.7	41339	27	AC005490	Homo sapiens UWGC:g139	4.98e+02
	648	15	11.7	41397	18	CEC50H2	Caenorhabditis elegans	4.98e+02
	649	15	11.7	41537	26	D87024	Human (lambda) DNA for	4.98e+02
	650	15	11.7	41645	18	CELC03H5	Caenorhabditis elegans	4.98e+02
	651	15	11.7	41813	26	AC004012	Human BAC clone GS285F	4.98e+02
c	652	15	11.7	41954	18	CEC15A11	Caenorhabditis elegans	4.98e+02
	653	15	11.7	41988	25	HSV362H12	Human DNA sequence fro	4.98e+02
c	654	15	11.7	42149	23	SPAC1B3	S.pombe chromosome I c	4.98e+02

c	655	15	11.7	42168	18	CELC04E6	Caenorhabditis elegans	4.98e+02
c	656	15	11.7	42366	18	CEF25H2	Caenorhabditis elegans	4.98e+02
	657	15	11.7	42460	24	SPCC1020	S.pombe chromosome III	4.98e+02
c	658	15	11.7	42681	26	AC004560	Homo sapiens chromosom	4.98e+02
c	659	15	11.7	43394	26	HSAC000359	Human cosmid gl572c264	4.98e+02
	660	15	11.7	43455	18	CELT07A9	Caenorhabditis elegans	4.98e+02
	661	15	11.7	43536	18	CELC10H11	Caenorhabditis elegans	4.98e+02
	662	15	11.7	44943	18	CELC44B7	Caenorhabditis elegans	4.98e+02
c	663	15	11.7	45713	18	CELB0511	Caenorhabditis elegans	4.98e+02
c	664	15	11.7	46074	26	AC005271	Homo sapiens chromosom	4.98e+02
	665	15	11.7	46423	18	CEC47A10	Caenorhabditis elegans	4.98e+02
	666	15	11.7	47167	27	HS757P12	Human DNA sequence fro	4.98e+02
c	667	15	11.7	47202	18	AC004724	Drosophila melanogaste	4.98e+02
c	668	15	11.7	47329	27	HS41P2A	Human DNA sequence fro	4.98e+02
c	669	15	11.7	47338	28	AF026469	Mus musculus ubiquitin	4.98e+02
c	670	15	11.7	47558	18	AF030694	Plasmodium falciparum	4.98e+02
c	671	15	11.7	48351	26	HS397C4	Human DNA sequence fro	4.98e+02
	672	15	11.7	50651	26	AC004762	Homo sapiens chromosom	4.98e+02
c	673	15	11.7	52616	26	HS125H23	Human DNA sequence fro	4.98e+02
	674	15	11.7	53561	15	AE000790	Borrelia burgdorferi p	4.98e+02
	675	15	11.7	56105	15	SS56KBFR	S.solfataricus 56 kbp	4.98e+02
c	676	15	11.7	56150	17	CEP56P12	*** SEQUENCING IN PROG	4.98e+02
c	677	15	11.7	57186	26	AC004042	Homo sapiens chromosom	4.98e+02
c	678	15	11.7	63916	26	HS69E11	Homo sapiens DNA seque	4.98e+02
c	679	15	11.7	66021	17	AC005429	*** SEQUENCING IN PROG	4.98e+02
c	680	15	11.7	69034	18	AF007261	Reclinomonas americana	4.98e+02
	681	15	11.7	69591	26	AC003984	Human PAC clone DJ0566	4.98e+02
	682	15	11.7	72872	24	F1104	Arabidopsis thaliana B	4.98e+02
c	683	15	11.7	73840	23	AB007644	Arabidopsis thaliana g	4.98e+02
c	684	15	11.7	74155	26	AC004656	Homo sapiens Xp22 BAC	4.98e+02
	685	15	11.7	74935	27	AC005172	Homo sapiens chromosom	4.98e+02
	686	15	11.7	75475	23	AB005235	Arabidopsis thaliana g	4.98e+02
c	687	15	11.7	75793	26	HS321I15	Human DNA sequence fro	4.98e+02
	688	15	11.7	76619	18	AC004722	Drosophila melanogaste	4.98e+02
c	689	15	11.7	77183	26	HS299C21	Human DNA sequence fro	4.98e+02
	690	15	11.7	78419	18	AC004639	Drosophila melanogaste	4.98e+02
c	691	15	11.7	78756	17	AC004358	*** SEQUENCING IN PROG	4.98e+02
c	692	15	11.7	79204	26	AC003088	Human BAC clone RG119P	4.98e+02
	693	15	11.7	79248	18	AC004640	Drosophila melanogaste	4.98e+02
	694	15	11.7	79914	26	AC005159	Homo sapiens BAC clone	4.98e+02
c	695	15	11.7	80115	23	AB012248	Arabidopsis thaliana g	4.98e+02
	696	15	11.7	80124	18	AC002444	Drosophila melanogaste	4.98e+02
c	697	15	11.7	80446	18	AC004288	Drosophila melanogaste	4.98e+02
c	698	15	11.7	80497	26	HS27K14	Human DNA sequence fro	4.98e+02
c	699	15	11.7	80846	17	HS501A4	Human DNA sequence ***	4.98e+02
c	700	15	11.7	81004	23	ATAC002521	Arabidopsis thaliana c	4.98e+02
	701	15	11.7	81602	18	AC004349	Drosophila melanogaste	4.98e+02
	702	15	11.7	81786	26	AC003002	Human DNA from overlap	4.98e+02
c	703	15	11.7	82261	25	HSAC000065	Human BAC clone RG085C	4.98e+02
	704	15	11.7	82621	17	AC005467	*** SEQUENCING IN PROG	4.98e+02
c	705	15	11.7	83549	26	AC004752	Homo sapiens chromosom	4.98e+02

	706	15	11.7	85139	18	AC001657	Drosophila melanogaste	4.98e+02
	707	15	11.7	85338	18	AC004279	Drosophila melanogaste	4.98e+02
c	708	15	11.7	85444	26	AP000023	Homo sapiens genomic D	4.98e+02
c	709	15	11.7	85633	26	HSAC002088	Human BAC clone RG013P	4.98e+02
c	710	15	11.7	85785	24	ATF21P8	Arabidopsis thaliana D	4.98e+02
	711	15	11.7	86004	27	HS51J23	Human DNA sequence fro	4.98e+02
	712	15	11.7	86383	23	AB010697	Arabidopsis thaliana g	4.98e+02
	713	15	11.7	86491	26	AC004810	Homo sapiens chromosom	4.98e+02
c	714	15	11.7	86699	23	AB009050	Arabidopsis thaliana g	4.98e+02
c	715	15	11.7	87013	26	HS528L19	Human DNA sequence fro	4.98e+02
	716	15	11.7	87325	18	AC004375	Drosophila melanogaste	4.98e+02
c	717	15	11.7	87434	23	AB010698	Arabidopsis thaliana g	4.98e+02
	718	15	11.7	87552	17	HS217D16	Human DNA sequence ***	4.98e+02
c	719	15	11.7	87978	26	AC003971	Human BAC 367D17 from	4.98e+02
	720	15	11.7	88071	26	AC002316	Homo sapiens chromosom	4.98e+02
	721	15	11.7	88398	17	AB018117	*** SEQUENCING IN PROG	4.98e+02
	722	15	11.7	88871	28	MMU58105	Mus musculus Btk locus	4.98e+02
	723	15	11.7	90551	23	ATT5K18	Arabidopsis thaliana D	4.98e+02
	724	15	11.7	91286	18	CEY102A5C	Caenorhabditis elegans	4.98e+02
	725	15	11.7	93932	18	AC004766	Drosophila melanogaste	4.98e+02
	726	15	11.7	94482	24	ATF26P21	Arabidopsis thaliana D	4.98e+02
c	727	15	11.7	94590	17	AC004574	*** SEQUENCING IN PROG	4.98e+02
	728	15	11.7	95968	18	CEY47D3B	Caenorhabditis elegans	4.98e+02
	729	15	11.7	96180	17	AC005065	*** SEQUENCING IN PROG	4.98e+02
c	730	15	11.7	96217	26	AC002410	Human BAC clone RG264L	4.98e+02
c	731	15	11.7	96554	27	AC004783	Homo sapiens chromosom	4.98e+02
	732	15	11.7	96675	17	AC004287	*** SEQUENCING IN PROG	4.98e+02
c	733	15	11.7	96792	26	HSU95741	Human chromosome 16p13	4.98e+02
	734	15	11.7	96801	23	ATAC003033	Arabidopsis thaliana c	4.98e+02
	735	15	11.7	97137	26	AC004745	Homo sapiens BAC clone	4.98e+02
c	736	15	11.7	98178	17	AC005300	*** SEQUENCING IN PROG	4.98e+02
	737	15	11.7	98641	17	AC003688	*** SEQUENCING IN PROG	4.98e+02
	738	15	11.7	98940	17	AC003118	*** SEQUENCING IN PROG	4.98e+02
c	739	15	11.7	99204	27	HS143G3	Human DNA sequence fro	4.98e+02
c	740	15	11.7	100000	26	AP000030	Homo sapiens genomic D	4.98e+02
c	741	15	11.7	100000	26	AP000031	Homo sapiens genomic D	4.98e+02
c	742	15	11.7	100000	26	AP000051	Homo sapiens genomic D	4.98e+02
c	743	15	11.7	100000	26	AP000046	Homo sapiens genomic D	4.98e+02
	744	15	11.7	100258	26	AC002462	Human BAC clone RG210I	4.98e+02
	745	15	11.7	100808	26	HS94G16	Human DNA sequence fro	4.98e+02
	746	15	11.7	100849	25	HUMNEUROF	Human oligodendrocyte	4.98e+02
c	747	15	11.7	100927	17	AC005006	*** SEQUENCING IN PROG	4.98e+02
c	748	15	11.7	101311	26	HUAC004513	Homo sapiens Chromosom	4.98e+02
c	749	15	11.7	101733	17	AC005518	*** SEQUENCING IN PROG	4.98e+02
	750	15	11.7	102204	26	AC003103	Homo sapiens chromosom	4.98e+02
c	751	15	11.7	102717	26	AC004079	Homo sapiens PAC clone	4.98e+02
c	752	15	11.7	102846	27	AC002483	Homo sapiens PAC clone	4.98e+02
c	753	15	11.7	105119	17	AC004856	*** SEQUENCING IN PROG	4.98e+02
c	754	15	11.7	105127	26	AC004548	Homo sapiens PAC clone	4.98e+02
c	755	15	11.7	106209	15	D90903	Synechocystis sp. PCC6	4.98e+02
	756	15	11.7	106793	26	AC002538	Homo sapiens chromosom	4.98e+02

c	757	15	11.7	106793	26	AC002538	Homo sapiens chromosom 4.98e+02
c	758	15	11.7	108248	26	HS326I13	Homo sapiens DNA seque 4.98e+02
c	759	15	11.7	108260	26	HS179P9	Human DNA sequence fro 4.98e+02
c	760	15	11.7	108393	24	ATAC005315	Arabidopsis thaliana c 4.98e+02
	761	15	11.7	108534	26	HS124O9	Human DNA sequence fro 4.98e+02
	762	15	11.7	108695	26	AC002119	Homo sapiens chromosom 4.98e+02
	763	15	11.7	109302	26	HS11C7	Human DNA sequence fro 4.98e+02
c	764	15	11.7	110000	17	HSY214H10	Human DNA sequence *** 4.98e+02
c	765	15	11.7	110000	17	CEY73P8	Caenorhabditis elegans 4.98e+02
	766	15	11.7	110192	26	AC004070	Human Chromosome X, co 4.98e+02
c	767	15	11.7	110514	23	F7G19	Sequence of BAC F7G19 4.98e+02
	768	15	11.7	111352	26	HS329E20	Homo sapiens DNA seque 4.98e+02
	769	15	11.7	112018	17	HS128O17	Human DNA sequence *** 4.98e+02
c	770	15	11.7	112863	17	HS482L3	Human DNA sequence *** 4.98e+02
	771	15	11.7	115601	17	HS981L23	Human DNA sequence *** 4.98e+02
c	772	15	11.7	117752	26	AC004743	Homo sapiens PAC clone 4.98e+02
	773	15	11.7	117911	26	HS117P19	Human DNA sequence fro 4.98e+02
c	774	15	11.7	117911	26	HS117P19	Human DNA sequence fro 4.98e+02
	775	15	11.7	118001	17	HS566L20	Human DNA sequence *** 4.98e+02
c	776	15	11.7	118357	27	HS1142C11	Human DNA sequence fro 4.98e+02
c	777	15	11.7	118426	26	HS431C21	Human DNA sequence fro 4.98e+02
	778	15	11.7	118521	26	HUAC004514	Homo sapiens Chromosom 4.98e+02
c	779	15	11.7	118595	26	HS288L1	Human DNA sequence fro 4.98e+02
c	780	15	11.7	118741	27	AC005319	Human Chromosome 15q26 4.98e+02
	781	15	11.7	119198	26	HS845O24	Homo sapiens DNA seque 4.98e+02
	782	15	11.7	119509	17	HS238G2	Human DNA sequence *** 4.98e+02
	783	15	11.7	120000	17	HSAC000005	*** SEQUENCING IN PROG 4.98e+02
	784	15	11.7	121019	26	AF042091	Homo sapiens chromosom 4.98e+02
c	785	15	11.7	121055	17	AC004823	*** SEQUENCING IN PROG 4.98e+02
c	786	15	11.7	121290	17	CEH27C14	Caenorhabditis elegans 4.98e+02
c	787	15	11.7	121524	24	AF041468	Guillardia theta compl 4.98e+02
c	788	15	11.7	121567	17	AC005727	*** SEQUENCING IN PROG 4.98e+02
	789	15	11.7	121576	17	AC003117	*** SEQUENCING IN PROG 4.98e+02
	790	15	11.7	122322	23	ATF6I18	Arabidopsis thaliana D 4.98e+02
	791	15	11.7	122571	26	AF064859	Homo sapiens chromosom 4.98e+02
c	792	15	11.7	122577	17	AC004825	*** SEQUENCING IN PROG 4.98e+02
	793	15	11.7	123211	26	AC005208	Homo sapiens chromosom 4.98e+02
	794	15	11.7	123336	26	AC003075	Human PAC clone DJ0658 4.98e+02
	795	15	11.7	123386	24	F12F1	Arabidopsis thaliana c 4.98e+02
c	796	15	11.7	123386	24	F12F1	Arabidopsis thaliana c 4.98e+02
c	797	15	11.7	123397	26	AC004808	Homo sapiens PAC clone 4.98e+02
c	798	15	11.7	124181	17	HS285F3	Human DNA sequence *** 4.98e+02
c	799	15	11.7	124752	17	AC004986	*** SEQUENCING IN PROG 4.98e+02
	800	15	11.7	125179	24	ATT4L20	Arabidopsis thaliana D 4.98e+02
	801	15	11.7	125345	17	AC004818	*** SEQUENCING IN PROG 4.98e+02
c	802	15	11.7	125457	17	AC005045	*** SEQUENCING IN PROG 4.98e+02
c	803	15	11.7	125590	18	CEY49E10	Caenorhabditis elegans 4.98e+02
c	804	15	11.7	125927	27	AC005754	Homo sapiens chromosom 4.98e+02
c	805	15	11.7	126474	26	HS181N1	Human DNA sequence fro 4.98e+02
c	806	15	11.7	126502	17	HS708P5	Human DNA sequence *** 4.98e+02
	807	15	11.7	127356	17	AC005063	*** SEQUENCING IN PROG 4.98e+02

	808	15	11.7	127506	26	AC005244	Homo sapiens chromosom 4.98e+02
	809	15	11.7	127816	17	AC004822	*** SEQUENCING IN PROG 4.98e+02
c	810	15	11.7	128230	26	HS313L4	Human DNA sequence fro 4.98e+02
	811	15	11.7	128473	27	AC004241	Homo sapiens 12q13.1 P 4.98e+02
	812	15	11.7	128765	26	HUAC002040	Homo sapiens Chromosom 4.98e+02
c	813	15	11.7	128779	26	HS298K1	Human DNA sequence fro 4.98e+02
c	814	15	11.7	128993	17	AC005068	*** SEQUENCING IN PROG 4.98e+02
	815	15	11.7	130223	26	AF064863	Homo sapiens chromosom 4.98e+02
	816	15	11.7	130467	27	HS798A17	Human DNA sequence *** 4.98e+02
c	817	15	11.7	130632	26	AC004547	Homo sapiens PAC clone 4.98e+02
c	818	15	11.7	131223	17	AC004853	*** SEQUENCING IN PROG 4.98e+02
c	819	15	11.7	131458	26	AC002386	Human BAC clone RG343H 4.98e+02
c	820	15	11.7	131855	17	AC002518	*** SEQUENCING IN PROG 4.98e+02
	821	15	11.7	131903	26	HS360E18	Human DNA sequence fro 4.98e+02
	822	15	11.7	132790	26	HS167A14	Human DNA sequence fro 4.98e+02
	823	15	11.7	132953	26	AC003015	Human BAC clone GS113H 4.98e+02
c	824	15	11.7	133840	23	AC003981	Complete sequence of A 4.98e+02
	825	15	11.7	133894	21	A48542	Sequence 1 from Patent 4.98e+02
	826	15	11.7	134137	27	HS3E5	Homo sapiens DNA seque 4.98e+02
c	827	15	11.7	134580	26	AC002526	Human PAC clone DJ0558 4.98e+02
	828	15	11.7	134898	17	HS11O7	Human DNA sequence *** 4.98e+02
	829	15	11.7	135259	26	AC004590	Homo sapiens chromosom 4.98e+02
c	830	15	11.7	135599	23	CPU30821	Cyanophora paradoxa cy 4.98e+02
	831	15	11.7	135705	17	HS97P20	Human DNA sequence *** 4.98e+02
c	832	15	11.7	135705	17	HS97P20	Human DNA sequence *** 4.98e+02
	833	15	11.7	135886	26	HSXIST2	Human Xq13 3' end of P 4.98e+02
	834	15	11.7	136460	17	AC005007	*** SEQUENCING IN PROG 4.98e+02
	835	15	11.7	136503	17	AC004866	*** SEQUENCING IN PROG 4.98e+02
c	836	15	11.7	136503	17	AC004866	*** SEQUENCING IN PROG 4.98e+02
	837	15	11.7	136568	17	AC004933	*** SEQUENCING IN PROG 4.98e+02
c	838	15	11.7	137022	26	AC004231	Homo sapiens chromosom 4.98e+02
c	839	15	11.7	137065	17	HS29M10	Human DNA sequence *** 4.98e+02
	840	15	11.7	137817	26	AC003091	Human BAC clone RG326G 4.98e+02
	841	15	11.7	138181	17	AC005171	*** SEQUENCING IN PROG 4.98e+02
	842	15	11.7	138201	17	AC005016	*** SEQUENCING IN PROG 4.98e+02
c	843	15	11.7	139166	27	AC003029	Homo sapiens Chromosom 4.98e+02
c	844	15	11.7	139190	26	HS929C8	Human DNA sequence fro 4.98e+02
	845	15	11.7	139642	17	HS20N2	Human DNA sequence *** 4.98e+02
	846	15	11.7	140000	17	AC004047	*** SEQUENCING IN PROG 4.98e+02
c	847	15	11.7	140230	17	AC004845	*** SEQUENCING IN PROG 4.98e+02
c	848	15	11.7	140377	17	HS291J10	Human DNA sequence *** 4.98e+02
	849	15	11.7	142396	25	HSAC000120	Human BAC clone RG161K 4.98e+02
c	850	15	11.7	142661	17	AC005482	*** SEQUENCING IN PROG 4.98e+02
c	851	15	11.7	143330	17	AC005193	*** SEQUENCING IN PROG 4.98e+02
	852	15	11.7	143701	27	HS29C1E	Homo sapiens DNA seque 4.98e+02
	853	15	11.7	143819	26	AC002994	Homo sapiens chromosom 4.98e+02
c	854	15	11.7	144763	17	CEY40D7	Caenorhabditis elegans 4.98e+02
c	855	15	11.7	146322	17	HS108K11	Human DNA sequence *** 4.98e+02
	856	15	11.7	146963	26	HS964D12	Homo sapiens DNA seque 4.98e+02
	857	15	11.7	147055	17	HS372K1	Human DNA sequence *** 4.98e+02
	858	15	11.7	147705	17	AC005681	*** SEQUENCING IN PROG 4.98e+02

c	859	15	11.7	147724	27	HS395P12	Human DNA sequence *** 4.98e+02
	860	15	11.7	148114	17	HS25J6	Human DNA sequence *** 4.98e+02
	861	15	11.7	149169	17	HS1071N3	Human DNA sequence *** 4.98e+02
	862	15	11.7	149267	17	HS447C4	Human DNA sequence *** 4.98e+02
	863	15	11.7	149420	17	HS5O6	Human DNA sequence *** 4.98e+02
c	864	15	11.7	149566	26	AC003081	Human BAC clone RG009H 4.98e+02
c	865	15	11.7	150613	23	AB001684	Chlorella vulgaris C-2 4.98e+02
c	866	15	11.7	151111	26	AC002308	Homo sapiens Chromosom 4.98e+02
	867	15	11.7	151874	26	AC004594	Homo sapiens PAC clone 4.98e+02
c	868	15	11.7	151874	26	AC004594	Homo sapiens PAC clone 4.98e+02
	869	15	11.7	152115	17	HS404K8	Human DNA sequence *** 4.98e+02
c	870	15	11.7	152290	26	AC002357	Homo sapiens Xp22 Cosm 4.98e+02
c	871	15	11.7	152392	17	HS1018D12	Human DNA sequence *** 4.98e+02
c	872	15	11.7	152550	17	HSJ181N11	*** SEQUENCING IN PROG 4.98e+02
	873	15	11.7	153687	17	HS441J1	Human DNA sequence *** 4.98e+02
	874	15	11.7	153735	26	AC004029	Human BAC clone RG035E 4.98e+02
c	875	15	11.7	154028	26	HSAC002087	Human BAC clone GS113D 4.98e+02
c	876	15	11.7	154252	26	AC004062	Homo sapiens chromosom 4.98e+02
	877	15	11.7	154619	16	D90917	Synechocystis sp. PCC6 4.98e+02
	878	15	11.7	155147	26	HS992D9	Human DNA sequence fro 4.98e+02
c	879	15	11.7	155529	17	HS738P11	Human DNA sequence *** 4.98e+02
c	880	15	11.7	155952	26	AC004534	Homo sapiens PAC clone 4.98e+02
	881	15	11.7	158293	26	AC004746	Homo sapiens BAC clone 4.98e+02
	882	15	11.7	159245	17	AC004819	*** SEQUENCING IN PROG 4.98e+02
	883	15	11.7	159424	26	AF064865	Homo sapiens chromosom 4.98e+02
c	884	15	11.7	159424	26	AF064865	Homo sapiens chromosom 4.98e+02
	885	15	11.7	160178	17	HS422G23	Human DNA sequence *** 4.98e+02
	886	15	11.7	160526	17	HS1033E15	Human DNA sequence *** 4.98e+02
	887	15	11.7	160850	17	AC004957	*** SEQUENCING IN PROG 4.98e+02
	888	15	11.7	161973	26	HUAC002038	Homo sapiens chromosom 4.98e+02
c	889	15	11.7	162455	17	AC005536	*** SEQUENCING IN PROG 4.98e+02
c	890	15	11.7	162485	17	AC004840	*** SEQUENCING IN PROG 4.98e+02
	891	15	11.7	162531	26	AC005249	Homo sapiens BAC clone 4.98e+02
c	892	15	11.7	162750	17	AC003061	*** SEQUENCING IN PROG 4.98e+02
	893	15	11.7	163712	27	AC004055	Homo sapiens chromosom 4.98e+02
	894	15	11.7	163730	17	AC004852	*** SEQUENCING IN PROG 4.98e+02
c	895	15	11.7	163974	27	AB011399	Homo sapiens gene for 4.98e+02
	896	15	11.7	164277	17	HS1052M9	Human DNA sequence *** 4.98e+02
c	897	15	11.7	164277	17	HS1052M9	Human DNA sequence *** 4.98e+02
	898	15	11.7	164511	27	AC004894	Homo sapiens PAC clone 4.98e+02
	899	15	11.7	165590	17	HS782D21	Human DNA sequence *** 4.98e+02
c	900	15	11.7	165590	17	HS782D21	Human DNA sequence *** 4.98e+02
c	901	15	11.7	166978	26	AC004098	Homo sapiens chromosom 4.98e+02
c	902	15	11.7	167034	17	HS77O6	Human DNA sequence *** 4.98e+02
	903	15	11.7	168629	17	HS1045J21	Human DNA sequence *** 4.98e+02
	904	15	11.7	168636	17	AC002060	*** SEQUENCING IN PROG 4.98e+02
	905	15	11.7	170000	17	AC005509	*** SEQUENCING IN PROG 4.98e+02
c	906	15	11.7	170000	17	AC004524	*** SEQUENCING IN PROG 4.98e+02
c	907	15	11.7	170342	17	AC003668	*** SEQUENCING IN PROG 4.98e+02
c	908	15	11.7	170392	17	HS740A11	Human DNA sequence *** 4.98e+02
c	909	15	11.7	171905	26	HSAC002080	Human BAC clone RG253B 4.98e+02

	910	15	11.7	172008	17	AC005011	*** SEQUENCING IN PROG 4.98e+02
	911	15	11.7	172241	26	AC004014	Human BAC clone RG022K 4.98e+02
c	912	15	11.7	172533	26	AC002381	Human BAC clone RG020D 4.98e+02
	913	15	11.7	172733	17	AC005082	*** SEQUENCING IN PROG 4.98e+02
	914	15	11.7	174505	17	HS197017	Human DNA sequence *** 4.98e+02
c	915	15	11.7	175488	17	AC004887	*** SEQUENCING IN PROG 4.98e+02
	916	15	11.7	176968	27	AC005669	Homo sapiens chromosom 4.98e+02
c	917	15	11.7	177821	17	AC005030	*** SEQUENCING IN PROG 4.98e+02
	918	15	11.7	177901	27	AC005084	Homo sapiens BAC clone 4.98e+02
c	919	15	11.7	177901	27	AC005084	Homo sapiens BAC clone 4.98e+02
c	920	15	11.7	179076	17	AC004890	*** SEQUENCING IN PROG 4.98e+02
c	921	15	11.7	179246	17	AC004854	*** SEQUENCING IN PROG 4.98e+02
	922	15	11.7	179707	17	HSAC001234	*** SEQUENCING IN PROG 4.98e+02
	923	15	11.7	180136	15	BAC180R	B. subtilis DNA, 180 k 4.98e+02
c	924	15	11.7	180565	17	HS149A16	Human DNA sequence *** 4.98e+02
c	925	15	11.7	181978	17	AC005058	*** SEQUENCING IN PROG 4.98e+02
	926	15	11.7	184289	17	AC004828	*** SEQUENCING IN PROG 4.98e+02
	927	15	11.7	184490	26	HSU82828	Homo sapiens ataxia te 4.98e+02
c	928	15	11.7	185364	17	AC005477	*** SEQUENCING IN PROG 4.98e+02
	929	15	11.7	185371	17	HS625H18	Human DNA sequence *** 4.98e+02
c	930	15	11.7	185893	26	AC005191	Homo sapiens PAC clone 4.98e+02
	931	15	11.7	186454	17	AC005678	*** SEQUENCING IN PROG 4.98e+02
c	932	15	11.7	187051	17	AC004688	*** SEQUENCING IN PROG 4.98e+02
c	933	15	11.7	188362	27	AC005019	Homo sapiens BAC clone 4.98e+02
	934	15	11.7	188636	26	HUAC002287	Homo sapiens Chromosom 4.98e+02
c	935	15	11.7	188741	26	AC002542	Human BAC clone RG114A 4.98e+02
	936	15	11.7	189009	17	AC005746	*** SEQUENCING IN PROG 4.98e+02
c	937	15	11.7	189999	17	AC004617	*** SEQUENCING IN PROG 4.98e+02
c	938	15	11.7	190444	17	AC005771	*** SEQUENCING IN PROG 4.98e+02
c	939	15	11.7	191423	17	AC004941	*** SEQUENCING IN PROG 4.98e+02
c	940	15	11.7	191857	17	CEY32B12	Caenorhabditis elegans 4.98e+02
c	941	15	11.7	192609	17	AC005345	*** SEQUENCING IN PROG 4.98e+02
c	942	15	11.7	192730	26	HUU95743	Homo sapiens chromosom 4.98e+02
	943	15	11.7	192802	26	AC005284	Homo sapiens chromosom 4.98e+02
c	944	15	11.7	193161	27	AC005500	, complete sequence. 4.98e+02
	945	15	11.7	193707	17	AC005035	*** SEQUENCING IN PROG 4.98e+02
c	946	15	11.7	196405	17	AC005053	*** SEQUENCING IN PROG 4.98e+02
	947	15	11.7	198285	27	HSEVMHC	H.sapiens DMA, DMB, HL 4.98e+02
	948	15	11.7	200000	17	AC004624	*** SEQUENCING IN PROG 4.98e+02
c	949	15	11.7	200000	17	AC004618	*** SEQUENCING IN PROG 4.98e+02
c	950	15	11.7	200000	17	AC005141	*** SEQUENCING IN PROG 4.98e+02
c	951	15	11.7	200216	23	ATFCA3	Arabidopsis thaliana D 4.98e+02
	952	15	11.7	202763	26	AC005181	Homo sapiens chromosom 4.98e+02
	953	15	11.7	203418	17	AC004947	*** SEQUENCING IN PROG 4.98e+02
c	954	15	11.7	203583	17	AC004940	*** SEQUENCING IN PROG 4.98e+02
c	955	15	11.7	204877	17	AC005375	*** SEQUENCING IN PROG 4.98e+02
c	956	15	11.7	205559	23	ATFCA4	Arabidopsis thaliana D 4.98e+02
	957	15	11.7	206553	17	AC005725	*** SEQUENCING IN PROG 4.98e+02
c	958	15	11.7	207674	23	ATFCA8	Arabidopsis thaliana D 4.98e+02
	959	15	11.7	207957	26	AC004470	Homo sapiens Xp22 BAC 4.98e+02
c	960	15	11.7	208396	26	AC003101	Homo sapiens chromosom 4.98e+02

	961	15	11.7	209165	23	ATFCA5	Arabidopsis thaliana D 4.98e+02
c	962	15	11.7	210000	17	AC005699	*** SEQUENCING IN PROG 4.98e+02
c	963	15	11.7	210000	17	AC004555	*** SEQUENCING IN PROG 4.98e+02
c	964	15	11.7	210000	17	AC005690	*** SEQUENCING IN PROG 4.98e+02
	965	15	11.7	211792	17	HS216E10	Human DNA sequence *** 4.98e+02
	966	15	11.7	213080	15	BSUB0001	Bacillus subtilis comp 4.98e+02
c	967	15	11.7	217538	17	HS109G6	Human DNA sequence *** 4.98e+02
	968	15	11.7	222930	26	HSU47924	Human chromosome 12p13 4.98e+02
	969	15	11.7	224129	17	CEY7A9	Caenorhabditis elegans 4.98e+02
c	970	15	11.7	224525	17	CEY70C5	Caenorhabditis elegans 4.98e+02
	971	15	11.7	226063	27	AC005406	, complete sequence. 4.98e+02
c	972	15	11.7	226835	17	CEY49A3	Caenorhabditis elegans 4.98e+02
c	973	15	11.7	227215	17	HS245G19	Human DNA sequence *** 4.98e+02
c	974	15	11.7	227892	17	AC005587	*** SEQUENCING IN PROG 4.98e+02
	975	15	11.7	228516	26	AC002523	Homo sapiens Xq28 BACs 4.98e+02
c	976	15	11.7	231000	17	AC000015	*** SEQUENCING IN PROG 4.98e+02
c	977	15	11.7	233000	16	AP000003	Pyrococcus horikoshii 4.98e+02
c	978	15	11.7	236281	26	AC004673	Homo sapiens Xp22 BAC 4.98e+02
c	979	15	11.7	238670	17	AC003059	*** SEQUENCING IN PROG 4.98e+02
	980	15	11.7	244362	17	HS618P1	Human DNA sequence *** 4.98e+02
c	981	15	11.7	244455	17	AC002470	*** SEQUENCING IN PROG 4.98e+02
c	982	15	11.7	244582	17	CEY40H7	Caenorhabditis elegans 4.98e+02
	983	15	11.7	245265	17	HS581F12	Human DNA sequence *** 4.98e+02
c	984	15	11.7	247274	17	HS347H13	Human DNA sequence *** 4.98e+02
	985	15	11.7	254200	17	CEY18D10	Caenorhabditis elegans 4.98e+02
c	986	15	11.7	254200	17	CEY18D10	Caenorhabditis elegans 4.98e+02
	987	15	11.7	259202	26	AC002366	Human Xp22 BAC CT-285I 4.98e+02
c	988	15	11.7	259474	26	HUAC004605	Homo sapiens Chromosom 4.98e+02
c	989	15	11.7	259894	26	HUAC002302	Homo sapiens Chromosom 4.98e+02
	990	15	11.7	267156	26	U66059	Human germline T-cell 4.98e+02
	991	15	11.7	269078	17	CEY76A2	Caenorhabditis elegans 4.98e+02
	992	15	11.7	286715	17	CEY45F3	Caenorhabditis elegans 4.98e+02
c	993	15	11.7	287986	17	CEY51H4	Caenorhabditis elegans 4.98e+02
	994	15	11.7	305444	17	CEY67H2	Caenorhabditis elegans 4.98e+02
	995	15	11.7	310213	17	AC004526	*** SEQUENCING IN PROG 4.98e+02
c	996	15	11.7	315339	23	SCCHRIII	S.cerevisiae chromosom 4.98e+02
	997	15	11.7	317311	17	AC004085	*** SEQUENCING IN PROG 4.98e+02
	998	15	11.7	323932	17	AC005682	*** SEQUENCING IN PROG 4.98e+02
	999	15	11.7	324816	26	AC002368	Homo sapiens Xq28 BAC 4.98e+02
	1000	15	11.7	337546	17	CEY47D3	Caenorhabditis elegans 4.98e+02

ALIGNMENTS

```
RESULT    1
LOCUS       AC005386   196475 bp    DNA             HTG       10-SEP-1998
DEFINITION  *** SEQUENCING IN PROGRESS *** Homo sapiens; HTGS phase 1, 1
            unordered pieces.
ACCESSION   AC005386
NID         g3581711
KEYWORDS    HTG; HTGS_PHASE1.
```



```
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 196475)
  AUTHORS   Smith,D.R.
  TITLE     Sequencing of Human Chromosome 10
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 196475)
  AUTHORS   Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-AUG-1998) Genome Therapeutics Corporation, 100 Beaver
            Street, Waltham, MA 02154, USA
COMMENT     On Sep 10, 1998 this sequence version replaced gi:3395412.
            ***                                                        ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                        ***
            * This sequence is unfinished. It consists of 1 contigs for
            * which the order is not known; their order in this record is
            * arbitrary. In some cases, the exact lengths of the gaps
            * between the contigs are also unknown; these gaps are  presented
            * as runs of N as a convenience only. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
            *         1   196475: contig of 196475 bp in length.
FEATURES              Location/Qualifiers
     source           1. .196475
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
BASE COUNT    55773 a  41494 c  40885 g  58319 t      4 others
ORIGIN

  Query Match           15.6%;  Score 20;  DB 17;  Length 196475;
  Best Local Similarity 100.0%;  Pred. No. 3.65e-01;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db 184376 ATTTTTAAAGCAGAATTTTG 184395
          ||||||||||||||||||||
Qy      8 ATTTTTAAAGCAGAATTTTG 27


RESULT    2
LOCUS       I70241       1523 bp    DNA             PAT       23-DEC-1997
DEFINITION  Sequence 25 from patent US 5679521.
ACCESSION   I70241
NID         g3006376
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1523)
```



```
  AUTHORS   Andersson,S. and Russell,D.W.
  TITLE     Steroid 5A-reductases
  JOURNAL   Patent: US 5679521-A 25 21-OCT-1997;
FEATURES              Location/Qualifiers
     source           1. .1523
                      /organism="unknown"
BASE COUNT      443 a    360 c    356 g    364 t
ORIGIN

  Query Match           14.8%;  Score 19;  DB 21;  Length 1523;
  Best Local Similarity 100.0%;  Pred. No. 1.71e+00;
  Matches    19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    567 AAAAGAATAAAGAATATCT 585
          |||||||||||||||||||
Qy     99 AAAAGAATAAAGAATATCT 117


RESULT    3
LOCUS       I12338       1523 bp    DNA             PAT       14-JUN-1995
DEFINITION  Sequence 25 from patent US 5422262.
ACCESSION   I12338
NID         g910361
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1523)
  AUTHORS   Andersson,S. and Russell,D.W.
  TITLE     Steroid 5a reductase nucleic acid segments and recombinant vectors
            and host cells
  JOURNAL   Patent: US 5422262-A 25 06-JUN-1995;
FEATURES              Location/Qualifiers
     source           1. .1523
                      /organism="unknown"
BASE COUNT      443 a    360 c    356 g    364 t
ORIGIN

  Query Match           14.8%;  Score 19;  DB 21;  Length 1523;
  Best Local Similarity 100.0%;  Pred. No. 1.71e+00;
  Matches    19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    567 AAAAGAATAAAGAATATCT 585
          |||||||||||||||||||
Qy     99 AAAAGAATAAAGAATATCT 117


RESULT    4
LOCUS       HUMSRD5AP    1645 bp    DNA             PRI       13-JAN-1995
DEFINITION  H.sapiens steroid reductase pseudogene.
ACCESSION   M60887
```



```
NID         g338467
KEYWORDS    androgen; pseudogene; sexual differentiation; steroid metabolism.
SOURCE      Homo sapiens DNA.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1645)
  AUTHORS   Jenkins,E.P., Hsieh,C.L., Milatovich,A., Normington,K.,
            Berman,D.M., Francke,U. and Russell,D.W.
  TITLE     Characterization and chromosomal mapping of a human steroid 5
            alpha-reductase gene and pseudogene and mapping of the mouse
            homologue
  JOURNAL   Genomics 11 (4), 1102-1112 (1991)
  MEDLINE   92147107
FEATURES              Location/Qualifiers
     source           1. .1645
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /map="Xq24-qter"
     repeat_region    1. .614
                      /note="line"
                      /rpt_type=other
     repeat_region    615. .626
                      /rpt_type=direct
     CDS              701. .1486
                      /gene="SRD5AP1"
                      /pseudo
                      /db_xref="GDB:G00-127-344"
                      /codon_start=1
     gene             701. .1537
                      /gene="SRD5AP1"
     polyA_site       1530. .1537
                      /gene="SRD5AP1"
                      /note="G00-127-344"
     repeat_region    1538. .1549
                      /rpt_type=direct
BASE COUNT      473 a    388 c    379 g    403 t      2 others
ORIGIN

  Query Match           14.8%;  Score 19;  DB 25;  Length 1645;
  Best Local Similarity 100.0%;  Pred. No. 1.71e+00;
  Matches    19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    567 AAAAGAATAAAGAATATCT 585
          |||||||||||||||||||
Qy     99 AAAAGAATAAAGAATATCT 117


RESULT    5
LOCUS       CEW04A8     36599 bp    DNA             HTG       27-NOV-1996
DEFINITION  *** SEQUENCING IN PROGRESS *** Caenorhabditis elegans cosmid W04A8;
```

```
            HTGS phase 1.
ACCESSION   Z82069
NID         g1695111
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Nematoda;
            Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditidae;
            Caenorhabditis.
REFERENCE   1  (bases 1 to 36599)
  AUTHORS   Unknown.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-NOV-1996) Nematode Sequencing Project, Sanger Centre,
            Hinxton, Cambridge CB10 1RQ, England and Department of Genetics,
            Washington University, St. Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     On Nov 29, 1996 this sequence version replaced gi:1666626.
            Order of segments is not known; 800 n's separate segments.
            Cosmid=W04A8; Contig ID=00972; Length=4928; Order=Unknown;
            Status=Cambridg
            e-Unfinished; Author=Kershaw JK; LastModification=Nov 25 09:51
            Cosmid=W04A8; Contig ID=00853; Length=1753; Order=Unknown;
            Status=Cambridg
            e-Unfinished; Author=Kershaw JK; LastModification=Nov 25 09:51
            Cosmid=W04A8; Contig ID=00162; Length=1015; Order=Unknown;
            Status=Cambridg
            e-Unfinished; Author=Kershaw JK; LastModification=Nov 25 09:51
            Cosmid=W04A8; Contig ID=00343; Length=1321; Order=Unknown;
            Status=Cambridg
            e-Unfinished; Author=Kershaw JK; LastModification=Nov 25 09:51
            Cosmid=W04A8; Contig ID=00792; Length=19343; Order=Unknown;
            Status=Cambrid
            ge-Unfinished; Author=Kershaw JK; LastModification=Nov 25 09:51
            Cosmid=W04A8; Contig ID=00769; Length=1101; Order=Unknown;
            Status=Cambridg
            e-Unfinished; Author=Kershaw JK; LastModification=Nov 25 09:51
            Cosmid=W04A8; Contig ID=00894; Length=2872; Order=Unknown;
            Status=Cambridg
            e-Unfinished; Author=Kershaw JK; LastModification=Nov 25 09:51.
            ***                                                        ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                        ***
            * This sequence is unfinished. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
FEATURES              Location/Qualifiers
     source           1. .36599
                      /organism="Caenorhabditis elegans"
                      /db_xref="taxon:6239"
                      /clone="W04A8"
                      /chromosome="I"
```



```
BASE COUNT    10280 a   5780 c   5467 g  10269 t   4803 others
ORIGIN

  Query Match           14.8%;  Score 19;  DB 17;  Length 36599;
  Best Local Similarity 100.0%;  Pred. No. 1.71e+00;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  28336 AAATTCTGCTTTAAAAATT 28354
          |||||||||||||||||||
Cp     25 AAATTCTGCTTTAAAAATT 7


RESULT    6
LOCUS       HS93C23     72245 bp    DNA              PRI       12-FEB-1998
DEFINITION  Human DNA sequence from PAC 93C23 on chromosome X. Contains steroid
            5-alpha-reductase pseudogene, ESTs and STS.
ACCESSION   AL008713
NID         g2887305
KEYWORDS    3-oxo-5-alpha-steroid delta(4)-dehydrogenase; dihydrotestosterone;
            pseudogene; X.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 72245)
  AUTHORS   Pavitt,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-FEB-1998) sanger.ac.uk/HGP/ChrX/) Sanger Centre,
            Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquires:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Feb 14, 1998 this sequence version replaced gi:2598473.
            IMPORTANT: This sequence is the entire insert of clone 93C23.
            During sequence assembly data are compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variations annotated may not be found in the sequence submission
            corresponding to the overlapping clone as we submit sequences with
            only a small overlap. This sequence was generated from part of
            bacterial clone contigs of human chromosome X, constructed by the
            Sanger Centre chromosome X mapping group. Further information can
            be found at http://www.sanger.ac.uk/HGP/ChrX/
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            The true left end of clone 93C23 is at 1 in this sequence. The true
            right end of clone 93C23 is at 72245.
            93C23 is from the library RPCI1 constructed at the Roswell Park
            Cancer Institute by the group of Pieter de Jong.
```



```
            For further details see http://bacpac.med.buffalo.edu/.
FEATURES              Location/Qualifiers
     source           1. .72245
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="X"
                      /map="X"
                      /clone="93C23"
                      /clone_lib="RPCI1"
     repeat_region    2. .1624
                      /note="L1 repeat: matches 3885. .2250 of consensus"
     repeat_region    1648. .1757
                      /note="L1 repeat: matches 2975. .3077 of consensus"
     repeat_region    2425. .3266
                      /note="L1MA9 repeat: matches 1030. .223 of consensus"
     repeat_region    3271. .3491
                      /note="L1PA8 repeat: matches 910. .691 of consensus"
     repeat_region    5973. .6090
                      /note="AluSq repeat: matches 1. .127 of consensus;
                      incomplete repeat"
     repeat_region    6316. .6600
                      /note="AluSx repeat: matches 1. .295 of consensus"
     repeat_region    6601. .6628
                      /note="14 copies of 2 mer 100 % conserved"
     repeat_region    7152. .7445
                      /note="AluSq repeat: matches 1. .303 of consensus"
     repeat_region    7853. .7937
                      /note="MLT1A1 repeat: matches 9. .92 of consensus"
     repeat_region    7940. .8159
                      /note="MLT1A1 repeat: matches 140. .363 of consensus"
     repeat_region    8389. .8484
                      /note="2 copies of 48 mer 89 % conserved"
     repeat_region    9295. .9520
                      /note="MIR repeat: matches 20. .254 of consensus"
     repeat_region    12225. .12286
                      /note="L1MA8 repeat: matches 951. .1012 of consensus"
     repeat_region    14161. .14351
                      /note="MIR repeat: matches 5. .200 of consensus"
     repeat_region    15520. .15774
                      /note="MIR repeat: matches 2. .260 of consensus"
     repeat_region    17763. .18023
                      /note="MIR repeat: matches 262. .1 of consensus"
     repeat_region    18541. .18579
                      /note="13 copies of 3 mer 85 % conserved"
     repeat_region    18752. .18841
                      /note="MIR2 repeat: matches 127. .35 of consensus"
     repeat_region    18842. .19284
                      /note="MLT2B repeat: matches 1. .444 of consensus"
     repeat_region    19286. .19541
                      /note="MLT2_internal repeat: matches 2. .272 of consensus"
     repeat_region    19536. .20413
```



```
                      /note="MLT2_internal repeat: matches 303. .1179 of
                      consensus"
     repeat_region    20391. .20780
                      /note="L1PA5 repeat: matches 511. .892 of consensus"
     repeat_region    20788. .25258
                      /note="MLT2_internal repeat: matches 1163. .5669 of
                      consensus"
     repeat_region    25260. .25375
                      /note="AluSx repeat: matches 182. .302 of consensus;
                      incomplete repeat"
     repeat_region    25839. .25877
                      /note="13 copies of 3 mer 85 % conserved"
     misc_feature     26906. .27049
                      /note="match: STS L41283"
     repeat_region    27051. .27226
                      /note="MIR repeat: matches 20. .214 of consensus"
     repeat_region    27287. .27363
                      /note="MIR2 repeat: matches 146. .64 of consensus"
     repeat_region    27375. .27474
                      /note="MIR repeat: matches 200. .97 of consensus"
     repeat_region    29553. .30293
                      /note="MLT2_internal repeat: matches 1739. .2482 of
                      consensus"
     repeat_region    30349. .31106
                      /note="L1MA2 repeat: matches 1055. .288 of consensus"
     repeat_region    31095. .34187
                      /note="L1 repeat: matches 5193. .2074 of consensus"
     repeat_region    34668. .34966
                      /note="L1 repeat: matches 1986. .1684 of consensus"
     prim_transcript  <35116. .>35420
                      /note="match: multiple ESTs; match: AA411794 N58388"
     repeat_region    35584. .35861
                      /note="MER43 repeat: matches 1. .272 of consensus"
     repeat_region    36469. .36533
                      /note="MER43 repeat: matches 208. .272 of consensus"
     repeat_region    37602. .37647
                      /note="23 copies of 2 mer 100 % conserved"
     repeat_region    37664. .38824
                      /note="MLT2_internal repeat: matches 2623. .3806 of
                      consensus"
     repeat_region    38845. .39127
                      /note="MLT2_internal repeat: matches 3859. .4139 of
                      consensus"
     repeat_region    39124. .39484
                      /note="THE1B repeat: matches 1. .363 of consensus"
     repeat_region    39486. .40097
                      /note="MLT2_internal repeat: matches 4131. .4757 of
                      consensus"
     repeat_region    40092. .40484
                      /note="MLT1B repeat: matches 1. .386 of consensus"
     repeat_region    40488. .40728
```

```
                     /note="KLT2_internal repeat: matches 4746. .5004 of
                     consensus"
     repeat_region   42115. .42255
                     /note="MIR repeat: matches 251. .110 of consensus"
     repeat_region   42666. .42845
                     /note="MER5B repeat: matches 178. .1 of consensus"
     repeat_region   42976. .43404
                     /note="L1MA5 repeat: matches 627. .1041 of consensus"
     repeat_region   43399. .43519
                     /note="MER5A repeat: matches 60. .189 of consensus"
     misc_feature    44752. .45014
                     /note="match: STS G03850"
     repeat_region   46546. .46681
                     /note="MER3 repeat: matches 209. .71 of consensus"
     repeat_region   46679. .46837
                     /note="MER3 repeat: matches 164. .7 of consensus"
     repeat_region   46868. .46912
                     /note="15 copies of 3 mer 82 % conserved"
     repeat_region   49055. .49151
                     /note="MIR repeat: matches 141. .45 of consensus"
     repeat_region   49903. .49952
                     /note="MIR2 repeat: matches 144. .94 of consensus"
     repeat_region   50377. .50485
                     /note="MIR2 repeat: matches 27. .141 of consensus"
     repeat_region   51949. .52046
                     /note="L1MB6 repeat: matches 919. .820 of consensus"
     repeat_region   53131. .53402
                     /note="MER42c repeat: matches 565. .286 of consensus"
     repeat_region   53211. .53595
                     /note="MER42B repeat: matches 484. .57 of consensus"
     repeat_region   54595. .54618
                     /note="12 copies of 2 mer 100 % conserved"
     repeat_region   55697. .56128
                     /note="L1MD1 repeat: matches 531. .68 of consensus"
     repeat_region   55810. .56043
                     /note="L1ME3A repeat: matches 392. .159 of consensus"
     repeat_region   56569. .58307
                     /note="L1 repeat: matches 4423. .2683 of consensus"
     repeat_region   58314. .63702
                     /note="L1 repeat: matches 4. .5390 of consensus"
     repeat_region   63553. .64445
                     /note="L1PA2 repeat: matches 1. .893 of consensus"
     repeat_region   64469. .64788
                     /note="L1MB7 repeat: matches 920. .601 of consensus"
     repeat_region   64796. .64893
                     /note="L1MB7 repeat: matches 284. .186 of consensus"
     repeat_region   64987. .65311
                     /note="L1 repeat: matches 1665. .1982 of consensus"
     repeat_region   66591. .67399
                     /note="L1 repeat: matches 860. .1690 of consensus"
     repeat_region   67888. .68609
```



```
                     /note="L1 repeat: matches 4671. .5390 of consensus"
     repeat_region   68460. .69351
                     /note="L1PA2 repeat: matches 1. .891 of consensus"
     mRNA            <69475. .>71390
                     /note="steroid 5-alpha-reductase mRNA; match: M32313;
                     match: PIR; A35856"
     CDS             69505. .70290
                     /note="steroid 5-alpha-reductase; match: M32313"
                     /codon_start=1
...
Note: remainder of annotations omitted.

  Query Match           14.8%;  Score 19;  DB 26;  Length 72245;
  Best Local Similarity 100.0%;  Pred. No. 1.71e+00;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  69370 AAAAGAATAAAGAATATCT 69388
          |||||||||||||||||||
Qy      99 AAAAGAATAAAGAATATCT 117


RESULT    7
LOCUS       CEY102A5C   91286 bp    DNA             INV       22-SEP-1998
DEFINITION  Caenorhabditis elegans cosmid Y102A5C, complete sequence.
ACCESSION   AL031627
NID         g3646746
KEYWORDS    HTG.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 91286)
  AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
  TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
  JOURNAL   Nature 368 (6466), 32-38 (1994)
  MEDLINE   94150718
REFERENCE   2  (bases 1 to 91286)
  AUTHORS   Gardner,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-NOV-1996) Nematode Sequencing Project, Sanger Centre,
```



```
            Hinxton, Cambridge CB10 1RQ, England and Department of Genetics,
            Washington University, St. Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
  COMMENT   Coding sequences below are predicted from computer analysis, using
            predictions from Genefinder (P. Green, U. Washington), and other
            available information.
            For a graphical representation of this sequence and its analysis
            see:-
            http://webace.sanger.ac.uk/cgi-
            bin/display?db=wormace&class=Sequence &object=Y102A5C
            Current sequence finishing criteria for the C. elegans genome
            sequencing consortium are that all bases are either sequenced
            unambiguously on both strands, or on a single strand with both a
            dye primer and dye terminator reaction, from distinct subclones.
            Exceptions are indicated by an explicit note.
            IMPORTANT: This sequence is NOT necessarily the entire insert of
            the specified clone.  It may be shorter because we only sequence
            overlapping sections once, or longer because we arrange for a small
            overlap between neighbouring submissions.
            IMPORTANT: This sequence is not the entire insert of clone Y102A5C.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we arrange for a small overlap between
            neighbouring submissions.
            The true left end of clone F49H6 is at 91184 in this sequence. The
            true right end of clone F09C6 is at 102 in this sequence. The start
            of this sequence (1. .102) overlaps with the end of sequence Z81496.
            The end of this sequence (91183. .91286) overlaps with the start of
            sequence Z81545.
  FEATURES  Location/Qualifiers
     source 1. .91286
            /organism="Caenorhabditis elegans"
            /db_xref="taxon:6239"
            /clone="Y102A5C"
     gene   770. .1535
            /gene="Y102A5C.2"
     CDS    join(770. .825,972. .1147,1199. .1330,1381. .1535)
            /gene="Y102A5C.2"
            /note="predicted using Genefinder"
            /codon_start=1
            /db_xref="PID:e1323285"
            /db_xref="PID:g3646747"
            /translation="MRTTIKVYLWLRYCLIFYCGFCAPKDTRKIIKYWFLTKILRRTV
            PRPQLDGIDNSVHQDDLNAGVDAERQPVPPGDAVEIEENQNDEVREAAENHSNREEDQ
            NVENQPIPPGDAEIRVQIEDENGENQDNNIIVINPPNRQNVEAEPDAEQPQFDEELTI
            RHDAIQVKPKRS"
     gene   complement(1770. .3311)
            /gene="Y102A5C.3"
     CDS    complement(join(1770. .2015,2064. .2279,2446. .2722,
            2981. .3064,3115. .3311))
            /gene="Y102A5C.3"
            /note="predicted using Genefinder"
```



```
            /codon_start=1
            /db_xref="PID:e1323286"
            /db_xref="PID:g3646748"
            /translation="MSASLRETELSIRTCILYEALDDKPIEESFKNFCRKVGKGAMTL
            YDFDTWFYRFHNGNHDLHHDRSKDPKQLTLSDMPIGVMPEILSNVDPIDRLVLRKVSQ
            GLRHVVKSIGWEFDTLTPIKDTRFPRIHFPSSQTIYNSKENGCSVTQWREFGDGKPSQ
            FRKFVPGGDKSELLYNDCKILLENKLVEPESLEELDFEIRNGSIGYEHVVWTEQWKM
            AKYLKSWEETSIPIEHLLHFTTPTVRLTEFTVEDARKIRDMLDTSGTFEKAKITVKTI
            DSIEIARVFVPDFELGADENGNEKFYYTTAKGRMFRIGFSSCGNVKKLTIKYLDSDDS
            EDSDY"
     gene   4240. .6246
            /gene="Y102A5C.4"
     CDS    join(4240. .4506,4603. .4969,5014. .5180,5228. .5356,
            5408. .5652,5769. .6246)
            /gene="Y102A5C.4"
            /note="predicted using Genefinder"
            /codon_start=1
            /db_xref="PID:e1323287"
            /db_xref="PID:g3646749"
            /translation="MLLNLFVFLVIGMFLKCIKRKCQLISEDVQQTLGSLTNVGEYVI
            ELADNVAPTNLTRKFVVLSNTIGKYLEGPKLQPRNELTLSTRVELEDLVVVVPKKYSK
            LDDLKYAKKFPTSSVEILKLTVVTVTAMIERCSKYTGIFMSEYVQLALGSIMTLGGIVI
            GWLLKPSEGLNEKMGELNKEFEKISKNMNSRLEDLQSSRFSPEFFKVPFSDFFLFSKS
            SKNSKKNQEAATAVTLMTYMQRTISHPSKENIRNFLRECEKTSPGHLGLKLAHMLRDT
            STNPLKMAMDAQKVPKKAVPEKWRETLDAVLTQCLVLEAFYNGMRGGGKDEMDMLNYK
            IKEHEEQIVCLRAMYHSRTTYWAYMQMNIHNVQDVPAVCSNESKAGILKDTLEEIMTN
            DAFYIIVFDYMSLFGQFIGVYKCYGFSYADEKNVLSSYNRGNCNVVLFRSVGPQRQKT
            YSQQLRTLLITHQPTLVNMPLYSWIPQNNYLYTTTATKDIRAKLGSNCGFVGIVCVAQ
            KPCVEWANCDPEFGPGHFTTEVLEPNNDAYEFSKYLMIAGIC"
     gene   13994. .15543
            /gene="Y102A5C.7"
     CDS    join(13994. .14072,14727. .14785,14847. .14952,15152. .15347,
            15396. .15443,15492. .15543)
            /gene="Y102A5C.7"
            /codon_start=1
            /db_xref="PID:e1323295"
            /db_xref="PID:g3646757"
            /translation="MFPAILIFLLLLPTLINACIPTQQVEIPCPTDWEQFERPSGTWC
            IRVFVGSGPRSNADSMCSAQGAVVSGIQNQEEIDFIVGSWISMYGSTGSMWVGAQRTA
            ACWTLQGQTAACSKTTSFEWTDNSATGTDGYVWNSAVEPNNVGGTEQCLVLTWTGLMS
            DQTCARTDYVGYVCGKAPM"
     gene   19263. .20472
            /gene="Y102A5C.8"
     CDS    join(19263. .19351,19402. .19568,19995. .20257,20305. .20472)
            /gene="Y102A5C.8"
            /codon_start=1
            /db_xref="PID:e1323296"
            /db_xref="PID:g3646758"
            /translation="MFFSAHAPSLSGMPLDIINRILDKLDPILRFSLRRVCRNLRNVI
            DNRDPEIEQVHVLFQSNQSVLTINSTFEFDYNLYAILDDLATLDVLTPLPLFKPGILE
            EIVITTFEPIENDEAFDIIETEQWKQAKSITLHSVNTLSYFTFPIEHLFHLCKFSVPL
```

```
                VELKISDAVKIRDVLLKSVHPEAGQINVKNKIQDDVLRVFNPEFNDNIFRYDNPELTF
                KLNSLNVKKI"
     gene       21019. .22134
                /gene="Y102A5C.9"
     CDS        join(21019. .21059,21247. .21889,21940. .22134)
                /gene="Y102A5C.9"
                /note="predicted using Genefinder"
                /codon_start=1
                /db_xref="PID:e1323288"
                /db_xref="PID:g3646750"
                /translation="MTDNLLNQLESCKICILRGVCKKLKNFIDNKDPGFKIVKITVHQ
                NSSDMFMDREGGYTFVNDLPHGIVGRARRMCKSKKNTHLDSLLGCLSVLMKNSELQLD
                SLDIIVTHNHNTIVKWFEELVQSKNQTHCKHVQLVAPGVQRMLSFLSHFKPEFLETII
                IAQNGVENSGEICKFIDIVESEQWKRAKSIYVYSIISSNFPIENLFHLTVPCVPLGEL
                AVSDAIKIRDILLKSAHFEYGQIDVKENIPQEVVRVFVPEFNDEIPIRYRSRGLEFKVD
                QFQLTFSFNSVIIKKT"
     gene       22946. .23952
                /gene="Y102A5C.10"
     CDS        join(22946. .23142,23194. .23277,23324. .23508,23897. .23952)
                /gene="Y102A5C.10"
                /codon_start=1
                /db_xref="PID:e1323297"
                /db_xref="PID:g3646759"
                /translation="MSAALCENELSIRTCILYEALDDIPIEESYKNFCRKVGKDAMTL
                YDFDTWFYRFHNGNHDLHHDRSKDSKQLTLSDMPIGIMPEILSNVDPIDRLVLQKVSK
                GLRQVVKDMKWEFDTLSFIKDNSLLYIHPPSSGTVYYSKENGGSVTHWTRSGDARLLA
                SSWTLTGHVFFDR"
     gene       complement(27647. .28334)
                /gene="Y102A5C.12"
     CDS        complement(join(27647. .27956,28003. .28086,28138. .28334))
                /gene="Y102A5C.12"
                /codon_start=1
                /db_xref="PID:e1323298"
                /db_xref="PID:g3646760"
                /translation="MSAALCENELSIRTCILYEALDDIPIEESYKNFCRKVGKDAMTL
                YDFDTWFYRFHNGNHDLHHDRSKDSKQLTLSDMPIGIMPEILSNVDPIDRLVLQKVSK
                GLRQVVKDMKWEFDTLSFIKDNSLLYIHPPSSGTVYYSKENGGSVTHWTRSGDGKPSQ
                IRRTFVEGCNSDKHVSKLISAKAMLMLKIDVISCPT"
     gene       complement(29146. .29938)
                /gene="Y102A5C.13"
     CDS        complement(join(29146. .29340,29391. .29871,29922. .29938))
                /gene="Y102A5C.13"
                /codon_start=1
                /db_xref="PID:e1323299"
                /db_xref="PID:g3646761"
                /translation="MDPKNKCILRGVCKSKKNTHLDSLLGCLSVLMKNSELQLDSLDI
                IVTHNHNTIVKWFEELVQSKNQTHCKHVQLVAPGVQRMLSFLSHFKPEFLETIIIAQN
                GVENSGEICKFIDIVESEQWKRAKSIYVYSIISSNFPIENLFHLTVPCVPLGELAVSD
                AIKIRDILLKSAHFEYGQIDVKENIPQEVVRVFVPEFNDEIPIRYRSRGLEFKVDQFQL
                TFSFNSVIIKKT"
     gene       complement(30647. .31662)
```



```
     CDS        /gene="Y102A5C.14"
                complement(join(30647. .30814,30863. .31442,31492. .31569,
                31622. .31662))
                /gene="Y102A5C.14"
                /note="predicted using Genefinder"
                /codon_start=1
                /db_xref="PID:e1323289"
                /db_xref="PID:g3646751"
                /translation="MTENLTKQLNSCKIAHAPSLSGMPLDIINRILDKLDPIFRFSLR
                RVCRNLRNVIDNRDPEIEQVHVLFQSNQSVLTINSTFEFDYNLYAILDDLASLWKNPK
                LQLKSFYASTSGKHQETVVDSFEEIARLDYQIHCKHFGLYLFEFDDVLTFLPLFKPGI
                LEEIVITTFEQVGNDEAFDIIETEQWKQAKSITLHSMNKLSYFTFPIEDLFHLCKFSV
                SLEELTVFDALKIRNILLKSVHPEAGQINVKNKIQDDVLRVFNPEFNDNIFRYDNPEL
                TFKLNSLNVKKI"
     gene       34040. .35336
                /gene="Y102A5C.17"
     CDS        join(34040. .34184,34540. .34598,34640. .34745,34945. .35140,
                35189. .35236,35285. .35336)
...
Note: remainder of annotations omitted.

  Query Match           14.8%;  Score 19;  DB 18;  Length 91286;
  Best Local Similarity 100.0%;  Pred. No. 1.71e+00;
  Matches    19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  42101 CAAAATTCTGCTTTAAAAA 42119
          ||||||||||||||||||
Cp     27 CAAAATTCTGCTTTAAAAA 9


RESULT    8
LOCUS       AC004130   116215 bp    DNA             PRI       06-FEB-1998
DEFINITION  Homo sapiens BAC clone RG293F17 from 7p15-p21, complete sequence.
ACCESSION   AC004130
NID         g2842785
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 116215)
  AUTHORS   Tin-Wollam,A., Sutterer,C. and Fronick,B.
  TITLE     The sequence of Homo sapiens BAC clone RG293F17
  JOURNAL   Unpublished (1998)
REFERENCE   2  (bases 1 to 116215)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-FEB-1998) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     SUBMITTED BY:  WUGSC
            Genome Sequencing Center
```



```
            Department of Genetics
            Washington University
            St. Louis MO 63108, USA
            http://genome.wustl.edu/gsc
            mailto:sapiens@watson.wustl.edu

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded or sequenced with an alternate
            chemistry; an attempt was made to resolve all sequencing problems,
            such as compressions and repeats; all regions were covered by
            sequence from more than one subclone; and the assembly was
            confirmed by restriction digest.

            MAPPING INFORMATION:
            The sequence of this clone was established as part of a mapping and
            sequencing collaboration between the NHGRI Chromosome 7 Mapping
            Project (Eric D. Green, Director), John D. McPherson in the
            Department of Genetics (Washington University), and the Washington
            University Genome Sequencing Center.  For additional information
            about the map position of this sequence, see
            http://www.nhgri.nih.gov/DIR/GTB/CHR7 or send
            mailto:egreen@nhgri.nih.gov

            SOURCE INFORMATION:
            Clone RG293F17 is from a release of the human BAC library
            CITB-HS-A.  The library contains cloned DNA from human sperm.  See:
             Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J.
            Kim et al., Genomics 34:213-8 (1996).  The clone is available from
            Research Genetics, Inc. (http://www.resgen.com).
            VECTOR:  pBeloBAC11
            Selection:  chloramphenicol

            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the right is RG326G04, 200 bp overlap.
            Actual start of this clone is at base position 1 of RG293F17;
            actual end is at 116215 of RG293F17.

            This clone contains STS sWSS2601 (NID:g1222836).

            This clone RG293F17 contains a bacterial transposon from base
            position 4508-5845, which has been cut from the submitted sequence.
FEATURES              Location/Qualifiers
     source           1. .116215
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="7"
```



```
                      /clone="RG293F17"
                      /clone_lib="CITB-HS-A"
                      /map="7p15-p21"
     repeat_region    1078. .1302
                      /rpt_family="MIR"
     repeat_region    1939. .2240
                      /rpt_family="ALU"
     repeat_region    4266. .4475
                      /rpt_family="L1"
     repeat_region    6166. .6460
                      /rpt_family="ALU"
     repeat_region    8190. .9034
                      /rpt_family="L1"
     repeat_region    9046. .9128
                      /rpt_family="MIR"
     repeat_region    9629. .9780
                      /rpt_family="RETROVIRAL"
     repeat_region    10256. .10552
                      /rpt_family="ALU"
     repeat_region    10693. .10742
                      /rpt_family="MIR"
     repeat_region    11122. .11270
                      /rpt_family="MER1_TYPE"
     repeat_region    11568. .11677
                      /rpt_family="ALU"
     misc_feature     14096. .15596
                      /note="CpG_island (%GC=67.1, o/e=0.70, #CpGs=138)"
     gene             15522. .93712
                      /gene="ITB8"
     CDS              join(15522. .15648,47352. .47437,50727. .50901,62766. .63012,
                      64381. .64546,65442. .65600,75118. .75213,78611. .78700,
                      82575. .82709,85436. .85841,88343. .88568,89777. .89886,
                      93330. .93493,93590. .93712)
                      /gene="ITB8"
                      /note="match to P26012 (PID:g124975); H_RG293F17.1"
                      /codon_start=1
                      /product="INTEGRIN BETA-8 SUBUNIT PRECURSOR"
                      /db_xref="PID:g2842786"
                      /translation="MCGSALAFFTAAFVCLQNDRRGPASFLWAAWVFSLVLGLGQGED
                      NRCASSNAASCARCLALGPECGWCVQEDFISGGSRSERCDIVSNLISKGCSVDSIEYP
                      SVHVIIPTENEINTQVTPGEVSIQLRPGAEANFMLKVHPLKKYPVDLYYLVDVSASMH
                      NNIEKLNSVGNDLSRKMAFFSRDFRLGFGSYVDKTVSPYISIRPERIHNQCSDYNLDC
                      MPPHGYIHVLSLTENITEFEKAVHRQKISGNIDTPEGGFDAMLQAAVCESHIGWRKEA
                      KRLLLVMTDQTSHLALDSKLAGIVVPNDGNCHLKNNVYVKSTTMEHPSLGQLSEKLID
                      NNINVIFAVQGKQFHWYKDLLPLLPGTIAGEIESKAANLNNLVVEAYQKLISEVKVQV
                      ENQVQGIYFNITAICPDGSRKPGMEGCRNVTSNDEVLFNVTVTMKKCDVTNGKNVAII
                      KPIGFNETAKIHIHRNCSCQCEDNRGPKGKCVDETFLDSKCFQCDEHKCHFDEDQFSS
                      ESCKSHKDQPVCSGRGVCVCGKCSCHKIKLGKVYGKYCEKDDFSCPYHHGNLCAGHGE
                      CEAGRCQCFSGWEGDRCQCPSAAAQHCVNSKGQVCSGRGTCVCGRCECTDPRSIGRFC
                      EHCPTCYTACKENWNCMQCLHPHNLSQAILDQCRTSCALMEQQHYVDQTSECFSSPSY
                      LRIFFIIFIVTFLIGLLKVLIIRQVILQWNSNKIKSSSDYRVSASKKDKLILQSVCTR
```

```
                     AVTYRREKPEEIKMDISKLNAHETFRCNF"
     repeat_region   16367. .16411
                     /rpt_family="MIR"
     repeat_region   19318. .19695
                     /rpt_family="L2"
     repeat_region   19719. .20019
                     /rpt_family="ALU"
     repeat_region   20171. .20303
                     /rpt_family="MIR"
     repeat_region   21192. .21371
                     /rpt_family="MIR"
     repeat_region   22060. .22362
                     /rpt_family="ALU"
     repeat_region   22627. .22923
                     /rpt_family="ALU"
     repeat_region   23461. .23764
                     /rpt_family="ALU"
     repeat_region   25495. .25561
                     /rpt_family="MARINER"
     repeat_region   28545. .28641
                     /rpt_family="L2"
     repeat_region   28642. .28695
                     /rpt_family="RETROVIRAL"
     repeat_region   28981. .29310
                     /rpt_family="L2"
     repeat_region   29362. .29706
                     /rpt_family="L2"
     repeat_region   30602. .30960
                     /rpt_family="MALR"
     repeat_region   33045. .33124
                     /rpt_family="MARINER"
     repeat_region   33574. .33727
                     /rpt_family="L1"
     repeat_region   33861. .34162
                     /rpt_family="ALU"
     repeat_region   35013. .35240
                     /rpt_family="MIR"
     repeat_region   37367. .37666
                     /rpt_family="ALU"
     repeat_region   38876. .39009
                     /rpt_family="RETROVIRAL"
     repeat_region   41291. .41433
                     /rpt_family="MIR"
     repeat_region   41679. .41977
                     /rpt_family="ALU"
     repeat_region   42244. .42417
                     /rpt_family="ALU"
     repeat_region   42430. .42876
                     /rpt_family="L1"
     repeat_region   44091. .44913
                     /rpt_family="L1"
```



```
     repeat_region   45045. .45375
                     /rpt_family="MALR"
     repeat_region   49090. .49284
                     /rpt_family="MIR"
     repeat_region   50198. .50495
                     /rpt_family="ALU"
     repeat_region   51157. .51274
                     /rpt_family="L2"
     repeat_region   52969. .53361
                     /rpt_family="MALR"
     repeat_region   53486. .54129
                     /rpt_family="L1"
     repeat_region   54325. .54628
                     /rpt_family="ALU"
     repeat_region   54702. .54921
                     /rpt_family="L1"
     repeat_region   56264. .56393
                     /rpt_family="L2"
     misc_feature    complement(62953. .63034)
                     /gene="ITB8"
                     /note="match to EST AA644567 (NID:g2569785) ab63h07.s1"
     repeat_region   64883. .65176
                     /rpt_family="ALU"
     misc_feature    complement(65432. .65781)
                     /gene="ITB8"
                     /note="match to EST AA644567 (NID:g2569785) ab63h07.s1"
     repeat_region   68145. .68326
...
Note: remainder of annotations omitted.

  Query Match           14.8%;  Score 19;  DB 26;  Length 116215;
  Best Local Similarity 100.0%;  Pred. No. 1.71e+00;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  17794 AATTTTTAAAGCAGAATTT 17812
          |||||||||||||||||||
Qy      7 AATTTTTAAAGCAGAATTT 25


RESULT    9
LOCUS       CEY6B3     253516 bp    DNA              HTG       30-JUL-1998
DEFINITION  Caenorhabditis elegans DNA *** SEQUENCING IN PROGRESS *** from
            clone Y6B3; HTGS phase 1.
ACCESSION   Z92865
NID         g3367655
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 253516)
```



```
  AUTHORS   White,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JUL-1998) Nematode Sequencing Project, Sanger Centre,
            Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics,
            Washington University, St. Louis, MO 63110, USA. E-mail:
            jea@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     On Aug 1, 1998 this sequence version replaced gi:1888382.
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and the release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc. Order of segments is not known; 800 n's separate
            segments.
            ***                                                          ***
            *** WARNING: Phase 1 High Throughput Genome Sequence         ***
            ***                                                          ***
            * This sequence is unfinished. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
FEATURES              Location/Qualifiers
     source           1. .253516
                      /organism="Caenorhabditis elegans"
                      /db_xref="taxon:6239"
                      /chromosome="I"
                      /clone="Y6B3"
BASE COUNT    75978 a  42503 c  42374 g  76407 t  16254 others
ORIGIN

  Query Match           14.8%;  Score 19;  DB 17;  Length 253516;
  Best Local Similarity 100.0%;  Pred. No. 1.71e+00;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   7143 AATTTTTAAAGCAGAATTT 7161
          |||||||||||||||||||
Qy      7 AATTTTTAAAGCAGAATTT 25


RESULT   10
LOCUS       HSPA20F7      552 bp    DNA              PRI       23-AUG-1996
DEFINITION  H.sapiens flow-sorted chromosome 6 HindIII fragment, SC6pA20F7.
ACCESSION   Z79020
NID         g1508298
KEYWORDS    Anonymous marker; single read.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 552)
  AUTHORS   Mungall,A.J., Huckle,E., Langford,C., Ross,M.T. and Rice,C.M.
  TITLE     Direct Submission
```



```
  JOURNAL   Submitted (22-AUG-1996) The Sanger Centre, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail contact:
            humquery@sanger.ac.uk
COMMENT     Vector: pBSIISK+.
FEATURES              Location/Qualifiers
     source           1. .552
                      /organism="Homo sapiens"
                      /note="The estimated purity of the flow-sorted chromosome
                      6 library is >97%"
                      /db_xref="taxon:9606"
                      /dev_stage="adult"
                      /sex="female"
                      /tissue_type="EBV lymphoblastoid cell line"
                      /clone_lib="SC6pA"
                      /clone="SC6pA20F7"
                      /chromosome="6"
BASE COUNT      190 a     97 c    122 g    143 t
ORIGIN

  Query Match           14.1%;  Score 18;  DB 25;  Length 552;
  Best Local Similarity 100.0%;  Pred. No. 7.65e+00;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    341 TTCTTTATTCTTTTAAAA 358
          ||||||||||||||||||
Cp    112 TTCTTTATTCTTTTAAAA 95


RESULT   11
LOCUS       U92952       2286 bp    DNA              PRI       02-AUG-1997
DEFINITION  Trachypithecus francoisi NADH dehydrogenase subunit 3 (ND3) gene,
            partial cds, tRNA-Arg gene, complete sequence, NADH dehydrogenase
            subunit 4L (ND4L) and NADH dehydrogenase subunit 4 (ND4) genes,
            complete cds, and tRNA-His, tRNA-Ser and tRNA-Leu genes, complete
            sequence, mitochondrial genes encoding mitochondrial products.
ACCESSION   U92952
NID         g2290421
KEYWORDS    .
SOURCE      Trachypithecus francoisi.
  ORGANISM  Mitochondrion Trachypithecus francoisi
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini;
            Cercopithecidae; Colobinae; Trachypithecus.
REFERENCE   1  (bases 1 to 2286)
  AUTHORS   Wang,W., Forstner,M.R.J., Zhang,Y., Lui,Z., Wei,Y., Huang,H.,
            Hu,H., Xie,Y., Wu,D. and Melnick,D.J.
  TITLE     A Phylogeny of Chinese Leaf Monkeys Using Mitochondrial ND3-ND4
            Gene Sequences
  JOURNAL   Int. J. Primatol. (1997) In press
REFERENCE   2  (bases 1 to 2286)
  AUTHORS   Wang,W., Forstner,M.R.J. and Melnick,D.J.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (11-MAR-1997) CERC-mailcode5557, Columbia University,
            10th Floor Schermerhorn Ext. 119th Street & Amsterdam Avenue, New
            York, NY 10027, USA
FEATURES             Location/Qualifiers
     source          1. .2286
                     /organism="Trachypithecus francoisi"
                     /mitochondrion
                     /isolate="HY1"
                     /db_xref="taxon:54180"
     gene            <1. .346
                     /gene="ND3"
     CDS             <1. .346
                     /gene="ND3"
                     /codon_start=1
                     /exception="TAA stop codon is created by polyadenylation"
                     /transl_table=2
                     /product="NADH dehydrogenase subunit 3"
                     /db_xref="PID:g2290422"
                     /translation="MNLVLALTINTLLTLLLMIIMFWLPLLNPYAEKADPYECGFDPL
                     NSARIPFSMKFFLVAITFLLFDLEIALLLPLPWALQTTNLPVMIKSSMMLIIILTLGL
                     AYEWTQKGLDWTE"
     tRNA            347. .411
                     /product="tRNA-Arg"
     gene            412. .708
                     /gene="ND4L"
     CDS             412. .708
                     /gene="ND4L"
                     /codon_start=1
                     /transl_table=2
                     /product="NADH dehydrogenase subunit 4L"
                     /db_xref="PID:g2290423"
                     /translation="MPIIYMNIMLAFLISLLGMLIYRSHLMSSLLCLEGMMLSLFIMS
                     TLMALNMHFPLANIMPIALLVFAACEAAVGLALLVSISNTYGLDYIHNLNLLQC"
     gene            702. .2079
                     /gene="ND4"
     CDS             702. .2079
                     /gene="ND4"
                     /codon_start=1
                     /exception="TAA stop codon is created by polyadenylation"
                     /transl_table=2
                     /product="NADH dehydrogenase subunit 4"
                     /db_xref="PID:g2290424"
                     /translation="MLKMILPTIMLLPTTWFSKNNMLWINLTMHSLIISLIPLMFFNQ
                     TSNNLTPSYSTTLSLDPLSTPLLTLTAWLLPLMIMASQYHLSNEWSLSRKKLYLSMMIFL
                     QISLIMTFMSTELILFYILFETTLIPTLIIITRWGNQAERLNASTYFLFYTLAGSLPL
                     LIMLMFTYNNMGSLNIMLLTLTAQKLTTTWSHNLTWLACMMAFMVKMPLYGLHLWLPK
                     AHVEAPIAGSMVLAAVLLKLGGYGMMRLTSILNPLTETMAYPFLMLSLWGMIMTSSTC
                     LRQTDLKALIAYSSVSHMALVITASLIQTPWSFTGAIILMIAHGLTSSMLFCLANSNY
                     ERTHSRIMLFSRGLQTLLPLMASWWFVANLTNLALPPTINLIGELLVMATLPSWSHVT
                     IMLTGLNMLITALYSLYMPITTQRGTLTSHIMNMKPSFTRENMLMFMHISPIILLSLN
```



```
                     PNIIMGFTS"
     tRNA            2080. .2148
                     /product="tRNA-His"
     tRNA            2149. .2207
                     /product="tRNA-Ser"
     tRNA            2208. .2286
                     /product="tRNA-Leu"
BASE COUNT      739 a    612 c    238 g    697 t
ORIGIN

  Query Match           14.1%;  Score 18;  DB 26;  Length 2286;
  Best Local Similarity 100.0%;  Pred. No. 7.65e+00;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2202 ATCTCAACTTTTAAAAGA 2219
            ||||||||||||||||||
Qy     87 ATCTCAACTTTTAAAAGA 104


RESULT   12
LOCUS       RSU08214     2370 bp    mRNA            ROD       29-NOV-1995
DEFINITION  Rattus sp. DNA binding protein (URE-B1) mRNA, complete cds.
ACCESSION   U08214
NID         g475515
KEYWORDS    .
SOURCE      rat.
  ORGANISM  Rattus sp.
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;
            Murinae; Rattus.
REFERENCE   1  (bases 1 to 2370)
  AUTHORS   Gu,J., Ren,K., Dubner,R. and Iadarola,M.J.
  TITLE     Cloning of a DNA binding protein that is a tyrosine kinase
            substrate and recognizes an upstream initiator-like sequence in the
            promoter of the preprodynorphin gene
  JOURNAL   Brain Res. Mol. Brain Res. 24 (1-4), 77-88 (1994)
  MEDLINE   95058008
REFERENCE   2  (bases 1 to 2370)
  AUTHORS   Gu,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAR-1994) Jun Gu, Neurobiology and Anesthesiology
            Branch, The National Institutes of Health, NIDR, Building 49, Room
            1A11, 9000 Rockville Pike, Bethesda, MD 20892, USA
FEATURES             Location/Qualifiers
     source          1. .2370
                     /organism="Rattus sp."
                     /db_xref="taxon:10118"
                     /clone_lib="Stratagene cDNA rat brain expression library"
                     /tissue_type="brain"
     gene            945. .1877
                     /gene="URE-B1"
```



```
     CDS             945. .1877
                     /gene="URE-B1"
                     /note="DNA binding protein"
                     /codon_start=1
                     /db_xref="PID:g475516"
                     /translation="MISREMFNPMYALFRTSPGDRVTTRSIHLPTVTPTTSATSSLA
                     GRYVASVMTTALLECYFTRSFYKHILGKSVRTTDMESEDYHFYQGLVYLLENDVSTL
                     GYDLTFSTEVQEPGVCEVRDLKPNGANILVTEENKKETVHLVCQMRMTGAIRKQLAAF
                     LEGFYEIIPKRLISIFTEQELEELLTGLPTIDIDDLKSNTEYHKYQSNSIQIQWFWRA
                     LRSFDQADRAKFLQFVTGTSKVPLQGFAALEGMNGIQKFQIHRDDRSTDRLPSAHTCF
                     NQLDLPAYESFEKLRHMLLLAIQECSEGFGLANK"
     polyA_site      2370
                     /note="9 A nucleotides"
BASE COUNT      723 a    531 c    462 g    654 t
ORIGIN

  Query Match           14.1%;  Score 18;  DB 28;  Length 2370;
  Best Local Similarity 100.0%;  Pred. No. 7.65e+00;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    470 ACAAAATTCTGCTTTAAA 487
          ||||||||||||||||||
Cp     28 ACAAAATTCTGCTTTAAA 11


RESULT   13
LOCUS       HSTMPO1      2509 bp    DNA             PRI       02-JUL-1995
DEFINITION  Human thymopoietin (TMPO) gene, exon 1.
ACCESSION   U18266
NID         g885676
KEYWORDS    .
SEGMENT     1 of 6
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2509)
  AUTHORS   Harris,C.A., Andryuk,P.J., Cline,S.W., Siekierka,J.J. and
            Goldstein,G.
  TITLE     Structure and mapping of the human thymopoietin (TMPO) gene and
            relationship of TMPO beta to rat lamin-associated polypeptide 2
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2509)
  AUTHORS   Harris,C.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-DEC-1994) Crafford A. Harris, Immunobiology Research
            Institute, Route 22 East, Annandale, NJ 08801-0999, USA
FEATURES             Location/Qualifiers
     source          1. .2509
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
```



```
                     /clone="P1.516 (DMPC-HFFno.1B-0943F)"
                     /clone_lib="DuPont Merck Hum Fibroblast P1 Library no.1
                     Series B (compressed) (Genome Systems Inc)"
                     /chromosome="12"
                     /map="12q22; 64% (% distance from centromere to telomere)"
     5'UTR           one-of(1888,1901). .2200
                     /gene="TMPO"
     exon            one-of(1888,1901). .2479
                     /gene="TMPO"
                     /number=1
BASE COUNT      667 a    587 c    656 g    599 t
ORIGIN

  Query Match           14.1%;  Score 18;  DB 26;  Length 2509;
  Best Local Similarity 100.0%;  Pred. No. 7.65e+00;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    250 AGAATCTGAAAAGTGTAA 267
          ||||||||||||||||||
Cp     75 AGAATCTGAAAAGTGTAA 58


RESULT   14
LOCUS       ATU87834     4377 bp    DNA             PLN       18-MAR-1997
DEFINITION  Arabidopsis thaliana zinc-finger protein Lsd1 (LSD1) gene, complete
            cds.
ACCESSION   U87834
NID         g1872522
KEYWORDS    .
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryotae; mitochondrial eukaryotes; Viridiplantae;
            Charophyta/Embryophyta group; Embryophyta; Magnoliophyta;
            Magnoliopsida; Capparales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 4377)
  AUTHORS   Jaba,T., Dietrich,R.A. and Dangl,J.L.
  TITLE     Initiation of runaway cell death in an Arabidopsis mutant by
            extracellular superoxide
  JOURNAL   Science 273 (5283), 1853-1856 (1996)
  MEDLINE   96386393
REFERENCE   2  (bases 1 to 4377)
  AUTHORS   Dietrich,R.A., Richberg,M.H., Schmidt,R., Dean,C. and Dangl,J.L.
  TITLE     A novel zinc finger protein is encoded by the Arabidopsis LSD1 gene
            and functions as a negative regulator of plant cell death
  JOURNAL   Cell 88 (5), 685-694 (1997)
  MEDLINE   97207015
REFERENCE   3  (bases 1 to 4377)
  AUTHORS   Dietrich,R.A. and Dangl,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JAN-1997) Biology Department, University of North
            Carolina, 108 Coker Hall, Chapel Hill, NC 27599, USA
```

```
FEATURES             Location/Qualifiers
     source          1. .4377
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /chromosome="4"
                     /map="between g3883 and g13838"
     mRNA            join(1989. .2494,2584. .2688,3178. .3317,3418. .3489,
                     3582. .3653,3741. .3985)
                     /gene="LSD1"
                     /product="Lsd1"
     gene            1989. .3985
                     /gene="LSD1"
     CDS             join(2356. .2494,2584. .2688,3178. .3317,3418. .3489,
                     3582. .3653,3741. .3782)
                     /gene="LSD1"
                     /function="negative regulator of cell death"
                     /codon_start=1
                     /product="zinc-finger protein Lsd1"
                     /db_xref="PID:g1872523"
                     /translation="MKVADMQDQLVCHGCRNLLMYPRGASNVRCALCNTINMVPPPPP
                     PHDMAHIICGGCRTMLMYTRGASSVRCSCCQTTNLVPAHSNQVAHAPSSQVAQINCGH
                     CRTTLMYPYGASSVKCAVCQPVTNVNMSNGRVPLPTNRPMGTACPPSTSTSTPPSQTQ
                     TVVVENPMSVDESGKLVSNVVVGVTTDKK"
BASE COUNT     1351 a    787 c    736 g   1503 t
ORIGIN

  Query Match           14.1%;  Score 18;  DB 24;  Length 4377;
  Best Local Similarity 100.0%;  Pred. No. 7.65e+00;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1714 AAAGGGGTAGACAAAAAA 1731
          ||||||||||||||||||
Cp     58 AAAGGGGTAGACAAAAAA 41


RESULT   15
LOCUS       CEY17D7A    21494 bp    DNA             INV       22-SEP-1998
DEFINITION  Caenorhabditis elegans cosmid Y17D7A, complete sequence.
ACCESSION   AL021470
NID         g2815047
KEYWORDS    HTG.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 21494)
  AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
```



```
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
  TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
  JOURNAL   Nature 368 (6466), 32-38 (1994)
  MEDLINE   94150718
REFERENCE   2  (bases 1 to 21494)
  AUTHORS   Kershaw,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-NOV-1996) Nematode Sequencing Project, Sanger Centre,
            Hinxton, Cambridge CB10 1RQ, England and Department of Genetics,
            Washington University, St. Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     Coding sequences below are predicted from computer analysis, using
            predictions from Genefinder (P. Green, U. Washington), and other
            available information.
            For a graphical representation of this sequence and its analysis
            see:-
            http://webace.sanger.ac.uk/cgi-
            bin/display?db=wormace&class=Sequence &object=Y17D7A
            Current sequence finishing criteria for the C. elegans genome
            sequencing consortium are that all bases are either sequenced
            unambiguously on both strands, or on a single strand with both a
            dye primer and dye terminator reaction, from distinct subclones.
            Exceptions are indicated by an explicit note.
            IMPORTANT:  This sequence is NOT necessarily the entire insert of
            the specified clone.  It may be shorter because we only sequence
            overlapping sections once, or longer because we arrange for a small
            overlap between neighbouring submissions.
            IMPORTANT: This sequence is not the entire insert of clone Y17D7A.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we arrange for a small overlap between
            neighbouring submissions.
            The true left end of clone Y17D7 is at 1 in this sequence. The true
            left end of clone C54E10 is at 21382 in this sequence. The end of
            this sequence (21382. .21494) overlaps with the start of sequence
            Z81487.
FEATURES             Location/Qualifiers
     source          1. .21494
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /chromosome="V"
                     /clone="Y17D7A"
     gene            complement(4418. .6309)
                     /gene="Y17D7A.2"
     CDS             complement(join(4418. .4706,5258. .5457,5980. .6201,
                     6289. .6309))
```



```
                     /gene="Y17D7A.2"
                     /codon_start=1
                     /db_xref="PID:e1247373"
                     /db_xref="PID:g2815049"
                     /db_xref="SPTREMBL:O45906"
                     /translation="MRRRFQKGSESPLYTKSNLQKLAEALQKIDICHDSSPSQPVKLV
                     TKYGKEEVPSFFEQDFLKATRWPTYFDEFQDLDQEQKLELIQAIWHVWSRLYKLSTSA
                     IGKRRQICEDKMLMISHGSEYTVMDLTKIEFDPSWCTKYTNEQMQHFIDNTQNSFLYK
                     LVDYMMDLKPSDVELSFMICQACFHYAGQRFHGKMLEICENFQQKLADELHEFYVDEW
                     RMPNYSGRLSQLLRINNRIRVEDFL"
     gene            complement(7086. .8971)
                     /gene="Y17D7A.1"
     CDS             complement(join(7086. .7244,7288. .7340,7402. .7541,
                     8838. .8971))
                     /gene="Y17D7A.1"
                     /note="similar to Zinc finger, C4 type (two domains)"
                     /codon_start=1
                     /db_xref="PID:e1247372"
                     /db_xref="PID:g2815048"
                     /db_xref="SPTREMBL:O45905"
                     /translation="MDLIASTSILSQEESSELICAVCSQPARGRHPGAVACRACAAFF
                     RRADAAKTTVKPCKKGGNCQNLLNNNGWFDCKPCRLQKCYEIGMNSANFQFDRDPIST
                     KLTNVPNSMEGFLGRPHCVIFVDPEVVRCEKDLIDCTDLLRKAGRILKEVSVRSTKKV
                     F"
     gene            complement(10500. .16470)
                     /gene="nhr-65"
     CDS             complement(join(10500. .10607,11714. .11984,13048. .13247,
                     13975. .14184,14464. .14589,15142. .15206,15251. .15363,
                     16358. .16470))
                     /gene="nhr-65"
                     /note="similar to Zinc finger, C4 type (two domains); cDNA
                     EST yk480h4.5 comes from this gene; cDNA EST yk480h4.3
                     comes from this gene"
                     /codon_start=1
                     /db_xref="PID:e1247374"
                     /db_xref="PID:g2815050"
                     /db_xref="SPTREMBL:O45907"
                     /translation="MELEPCSSSPERCKVCGDTGNGMHPGAFTCRACAAFFRRAASRK
                     FLRKCENHLIFALKCKNCRLQRCYEAGMSSENFQFCRDLIGAKGAIPKLKVPESPEQT
                     VGRPVFVLQCDPEVLFLRRNIIDCVPLLEKAEKLIEPGSESPVFSKNRLLKLAQGLQQ
                     FQDAKSNQVKFVQKMGQKEIMSFFETDFLCATKWFTYLDEFQFLDKNQQLTLMQGIWH
                     VWSRLHKLAVSAMGRRRGICDKNTVMVSHQNEFAVCDLNEIEVDMSWCTNFSNEQMRY
                     FLDTSHDSYVYQVMDEMLGLKPNDVELSYMMCQLCLQYAGQRFQGEILEPCEKMLGPL
                     ADDLHSYYVKQLRMPNYAARLAKLMKINNRIKVDMLKMRQRQEISRVFDIWTIDPSHP
                     EFIQDANAC"
     CDS             join(18591. .18789,18908. .19134,20092. .20226,20855. .21153,
                     Z81487:673. .1021,Z81487:1802. .2080)
                     /gene="Y17D7A.4"
                     /note="similar to Cytochrome P450"
                     /codon_start=1
                     /db_xref="PID:e1323332"
```



```
                     /db_xref="PID:g3646788"
                     /translation="MLLLLIFSSIFLYFFYHFHWKRRNLPPGPRPLPFLGNLLSLKTL
                     KPGYEAPSNWKKEYGPIFTFWMANKPFVIIASYEKMKETPVKDGDTYVDKQLTHTEKE
                     RLGENYGVLDTNGHMWKEHRRFTLTQLRDLGLGKDLMQEKILMEVEELFKELDAHGGE
                     EIDLPKLIDRSVGNVINLTLFNKRFDMDKRDEFAHLKSLIDGMRNVTSQFRYLIQYLV
                     PWTSTVLPGPTLSEKVRAKREELDDFFYSQIDEHRNEIDFDNTENLDFVEAYLKEQKK
                     REEDGDFKTFCNKQLCAMLFDLWIAGLMTTTMTMTWGLSYYLYNPEVQRKIREELDKV
                     IGNDRLISTADKNDLPYLQAFVTETQRTANIIPLNLIHMTTRDTVIDGFPIQKGTGVI
                     AQISTVMYDEKVFPEPYKFKPERFIENGKFKKVDEVIPFSIGKRQCLGEGLARIELFL
                     FFANIFNRYDVMPDFSGSLPDLDKSKDNFVIPRKFKAVLKRRYMD"
BASE COUNT     7185 a   3625 c   3691 g   6993 t
ORIGIN

  Query Match           14.1%;  Score 18;  DB 18;  Length 21494;
  Best Local Similarity 100.0%;  Pred. No. 7.65e+00;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  18061 TTTTTAAAGCAGAATTTT 18078
          ||||||||||||||||||
Qy      9 TTTTTAAAGCAGAATTTT 26


RESULT   16
LOCUS       CEF08G5     32784 bp    DNA             INV       21-SEP-1998
DEFINITION  Caenorhabditis elegans cosmid F08G5, complete sequence.
ACCESSION   Z70682
NID         g1256464
KEYWORDS    HTG; Collagen; Transfer RNA; tRNA-Arg; tRNA-Gln;
            UDP-glucuronosyltransferase like.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 32784)
  AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
  TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
  JOURNAL   Nature 368 (6466), 32-38 (1994)
  MEDLINE   94150718
REFERENCE   2  (bases 1 to 32784)
  AUTHORS   Steward,C.
```

```
TITLE     Direct Submission
JOURNAL   Submitted (06-NOV-1996) Louis, MO 63110, USA. E-mail:
          jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT   Coding sequences below are predicted from computer analysis, using
          predictions from Genefinder (P. Green, U. Washington), and other
          available information.
          For a graphical representation of this sequence and its analysis
          see:-
          http://webace.sanger.ac.uk/cgi-
          bin/display?db=wormace&class=Sequence &object=F08G5
          Current sequence finishing criteria for the C. elegans genome
          sequencing consortium are that all bases are either sequenced
          unambiguously on both strands, or on a single strand with both a
          dye primer and dye terminator reaction, from distinct subclones.
          Exceptions are indicated by an explicit note.
          IMPORTANT:  This sequence is NOT necessarily the entire insert of
          the specified clone.  It may be shorter because we only sequence
          overlapping sections once, or longer because we arrange for a small
          overlap between neighbouring submissions.
          IMPORTANT: This sequence is not the entire insert of clone F08G5.
          It may be shorter because we only sequence overlapping sections
          once, or longer because we arrange for a small overlap between
          neighbouring submissions.
          The true left end of clone F08G5 is at 1 in this sequence. The true
          right end of clone F08G5 is at 7448 in
          sequence Z70680.
          The true left end of clone C25G4 is at 32680 in this sequence. The
          true right end of clone F28D1 is at 725 in this sequence. The start
          of this sequence (1. .103) overlaps with the end of sequence Z70684.
          The end of this sequence (32680. .32784) overlaps with the start of
          sequence Z70680.
FEATURES             Location/Qualifiers
     source          1. .32784
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /chromosome="IV"
                     /clone="F08G5"
     gene            2838. .4101
                     /gene="F08G5.1"
     CDS             join(2838. .2935,2985. .3151,3200. .3582,3633. .3950,
                     4009. .4101)
                     /gene="F08G5.1"
                     /note="predicted using Genefinder"
                     /codon_start=1
                     /db_xref="PID:e348711"
                     /db_xref="PID:g1256467"
                     /db_xref="SPTREMBL:Q19220"
                     /translation="MEVNSRRLTVIHGVEPIETVYCNMEVSRYPSLKMKNTNLFVIVN
                     NQAQGFRLTLRGEDRENFLSTVRKFAYISETPVKDHLNRSSTNTDVWYNSVEKRRGGS
                     SSQPTSFSQPCDIDIGGSVKKSAARNLSHTTSTNIGESSRMPALTANEPFSQPVYNPY
                     NRPASSASTVSSSIGSAYSVLHDDSPFSGPSPNSNHSLQFPCPSPSHSSSPSGFSQSS
```



```
                     SHSSQLSSSPANAPSFPDFHSPPSANDIQIDQLNPVTVAPQIKQTADKCVQTDKQYVD
                     ERFDDPAFLRRYIKRVMSDSKMARLVGMMRSEIRKLPNDEFDAFYKKVKDHQRTAPSS
                     SSKIGPPAIVQQTSDMEP"
     gene            4263. .5943
                     /gene="F08G5.2"
     CDS             join(4263. .4329,4376. .4555,4855. .5113,5309. .5738,
                     5788. .5943)
                     /gene="F08G5.2"
                     /note="predicted using Genefinder"
                     /codon_start=1
                     /db_xref="PID:e349134"
                     /db_xref="PID:g1256468"
                     /db_xref="SPTREMBL:Q19221"
                     /translation="MEAENDNKKQGWTLIDNVISCILLVYAGIVTAFICPIVFLCFFY
                     TAAPIMLFSTEFRKFVMKIMARAGNLQFITCVRITGIKVMMSGDGLERLTDCRALLLP
                     NHLGLFDHPIFMTAADSPCVNAVGRWIFVIYNMWIYSPLGWLNSSYGNTFIDSVPPHK
                     RAQTLHSLRRHFDRIYHEVDLRWLCLYPEGSRLFLIQKRNSEFEKKRGLPPLTNCAYP
                     RLGAALSAIKVLGPDDPNNPLKSNNGKGPPLEYLVDVTLGYPDGNILPLNEIFMSAWRN
                     GGGKPFAIHYEIFKMDPKWKDEEELKNFLFDRYQKKDKLLESYYKTGSPSSESKEAQF
                     PSDMILIGFHILIMGLFALLVIFLLSPVY"
     gene            complement(6848. .8473)
                     /gene="F08G5.3"
     CDS             complement(join(6848. .6998,7613. .7761,7964. .8053,
                     8357. .8473))
                     /gene="F08G5.3"
                     /note="predicted using Genefinder"
                     /codon_start=1
                     /db_xref="PID:e348710"
                     /db_xref="PID:g1256466"
                     /db_xref="SPTREMBL:Q19219"
                     /translation="MSYQHEVVKYAPTEDNEIHALIRGKPSFCKANIRVNTQQPQERL
                     TYLDKRYDHLRKMTHSLRKKVNELEDIMRLDNDEENMDTIQKLLDEIKREKQLMRDEA
                     HIIKGELSKTMYNEDLRRRLAGEWKKIDADRKSSAAESEASSSMIGDDDTPSTSSDGH
                     KPYRETSM"
     tRNA            complement(10702. .10773)
                     /gene="F08G5.t3"
                     /note="CTG Gln Q-tRNA; predicted using tRNAscan-SE-1.11;
                     preliminary prediction; similar to tRNA-Gln"
     gene            complement(10702. .10773)
                     /gene="F08G5.t3"
     tRNA            complement(11009. .11080)
                     /gene="F08G5.t2"
                     /note="TTG Gln Q-tRNA; predicted using tRNAscan-SE-1.11;
                     preliminary prediction; similar to tRNA-Gln"
     gene            complement(11009. .11080)
                     /gene="F08G5.t2"
     tRNA            complement(11686. .11758)
                     /gene="F08G5.t1"
                     /note="ACG Arg R-tRNA; predicted using tRNAscan-SE-1.11;
                     preliminary prediction; similar to tRNA-Arg"
     gene            complement(11686. .11758)
```



```
                     /gene="F08G5.t1"
     gene            complement(21306. .22257)
                     /gene="F08G5.4"
     CDS             complement(join(21306. .22064,22117. .22257))
                     /gene="F08G5.4"
                     /note="predicted using Genefinder; similar to collagen;
                     cDNA EST yk204e8.3 comes from this gene; cDNA EST
                     yk204e8.5 comes from this gene"
                     /codon_start=1
                     /db_xref="PID:e349133"
                     /db_xref="PID:g1256465"
                     /db_xref="SPTREMBL:Q19218"
                     /translation="MSAKYVVTIASCPSGLAIVACLFTVGAIFKDINDLYDNVMDDMD
                     EFRHFANNAWKDMIPVTRPSLDNQSNLRAIPGREKRQAGQCNCGAQPNNCPPGPPGPP
                     GAPGAPGDDGHAGEAGKTGINGISLISHEGESGCIKCPAGEAGPAGPDGAPGAPGPDG
                     QPGQDGPAGQPGAPGPAGPEGDAGAPGDAGAPGAPGAPGQDGQRGTGLPGAPGAPGPQ
                     GPSGNPGQDGAAGAPGQAGPAGPAGPDGQPGQAGQDGEAGPEGNAGQPGADAAYCPCP
                     ARTGAVENKPETSGYRRRVSKVV"
     gene            23649. .28762
                     /gene="F08G5.5"
     CDS             join(23649. .23704,23750. .23840,24635. .25069,25364. .25666,
                     25854. .26015,26074. .26241,26402. .26465,26511. .26564,
                     27016. .27170,27239. .27395,27512. .27936,27983. .28079,
                     28542. .28762)
                     /gene="F08G5.5"
                     /note="Similarity to Rabbit UDP-glucuronosyltransferase
                     (SW:UDBE_RABIT)"
                     /codon_start=1
                     /db_xref="PID:e348712"
                     /db_xref="PID:g1256469"
                     /db_xref="SPTREMBL:Q19222"
                     /translation="MPVASKMRGYARNPVTVPNTFLLSSEVYEGNRFAFLTRTRRIFE
                     YFIDQDSYQCLGFLCFYFYKLSQNKFQITSGPVASLYSEASGKLLGLTSATKSKTFQN
                     SILTVNDFPDTFSFVQPRGNDLIRVGEHCYSSANLPSEFRDFVEDSMSKGTIYVAMGS
                     YLNLEDGPKGTVEAFIEALNVFEDYRVIWSHKGNVTGAKCHVKSVNWAPQKELLAHEK
                     TVAFITHGGLKSAEEGVCSGVPMLFLPFYGDQPRNAHRFVTNGIAEALYKKAITSLDI
                     QQKLEKLLVDPSYKNNVMKQAIAEIARIADSIAERYVPRTPVPKSPIKISFRNVVSLY
                     KNIYSLRYGPCGNMYVLSYYLDAPISSLDLGAFHISQVLRRTESQYSKFKRRSIPMNH
                     LQYLWLDIFLVILTRLMFVWLIPSHFVIHTPCNTNDHIEDFLRNKTAEDIIGLRQQYH
                     RAIYKKKKEESALADASQPNLVLADGAVLPCEAGWHQFPDTGCCYRVSDAESEWHGGT
                     AICKALNPDAEMASFHSQAESIFVANKYSTIHAWYWIISNSAPSDADSSCVEMMDGLL
                     GLIFALSLQKGQTNAYSCDESIQIICKYCPRETTSTTTVTPTTTKAATIKTSTTTVTP
                     TTVTTTKAPTTIKTSTSTTTPKFTTTTIQTTKKITTSPTSTTASIPITCTSNCPVPAV
                     NFKGLCYKMCRGLVKFEDSCTWCNGTMATISSGEENDFVSRVPGSNDETTRQIWIGNT
                     ESSGYLNWSQGQPTKPNDGLDYCISMDLSAGSTRGKYKYLPCQSTVINSLCVMNP"
     gene            29612. .31951
                     /gene="F08G5.6"
     CDS             join(29612. .29663,29708. .29817,29871. .29983,30029. .30147,
                     30196. .30308,30695. .31054,31781. .31951)
                     /gene="F08G5.6"
                     /codon_start=1
```



```
                     /db_xref="PID:e349135"
                     /db_xref="PID:g1256470"
                     /db_xref="SPTREMBL:Q19223"
                     /translation="MKTLLWFLVFVGVCVSNALDCTQIPAGGIVPRHNATIPAGAKEPVL
                     IPPNFNCLYNIKVPAFVYARVVLENGLNGNNDQITVTNEQNAKTFINSRSGKYATFYV
                     FPNTLTSPLVSTKSVNMHSSFRMVVDYYSLNNATVTYLDNPDMQYSRLYTLQAEIGWS
                     PRTIVAKERMSISIAYSGWTPYDNLYNFFLIEGDLENPLAIYRMTQFLYQNYISKGNK
                     VTVLGLDNTFSRSSIVFVPLSQAQQYDSYTAASGYMNLNQVDIDATVGEKKKKIVTMI
                     SMRDYVMIWKMEKTSNPDCVLKAITSPPTSESKVLLDFSVETQFPTNITRPSFCIIAE
                     NCAASIQFIVL"
     CDS             join(32440. .32521,32678. .32731,32781. .32784,
                     Z70680:106. .262,Z70680:531. .686,Z70680:930. .1673,
                     Z70680:1717. .1813,Z70680:2052. .2234,Z70680:3086. .3291,
                     Z70680:3339. .3506,Z70680:3842. .4117)
                     /gene="C25G4.1"
                     /note="similar to Lectin C-type domain short and long
                     forms (2 domains)"
                     /codon_start=1
                     /db_xref="PID:e1322063"
...
Note: remainder of annotations omitted.

  Query Match          14.1%;  Score 18;  DB 18;  Length 32784;
  Best Local Similarity 100.0%;  Pred. No. 7.65e+00;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   3153 TATTCTTTATTCTTTTAA 3170
         ||||||||||||||||||
Cp    114 TATTCTTTATTCTTTTAA 97


RESULT  17
LOCUS       AC005715    40753 bp    DNA             HTG       26-SEP-1998
DEFINITION  *** SEQUENCING IN PROGRESS *** DS03296; D437; Drosophila
            melanogaster; Chromosome 2R; Region 50E1-50E2; HTGS phase 1, 27
            unordered pieces.
ACCESSION   AC005715
NID         g3659415
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea;
            Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 40753)
  AUTHORS   Celniker,S.E., George,R.A., Galle,R., Svirskas,R.R., Hoskins,R.A.,
            Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Chavez,C.,
            Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A.,
            Hummasti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B.,
            Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S., Moshrefi,A.R.,
            Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,E.,
```

```
            Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R.,
            Zieran,L.L. and Kimmel,B.E.
  TITLE     Sequencing of Drosophila melanogaster
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 40753)
  AUTHORS   Celniker,S.E., George,R.A., Galle,R., Svirskas,R.R., Hoskins,R.A.,
            Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Chavez,C.,
            Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A.,
            Hummati,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B.,
            Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S., Moshrefi,A.R.,
            Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,E.,
            Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R.,
            Zieran,L.L. and Kimmel,B.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-SEP-1998) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://fruitfly.berkeley.edu/sequence/) or send
            email to drosophila@mhgc.lbl.gov. All contigs in this submission
            meet the following cutoffs: length >= 400 bases, phrap computed
            error rate <= 1/10.                                          ***
            ***                                                          ***
            *** WARNING: Phase 1 High Throughput Genome Sequence         ***
            ***                                                          ***
            * This sequence is unfinished. It consists of 27 contigs for
            * which the order is not known; their order in this record is
            * arbitrary. In some cases, the exact lengths of the gaps
            * between the contigs are also unknown; these gaps are  presented
            * as runs of N as a convenience only. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
            *       1     1133: contig of 1133 bp in length
            *    1134     1213: gap of unknown length
            *    1214     1937: contig of 724 bp in length
            *    1938     2017: gap of unknown length
            *    2018     2887: contig of 870 bp in length
            *    2888     2967: gap of unknown length
            *    2968     4005: contig of 1038 bp in length
            *    4006     4085: gap of unknown length
            *    4086     4928: contig of 843 bp in length
            *    4929     5008: gap of unknown length
            *    5009     5713: contig of 705 bp in length
            *    5714     5793: gap of unknown length
            *    5794     6841: contig of 1048 bp in length
            *    6842     6921: gap of unknown length
            *    6922     7658: contig of 737 bp in length
            *    7659     7738: gap of unknown length
            *    7739     9041: contig of 1303 bp in length
            *    9042     9121: gap of unknown length
            *    9122    10249: contig of 1128 bp in length
```



```
            *    10250    10329: gap of unknown length
            *    10330    11589: contig of 1260 bp in length
            *    11590    11669: gap of unknown length
            *    11670    12469: contig of 800 bp in length
            *    12470    12549: gap of unknown length
            *    12550    14147: contig of 1598 bp in length
            *    14148    14227: gap of unknown length
            *    14228    15639: contig of 1412 bp in length
            *    15640    15719: gap of unknown length
            *    15720    16961: contig of 1242 bp in length
            *    16962    17041: gap of unknown length
            *    17042    18408: contig of 1367 bp in length
            *    18409    18488: gap of unknown length
            *    18489    19581: contig of 1093 bp in length
            *    19582    19661: gap of unknown length
            *    19662    21011: contig of 1350 bp in length
            *    21012    21091: gap of unknown length
            *    21092    22382: contig of 1291 bp in length
            *    22383    22462: gap of unknown length
            *    22463    24504: contig of 2042 bp in length
            *    24505    24584: gap of unknown length
            *    24585    27028: contig of 2444 bp in length
            *    27029    27108: gap of unknown length
            *    27109    28862: contig of 1754 bp in length
            *    28863    28942: gap of unknown length
            *    28943    30953: contig of 2011 bp in length
            *    30954    31033: gap of unknown length
            *    31034    32926: contig of 1893 bp in length
            *    32927    33006: gap of unknown length
            *    33007    34436: contig of 1430 bp in length
            *    34437    34516: gap of unknown length
            *    34517    37632: contig of 3116 bp in length
            *    37633    37712: gap of unknown length
            *    37713    40753: contig of 3041 bp in length.
FEATURES             Location/Qualifiers
     source          1. .40753
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="DS03296; D437"
                     /chromosome="2"
                     /map="50E1-50E2"
BASE COUNT    10850 a   8355 c   8342 g  11124 t   2082 others
ORIGIN

  Query Match          14.1%;  Score 18;  DB 17;  Length 40753;
  Best Local Similarity 100.0%;  Pred. No. 7.65e+00;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  20502 TTTTAAAAGAATAAAGAA 20519
          ||||||||||||||||||
Qy     95 TTTTAAAAGAATAAAGAA 112
```



```
RESULT   18
LOCUS       CEC34D1     45022 bp    DNA             INV       21-SEP-1998
DEFINITION  Caenorhabditis elegans cosmid C34D1, complete sequence.
ACCESSION   Z78060
NID         g1487889
KEYWORDS    HTG; Guanine nucleotide-binding protein.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 45022)
  AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
  TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
  JOURNAL   Nature 368 (6466), 32-38 (1994)
  MEDLINE   94150718
REFERENCE   2  (bases 1 to 45022)
  AUTHORS   Baynes,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-NOV-1996) Nematode Sequencing Project, Sanger Centre,
            Hinxton, Cambridge CB10 1RQ, England and Department of Genetics,
            Washington University, St. Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     Coding sequences below are predicted from computer analysis, using
            predictions from Genefinder (P. Green, U. Washington), and other
            available information.
            For a graphical representation of this sequence and its analysis
            see:-
            http://webace.sanger.ac.uk/cgi-
            bin/display?db=wormace&class=Sequence &object=C34D1
            Current sequence finishing criteria for the C. elegans genome
            sequencing consortium are that all bases are either sequenced
            unambiguously on both strands, or on a single strand with both a
            dye primer and dye terminator reaction, from distinct subclones.
            Exceptions are indicated by an explicit note.
            IMPORTANT:  This sequence is NOT necessarily the entire insert of
            the specified clone.  It may be shorter because we only sequence
            overlapping sections once, or longer because we arrange for a small
            overlap between neighbouring submissions.
```



```
            This sequence is the entire insert of clone C34D1. The true right
            end of clone F57B1 is at 5273 in this sequence. The start of this
            sequence (1. .744) overlaps with the end of sequence Z78064.
            The end of this sequence (44923. .45022) overlaps with the start of
            sequence AL022271.
FEATURES             Location/Qualifiers
     source          1. .45022
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /chromosome="V"
                     /clone="C34D1"
     gene            complement(1721. .2078)
                     /gene="C34D1.1"
     CDS             complement(join(1721. .1802,1853. .1919,1976. .2078))
                     /gene="C34D1.1"
                     /note="predicted using Genefinder; Weak similarity to
                     Drosophila doublesex protein (SW:DSXF_DROME)"
                     /codon_start=1
                     /db_xref="PID:e348933"
                     /db_xref="PID:g1487893"
                     /db_xref="SPTREMBL:Q18436"
                     /translation="MTQIGLYPSALTKPFELSLSAIKSVLEPENPSYSKRVPNCQKCG
                     QHGRKSRLKGHKRSCPFRECPCAKVCVFSNKRSSRVSLE"
     gene            complement(12308. .12741)
                     /gene="C34D1.2"
     CDS             complement(join(12308. .12475,12522. .12598,12642. .12741))
                     /gene="C34D1.2"
                     /note="predicted using Genefinder; Weak similarity to
                     Drosophila doublesex protein (SW:DSXF_DROME)"
                     /codon_start=1
                     /db_xref="PID:e348105"
                     /db_xref="PID:g1487892"
                     /db_xref="SPTREMBL:Q18435"
                     /translation="MVADVLRQSVLVPISMYPQYPVPRRTTTSVFKEYQKKKKIYYCQ
                     RCLNHDVPRPRKNHKCECPYADCTCEKCGLVEKRRILNIRLQNYNQFNIENENDSKLI
                     IVEDDDKKKQGG"
     gene            complement(23020. .25671)
                     /gene="odr-3"
     CDS             complement(join(23020. .23091,23166. .23438,24223. .24352,
                     24569. .24697,25153. .25458,25511. .25671))
                     /gene="odr-3"
                     /note="predicted using Genefinder; ODR-3 Guanine
                     nucleotide-binding protein"
                     /codon_start=1
                     /db_xref="PID:e348104"
                     /db_xref="PID:g1487891"
                     /db_xref="SPTREMBL:Q18434"
                     /translation="MGSCQSNENSEGNARNKEIEKQLNADKRAGSSIVKLLLLGAGEC
                     GKSTVLKQMQILHSNGFTEEEVNEKRAIVYNNTVSAMCTILRAMDGVLHLPLENGQKE
                     AEKAIVMKVQENGEGEALTEEVSKAIQSLWADPGVKKAFEMRSEYQLPDSAKYFLDN
                     CQRISEPGYRPNDQDILYSRVATTGVVEVKFKIKELDFRVFDVGGQRSERRKWIHCFD
```

```
                     NVESIIFITAISEYDQVLFEDETTNRMIESMQLFNSICNSTWFLSTAMILFMNKKDLF
                     MEKIQRVNITTAFPDYEGGQNYEEAVSFIKQKFAELNLNPDKKTIYMHETCATDTNQV
                     QLVISSVIDTIIQKNLQKAGMM"
     gene            complement(33360. .35429)
                     /gene="C34D1.4"
     CDS             complement(join(33360. .33512,34036. .34299,34354. .34536,
                     35289. .35429))
                     /gene="C34D1.4"
                     /note="predicted using Genefinder"
                     /codon_start=1
                     /db_xref="PID:e348932"
                     /db_xref="PID:g1487890"
                     /db_xref="SPTREMBL:Q18433"
                     /translation="MSLASSSSTNTTNGTSSNRDALSSSPSSSSSPSPSSSCLIKCFL
                     MERHYRHLIFSLSLHTLIYLLLLRMCFSPRDLAQAYLADVIVLLLAAPGTARNLPILIF
                     PSVIVKFICFSLCAGVAMLSLSETSINMKTKEARMEFRRQRNYPEVVNIAFEDHPTVC
                     IWIHAMAVIVALDLNIGLKAFYSSFDCMKRPVVEEEKPPSYTVCMTNSSQTTLPPTYS
                     EALEARKKMQAQAFPTRKNQRNSGIPVV"
     CDS             join(44587. .44907,44957. .45022,AL022271:101. .127,
                     AL022271:378. .443,AL022271:865. .1035)
                     /gene="C34D1.5"
                     /note="similar to Basic region plus leucine zipper
                     transcription factors"
                     /codon_start=1
                     /db_xref="PID:e1322409"
                     /db_xref="PID:g3642081"
                     /translation="MSDYCSNDASPFGDYDLPSDEPSPKYQKLNKNIFSPLSPTFELK
                     FQSKLRVSGNSTSSTTSDDSDDESLHRSNQLKHLETLFKESILRTSFSRPGVSNTSCD
                     EKLDLVSEDEKKRLRNTEAARRCREKIKRKTDDLETELTRLTARNEVMNQHRIRLLSQ
                     VEEQMRMLENIKSRNPELEAEIRYEAKKIVDTYTHQVETKRKAIQDYFDMTCNKLY"
BASE COUNT    15786 a   7347 c   7838 g  14051 t
ORIGIN

  Query Match          14.1%;  Score 18;  DB 18;  Length 45022;
  Best Local Similarity 100.0%;  Pred. No. 7.65e+00;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  24628 AGAATAAAGAATATCTTG 24645
          ||||||||||||||||||
Qy    102 AGAATAAAGAATATCTTG 119


RESULT   19
LOCUS       AC005196    69060 bp    DNA             PRI       29-JUN-1998
DEFINITION  Homo sapiens chromosome 5, BAC clone 44J17 (LBNL H96), complete
            sequence.
ACCESSION   AC005196 AC002282 AC002283 AC001186 AC001187 AC001188 AC001189
            AC001190 AC001191 AC001192 AC001193 AC001194 AC001195 AC001196
            AC001197 AC001198 AC001199 AC001200 AC001617 AC001618 AC001619
            AC004635
NID         g3264842
```



```
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 69060)
  AUTHORS   Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
            Kadner,K., Miguel,T., Miller,C., Pitluck,S., Pollard,M.,
            Rojeski,H., Subramanian,S. and Martin,C.H.
  TITLE     Sequencing of human chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 69060)
  AUTHORS   Ricke,D.O.
  TITLE     Large Scale Sequence Analysis and Annotation with the Sequence
            Comparison Analysis (SCAN) System
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 69060)
  AUTHORS   Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
            Davis,C.A., Kadner,K., Miguel,T., Pitluck,S., Pollard,M.,
            Rojeski,H., Subramanian,S. and Martin,C.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUN-1998) Human Genome Center, DOE Joint Genome
            Institute, Lawrence Berkeley National Laboratory, MS 74-157,
            Berkeley, CA 94720, U.S.A.
COMMENT     Sequence submitted by:
            DOE Joint Genome Institute.
FEATURES             Location/Qualifiers
     source          1. .69060
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="5q"
                     /clone="BAC 44J17"
                     /chromosome="5"
                     /note="LBNL H96"
     repeat_region   complement(174. .476)
                     /rpt_family="L1"
     repeat_region   1136. .1992
                     /rpt_family="L1"
     repeat_region   2409. .2820
                     /rpt_family="L1"
     repeat_region   complement(3194. .4086)
                     /rpt_family="L1"
     repeat_region   4075. .5469
                     /rpt_family="L1"
     repeat_region   complement(7898. .8200)
                     /rpt_family="Alu"
     misc_feature    8842. .8955
                     /note="GRAIL 2 excellent exon, frame 1"
     prim_transcript complement(join(9002. .9363,9463. .9536))
                     /note="82% identity AL021307"
     repeat_region   11756. .11995
```



```
                     /rpt_family="L1"
     repeat_region   12828. .13494
                     /rpt_family="PAB"
     repeat_region   14565. .14885
                     /rpt_family="Alu"
     repeat_region   16261. .16810
                     /rpt_family="L1"
     repeat_region   16798. .17124
                     /rpt_family="Alu"
     repeat_region   17113. .17483
                     /rpt_family="L1"
     repeat_region   complement(19980. .20013)
                     /rpt_family="LOR1"
     repeat_region   20252. .20407
                     /rpt_family="MIR"
     misc_feature    20491. .20572
                     /note="GRAIL 2 excellent exon, frame 0"
     repeat_region   complement(20768. .21256)
                     /rpt_family="LOR1"
     repeat_region   complement(22360. .22592)
                     /rpt_family="L1"
     repeat_region   complement(23665. .24708)
                     /rpt_family="L1"
     repeat_region   24716. .25098
                     /rpt_family="THE1"
     repeat_region   25099. .26632
                     /rpt_family="MSTAR"
     repeat_region   26633. .26932
                     /rpt_family="Alu"
     repeat_region   26929. .27012
                     /rpt_family="MSTAR"
     repeat_region   27013. .27386
                     /rpt_family="THE1"
     repeat_region   complement(27396. .27487)
                     /rpt_family="L1"
     repeat_region   complement(28090. .28399)
                     /rpt_family="Alu"
     repeat_region   28092. .28116
                     /note="(T)25"
                     /rpt_type=tandem
                     /rpt_unit=T
     repeat_region   28421. .29116
                     /rpt_family="MER44C"
     repeat_region   30948. .31138
                     /rpt_family="MER7"
     repeat_region   31737. .32077
                     /rpt_family="Alu"
     repeat_region   complement(33556. .33996)
                     /rpt_family="MLT2B2"
     repeat_region   35058. .35085
                     /note="(GT)14"
```



```
                     /rpt_type=tandem
                     /rpt_unit=GT
     repeat_region   complement(37354. .37610)
                     /rpt_family="Alu"
     repeat_region   complement(38544. .38617)
                     /rpt_family="MIR2"
     misc_feature    39126. .39241
                     /note="GRAIL 2 excellent exon, frame 0"
     repeat_region   complement(39902. .40045)
                     /rpt_family="L1"
     repeat_region   complement(40468. .40555)
                     /rpt_family="MML1"
     repeat_region   complement(40574. .40874)
                     /rpt_family="Alu"
     repeat_region   complement(40988. .41305)
                     /rpt_family="Alu"
     repeat_region   41007. .41026
                     /note="(T)20"
                     /rpt_type=tandem
                     /rpt_unit=T
     repeat_region   complement(41641. .42859)
                     /rpt_family="L1"
     repeat_region   complement(43007. .44131)
                     /rpt_family="L1"
     repeat_region   45580. .45900
                     /rpt_family="Alu"
     repeat_region   complement(48105. .48659)
                     /rpt_family="L1"
     repeat_region   complement(48717. .49026)
                     /rpt_family="Alu"
     repeat_region   complement(49442. .50119)
                     /rpt_family="L1"
     repeat_region   53134. .53318
                     /rpt_family="L1"
     repeat_region   53552. .53664
                     /rpt_family="MER54"
     repeat_region   53667. .53745
                     /rpt_family="MADE1"
     repeat_region   53885. .54778
                     /rpt_family="L1"
     repeat_region   complement(57541. .57685)
                     /rpt_family="MLT1b"
     repeat_region   complement(57856. .57951)
                     /rpt_family="MLT1b"
     repeat_region   complement(58014. .58357)
                     /rpt_family="MLT"
     repeat_region   complement(58810. .58892)
                     /rpt_family="MLT"
     repeat_region   complement(58871. .58975)
                     /rpt_family="MST"
     repeat_region   59034. .59111
```

```
                     /rpt_family="MSTAR"
     repeat_region   complement(59138. .59375)
                     /rpt_family="MLT1r"
     repeat_region   complement(59390. .59526)
                     /rpt_family="Alu"
     repeat_region   complement(60334. .60486)
                     /rpt_family="Alu"
     repeat_region   60565. .60687
                     /rpt_family="Alu"
     repeat_region   complement(61684. .62017)
                     /rpt_family="L1"
     repeat_region   complement(62475. .62785)
                     /rpt_family="Alu"
     repeat_region   66752. .67072
                     /rpt_family="Alu"
     repeat_region   complement(68291. .68547)
                     /rpt_family="THE1"
     repeat_region   68594. .68631
                     /note="(GT)19"
                     /rpt_type=tandem
                     /rpt_unit=GT
     repeat_region   complement(68639. .69044)
                     /rpt_family="MLT2B2"
BASE COUNT    21672 a  12123 c  12510 g  22755 t
ORIGIN

  Query Match          14.1%;  Score 18;  DB 26;  Length 69060;
  Best Local Similarity 100.0%;  Pred. No. 7.65e+00;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   5645 TTATTCTTTTAAAAGTTG 5662
          ||||||||||||||||||
Cp    108 TTATTCTTTTAAAAGTTG 91


RESULT   20
LOCUS       HUAE000662  71153 bp    DNA             PRI       02-SEP-1997
DEFINITION  Homo sapiens T-cell receptor alpha delta locus from bases 1000498
            to 1071650 (section 5 of 5) of the Complete Nucleotide Sequence.
ACCESSION   AE000662 AE000521
NID         g2358068
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 71153)
  AUTHORS   Koop,B.F., Rowen,L., Wang,K., Kuo,C.L., Seto,D., Lenstra,J.A.,
            Howard,S., Shan,W., Deshpande,P. and Hood,L.
  TITLE     The human T-cell receptor TCRAC/TCRDC (C alpha/C delta) region:
            organization, sequence, and evolution of 97.6 kb of DNA
```



```
  JOURNAL   Genomics 19 (3), 478-493 (1994)
  MEDLINE   94245236
  REMARK    This citation covers from bases 966383-1064019
REFERENCE   2  (bases 1 to 71153)
  AUTHORS   Boysen,C., Simon,M.I. and Hood,L.
  TITLE     Analysis of the 1.1-Mb human alpha/delta T-cell receptor locus with
            bacterial artificial chromosome clones
  JOURNAL   Genome Res. 330, 330-338 (1997)
REFERENCE   3  (bases 1 to 71153)
  AUTHORS   Boysen,C., Inyoul,L., Smith,T.M., Smit,A., Wang,K., Rowen,L. and
            Hood,L.
  TITLE     T-Cell Receptor Alpha Delta Locus Complete Nucleotide Sequence
  JOURNAL   Unpublished (1997)
  REMARK    This citation covers bases 1-983545 and bases 1064020-1071650
REFERENCE   4  (bases 1 to 71153)
  AUTHORS   Boysen,C., Lee,I., Smith,T.M., Smit,A., Kai,W., Lee,R. and Leroy,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-1997) Department of Molecular Biotechnology,
            University of Washington, Box 357730, Seattle, Washington 98195,
            USA
  REMARK    Complete nucleotide sequence of the human T-cell receptor alpha
            delta locus
COMMENT     Sequencing method: high redundancy shotgun.  Interspersed Repeats
            were identified with RepeatMasker (available from
            http://ftp.genome.washington.edu/RM/RepeatMasker.html) Simple
            sequence repeats were identified with sputnik (available from
            http://serac.mbt.washington.edu/~chrisa/software/sputnik.html).
FEATURES             Location/Qualifiers
     source          1. .71153
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     repeat_region   2362. .2660
                     /rpt_family="AluSc"
     repeat_region   4274. .4574
                     /rpt_family="AluSx"
     repeat_region   9569. .9796
                     /rpt_family="MIR"
     repeat_region   10535. .10834
                     /rpt_family="AluY"
     repeat_region   10835. .11138
                     /rpt_family="AluJb"
     repeat_region   complement(13900. .14190)
                     /rpt_family="AluSq"
     repeat_region   complement(20506. .20810)
                     /rpt_family="AluY"
     repeat_region   20928. .21223
                     /rpt_family="AluSx"
     repeat_region   23083. .23335
                     /rpt_family="MIR"
     repeat_region   25320. .25618
                     /rpt_family="AluSq"
```



```
     repeat_region   complement(27167. .27240)
                     /rpt_family="MIR"
     repeat_region   27306. .27428
                     /rpt_family="MIR2"
     repeat_region   complement(30158. .30403)
                     /rpt_family="MIR"
     repeat_region   33727. .33878
                     /rpt_family="MIR"
     repeat_region   complement(33920. .34213)
                     /rpt_family="AluYa5"
     repeat_region   complement(35275. .35553)
                     /rpt_family="MIR2"
     repeat_region   complement(36848. .36886)
                     /rpt_family="L1MA3"
     repeat_region   39321. .39398
                     /rpt_family="MIR"
     repeat_region   40206. .40261
                     /rpt_family="MIR"
     repeat_region   complement(40433. .40659)
                     /rpt_family="MIR"
     repeat_region   complement(40683. .40848)
                     /rpt_family="AluSx"
     repeat_region   complement(41322. .41444)
                     /rpt_family="MIR"
     repeat_region   complement(45732. .45964)
                     /rpt_family="MIR"
     repeat_region   46927. .46992
                     /rpt_family="MIR2"
     repeat_region   complement(47281. .47388)
                     /rpt_family="MIR2"
     repeat_region   49066. .49073
                     /rpt_type=tandem
                     /rpt_unit=TAT
     repeat_region   49734. .50032
                     /rpt_family="AluY"
     repeat_region   51737. .51940
                     /rpt_family="MER3"
     repeat_region   complement(52892. .53047)
                     /rpt_family="MIR"
     repeat_region   54558. .54732
                     /rpt_family="MIR2"
     repeat_region   complement(54749. .55052)
                     /rpt_family="AluSq"
     repeat_region   55059. .55302
                     /rpt_family="MIR2"
     repeat_region   57276. .57570
                     /rpt_family="AluJo"
     repeat_region   59210. .59460
                     /rpt_family="MER46"
     repeat_region   complement(61706. .61769)
                     /rpt_family="MIR2"
```



```
     repeat_region   63506. .63580
                     /rpt_family="MIR"
     repeat_region   63695. .63862
                     /rpt_family="L1ME3"
     repeat_region   63931. .63951
                     /rpt_type=tandem
                     /rpt_unit=AC
     repeat_region   64131. .64281
                     /rpt_family="MIR"
     repeat_region   65845. .66137
                     /rpt_family="AluSx"
     repeat_region   complement(66237. .66480)
                     /rpt_family="MIR2"
     repeat_region   complement(68187. .68457)
                     /rpt_family="AluJo"
     repeat_region   complement(68480. .68777)
                     /rpt_family="AluSq"
     repeat_region   complement(68778. .68837)
                     /rpt_family="FAM"
     repeat_region   69326. .69628
                     /rpt_family="AluSp"
     repeat_region   complement(69688. .69766)
                     /rpt_family="MER58A"
     repeat_region   69845. .69897
                     /rpt_family="MER5A"
     repeat_region   complement(69907. .70048)
                     /rpt_family="MER5A"
     repeat_region   complement(70060. .70225)
                     /rpt_family="MER5B"
BASE COUNT    20125 a  15457 c  16410 g  19161 t
ORIGIN

  Query Match          14.1%;  Score 18;  DB 26;  Length 71153;
  Best Local Similarity 100.0%;  Pred. No. 7.65e+00;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   6175 TTCTTTATTCTTTTAAAA 6192
          ||||||||||||||||||
Cp    112 TTCTTTATTCTTTTAAAA 95


RESULT   21
LOCUS       HUMTCRADCV  97630 bp    DNA             PRI       21-NOV-1997
DEFINITION  Human Tcr-C-delta gene, exons 1-4; Tcr-V-delta gene, exons 1-2;
            T-cell receptor alpha (Tcr-alpha) gene, J1-J61 segments; and
            Tcr-C-alpha gene, exons 1-4.
ACCESSION   M94081
NID         g2627263
KEYWORDS    T-cell receptor C-alpha; T-cell receptor C-delta; T-cell receptor
            V-delta; T-cell receptor alpha.
SOURCE      human.
```

```
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 97630)
  AUTHORS   Koop,B.F., Rowen,L., Seto,D., Kuo,C.-l., Hood,L. and Wang,K.
  TITLE     Nucleotide sequence of the 3' terminal end of the T-cell receptor
            alpha/delta locus
  JOURNAL   Unpublished (1992)
COMMENT     On Nov 20, 1997 this sequence version replaced gi:339242.
FEATURES             Location/Qualifiers
     source          1. .97630
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /germline
     exon            3168. .3464
                     /gene="Tcr-C-delta"
                     /number=1
     gene            3168. .6809
                     /gene="Tcr-C-delta"
     exon            4028. .4098
                     /gene="Tcr-C-delta"
                     /number=2
     exon            4458. .4572
                     /gene="Tcr-C-delta"
                     /number=3
     exon            5787. .6809
                     /gene="Tcr-C-delta"
                     /number=4
     gene            9295. .9868
                     /gene="Tcr-V-delta"
     exon            9295. .9598
                     /gene="Tcr-V-delta"
                     /number=1
     exon            9832. .9868
                     /gene="Tcr-V-delta"
                     /number=2
     gene            <15555. .>85296
                     /gene="Tcr-alpha"
     CDS             <15555. .>15627
                     /gene="Tcr-alpha"
                     /standard_name="possible J 61 gene segment"
                     /note="putative"
                     /codon_start=3
                     /db_xref="PID:g553731"
                     /translation="SCGYRVNRKLTFGANTRGIMKLS"
     exon            16556. .16610
                     /gene="Tcr-alpha"
                     /standard_name="possible J 60 gene segment"
                     /note="putative"
     CDS             <16801. .>16850
                     /gene="Tcr-alpha"
```



```
                     /standard_name="possible J 59 gene segment"
                     /note="putative"
                     /codon_start=2
                     /db_xref="PID:g553732"
                     /translation="EGRKQEIYIWNGDASE"
     exon            17953. .18016
                     /gene="Tcr-alpha"
                     /standard_name="possible J 58 gene segment"
                     /note="putative"
     CDS             <19121. .>19181
                     /gene="Tcr-alpha"
                     /standard_name="possible J 57 gene segment"
                     /note="putative"
                     /codon_start=1
                     /db_xref="PID:g553733"
                     /translation="TQGGSEKLVFGKGTKLTVNP"
     CDS             <19765. .>19829
                     /gene="Tcr-alpha"
                     /standard_name="possible J 56 gene segment"
                     /note="putative"
                     /codon_start=2
                     /db_xref="PID:g553734"
                     /translation="CYTGANSKLTFGKGITLSVRP"
     CDS             <22523. .>22593
                     /gene="Tcr-alpha"
                     /standard_name="possible J 55 gene segment"
                     /note="putative"
                     /codon_start=1
                     /db_xref="PID:g553735"
                     /translation="VLWCNSGSPEAGIWPRNQADYQP"
     CDS             <23251. .>23316
                     /gene="Tcr-alpha"
                     /standard_name="possible J 54 gene segment"
                     /note="putative"
                     /codon_start=3
                     /db_xref="PID:g553736"
                     /translation="NSGGSNYKLTFGKGTLLTVNP"
     CDS             <26473. .>26542
                     /gene="Tcr-alpha"
                     /standard_name="possible J 53 gene segment"
                     /note="putative"
                     /codon_start=1
                     /db_xref="PID:g553737"
                     /translation="ANAGGTSYGKLTFGQGTILTVHP"
     exon            27429. .27491
                     /partial
                     /gene="Tcr-alpha"
                     /standard_name="possible J 52 gene segment"
                     /note="putative"
     CDS             <27841. .>27905
                     /gene="Tcr-alpha"
```



```
                     /standard_name="possible J 51 gene segment"
                     /note="putative"
                     /codon_start=3
                     /db_xref="PID:g339250"
                     /translation="KSLTFWSSCPCAGIVGQADST"
     CDS             <28839. .>28895
                     /gene="Tcr-alpha"
                     /standard_name="possible J 50 gene segment"
                     /note="putative"
                     /codon_start=2
                     /db_xref="PID:g339251"
                     /translation="ENLLRQGDIWARDKLISHS"
     CDS             <29734. .>29789
                     /gene="Tcr-alpha"
                     /standard_name="possible J 49 gene segment"
                     /note="putative"
                     /codon_start=2
                     /db_xref="PID:g553738"
                     /translation="NTGNQFYFGTGTSLTVIP"
     CDS             <30737. .>30799
                     /gene="Tcr-alpha"
                     /standard_name="possible J 48 gene segment"
                     /note="putative"
                     /codon_start=3
                     /db_xref="PID:g553739"
                     /translation="SNFGNEKLTFGTGTRLTIIP"
     CDS             <33096. .>33152
                     /gene="Tcr-alpha"
                     /standard_name="possible J 47 gene segment"
                     /note="putative"
                     /codon_start=2
                     /db_xref="PID:g339254"
                     /translation="GIWKQTGLWRRNHSESQVL"
     CDS             <33647. .>33709
                     /gene="Tcr-alpha"
                     /standard_name="possible J 46 gene segment"
                     /note="putative"
                     /codon_start=3
                     /db_xref="PID:g553740"
                     /translation="KKSSGDKLTFGTGTRLAVRP"
     CDS             <34168. .>34234
                     /gene="Tcr-alpha"
                     /standard_name="possible J 45 gene segment"
                     /note="putative"
                     /codon_start=1
                     /db_xref="PID:g553741"
                     /translation="VYSGGGADGLTFGKGTHLIIQP"
     CDS             <35064. .>35126
                     /gene="Tcr-alpha"
                     /standard_name="possible J 44 gene segment"
                     /note="putative"
```



```
                     /codon_start=3
                     /db_xref="PID:g553742"
                     /translation="NTGTASKLTFGTGTRLQVTL"
     CDS             <36139. .>36207
                     /gene="Tcr-alpha"
                     /standard_name="possible J 43 gene segment"
                     /note="putative"
                     /codon_start=3
                     /db_xref="PID:g553743"
                     /translation="MYYCDNNNDMRPGAGTRLTVKP"
     CDS             <37130. .>37195
                     /gene="Tcr-alpha"
                     /standard_name="possible J 42 gene segment"
                     /note="putative"
                     /codon_start=3
                     /db_xref="PID:g553744"
                     /translation="NYGGSQGNLIFGKGTKLSVKP"
     CDS             <37899. .>37961
                     /gene="Tcr-alpha"
                     /standard_name="possible J 41 gene segment"
                     /note="putative"
                     /codon_start=1
                     /db_xref="PID:g339260"
                     /translation="GTQIPGMHSTSAKAPRCWSHP"
     CDS             <39930. .>39990
                     /gene="Tcr-alpha"
                     /standard_name="possible J 40 gene segment"
                     /note="putative"
                     /codon_start=1
                     /db_xref="PID:g553745"
                     /translation="TTSGTYKYIFGTGTRLKVLA"
     CDS             <41843. .>41905
                     /gene="Tcr-alpha"
                     /standard_name="possible J 39 gene segment"
                     /note="putative"
                     /codon_start=3
                     /db_xref="PID:g553746"
                     /translation="NNNAGNMLTFGGGTRLMVKP"
...
Note: remainder of annotations omitted.

  Query Match           14.1%;  Score 18;  DB 26;  Length 97630;
  Best Local Similarity 100.0%;  Pred. No. 7.65e+00;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  40286 TTCTTTATTCTTTTAAAA 40303
          ||||||||||||||||||
Cp    112 TTCTTTATTCTTTTAAAA 95


RESULT   22
```

```
LOCUS       HS39B17    112690 bp    DNA              HTG       29-JUN-1998
DEFINITION  Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone 39B17;
            HTGS phase 1.
ACCESSION   AL023656
NID         g3281962
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 112690)
  AUTHORS   Milne,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUN-1998) Wellcome Trust Genome Campus, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquires:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 1, 1998 this sequence version replaced gi:3255973.
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and the release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc. Order of segments is not known; 800 n's separate
            segments. Unfinished: dJ39B17  Contig_ID: 00971  acc=AL023656
            Length: 20620 bp Unfinished: dJ39B17  Contig_ID: 01818
            acc=AL023656  Length: 91270 bp.                          ***
            ***                                                      ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                      ***
            * This sequence is unfinished. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
FEATURES             Location/Qualifiers
     source          1. .112690
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="39B17"
BASE COUNT    35437 a  20655 c  20265 g  35532 t    801 others
ORIGIN

  Query Match           14.1%;  Score 18;  DB 17;  Length 112690;
  Best Local Similarity 100.0%;  Pred. No. 7.65e+00;
  Matches    18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  14750 ATTTTTAAAGCAGAATTT 14767
          ||||||||||||||||||
Qy      8 ATTTTTAAAGCAGAATTT 25

RESULT   23
```



```
LOCUS       CEY17D7    148583 bp    DNA              HTG       16-OCT-1997
DEFINITION  Caenorhabditis elegans DNA *** SEQUENCING IN PROGRESS *** from
            clone Y17D7; HTGS phase 1.
ACCESSION   Z92818
NID         g2546912
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryotae; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 148583)
  AUTHORS   Kershaw,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-OCT-1997) Nematode Sequencing Project, Sanger Centre,
            Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics,
            Washington University, St. Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     On Oct 18, 1997 this sequence version replaced gi:2511619.
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and
            the release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated
            with foreign sequence from E.coli, yeast, vector, phage etc. Order
            of segments is not known; 800 n's separate segments.
            ***                                                      ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                      ***
            * This sequence is unfinished. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
FEATURES             Location/Qualifiers
     source          1. .148583
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /clone="Y17D7"
                     /chromosome="V"
BASE COUNT    48332 a  25820 c  26352 g  48079 t
ORIGIN

  Query Match           14.1%;  Score 18;  DB 17;  Length 148583;
  Best Local Similarity 100.0%;  Pred. No. 7.65e+00;
  Matches    18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  18061 TTTTTAAAGCAGAATTTT 18078
          ||||||||||||||||||
Qy      9 TTTTTAAAGCAGAATTTT 26

RESULT   24
```



```
LOCUS       AC005061   161824 bp    DNA              HTG       12-JUN-1998
DEFINITION  *** SEQUENCING IN PROGRESS *** Homo sapiens clone RG087E15; HTGS
            phase 1, 8 unordered pieces.
ACCESSION   AC005061
NID         g3212930
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 161824)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 161824)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-1998) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     ***                                                      ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                      ***
            * This sequence is unfinished. It consists of 8 contigs for
            * which the order is not known; their order in this record is
            * arbitrary. In some cases, the exact lengths of the gaps
            * between the contigs are also unknown; these gaps are  presented
            * as runs of N as a convenience only. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
            *     1     1774: contig of 1774 bp in length
            *  1775     1791: gap of unknown length
            *  1792     3521: contig of 1730 bp in length
            *  3522     3538: gap of unknown length
            *  3539     8331: contig of 4793 bp in length
            *  8332     8348: gap of unknown length
            *  8349    15298: contig of 6950 bp in length
            * 15299    15315: gap of unknown length
            * 15316    32283: contig of 16968 bp in length
            * 32284    32300: gap of unknown length
            * 32301    58572: contig of 26272 bp in length
            * 58573    58589: gap of unknown length
            * 58590    94140: contig of 35551 bp in length
            * 94141    94157: gap of unknown length
            * 94158   161824: contig of 67667 bp in length.
FEATURES             Location/Qualifiers
     source          1. .161824
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RG087E15"
BASE COUNT    46538 a  34103 c  33454 g  47610 t    119 others
```



```
ORIGIN

  Query Match           14.1%;  Score 18;  DB 17;  Length 161824;
  Best Local Similarity 100.0%;  Pred. No. 7.65e+00;
  Matches    18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db 115893 TTGAGAATCTGAAAAGTG 115910
          ||||||||||||||||||
Cp     78 TTGAGAATCTGAAAAGTG 61


RESULT   25
LOCUS       HS107N3 .   174045 bp    DNA              PRI       09-JUL-1996
DEFINITION  Human DNA sequence from PAC 107N3, between markers DXS6791 and
            DXS8038 on chromosome X.
ACCESSION   Z75741
NID         g1419560
KEYWORDS    X.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 174045)
  AUTHORS   Deadman,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JUL-1996) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1RQ, UK. E-mail enquires: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     IMPORTANT: This sequence is the entire insert of clone 107N3. The
            true left end of clone 107N3 is at 1 in this sequence. The true
            right end of clone 107N3 is at 174045.
            107N3 is from the human PAC library.
FEATURES             Location/Qualifiers
     source          1. .174045
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="X"
                     /clone="107N3"
     repeat_region   5. .233
                     /note="L1 element fragment"
     repeat_region   294. .371
                     /note="MIR2 element fragment"
     repeat_region   849. .1289
                     /note="MER28 element fragment"
     repeat_region   1043. .1225
                     /note="TIGGER2 element fragment"
     repeat_region   4331. .4395
                     /note="MER43 element fragment"
     repeat_region   5008. .5085
                     /note="MER43 element fragment"
```

```
     repeat_region   5195. .5275
                     /note="MLT2B2 element fragment"
     repeat_region   5198. .5457
                     /note="MLT2A1 element fragment"
     repeat_region   5536. .5592
                     /note="MLT2A1 element fragment"
     repeat_region   5707. .5815
                     /note="MER43 element fragment"
     repeat_region   6411. .6542
                     /note="MER25 element fragment"
     repeat_region   6609. .6914
                     /note="L1 element fragment"
     repeat_region   7548. .7640
                     /note="L1 element fragment"
     repeat_region   7683. .8143
                     /note="L1 element fragment"
     repeat_region   8240. .8360
                     /note="L1 element fragment"
     repeat_region   8598. .8859
                     /note="L1 element fragment"
     repeat_region   9043. .9475
                     /note="L1 element fragment"
     repeat_region   9541. .9892
                     /note="L1 element fragment"
     repeat_region   10018. .11250
                     /note="L1 element fragment"
     repeat_region   11273. .11320
                     /note="4 copies of 12 mer 94 % conserved"
     repeat_region   11273. .11320
                     /note="24 copies of 2 mer 94 % conserved"
     repeat_region   11853. .13208
                     /note="L1 element fragment"
     repeat_region   13732. .13803
                     /note="MSTA element fragment"
     repeat_region   13738. .13925
                     /note="THE1B element fragment"
     repeat_region   13994. .14080
                     /note="THE1B element fragment"
     repeat_region   15476. .15527
                     /note="2 copies of 26 mer 96 % conserved"
     repeat_region   18439. .18595
                     /note="THE1BR element fragment"
     repeat_region   18822. .18873
                     /note="MSTA element fragment"
     repeat_region   18860. .18926
                     /note="THE1B element fragment"
     repeat_region   19342. .19630
                     /partial
                     /note="Alu repeat: matches 308. .1 of consensus"
     repeat_region   19686. .19967
                     /partial
```



```
                     /note="Alu repeat: matches 297. .1 of consensus"
     repeat_region   20027. .20095
                     /note="THE1B element fragment"
     repeat_region   20257. .20338
                     /note="THE1B element fragment"
     repeat_region   20269. .20363
                     /note="MSTA element fragment"
     repeat_region   25410. .26007
                     /note="L1 element fragment"
     repeat_region   26232. .26513
                     /partial
                     /note="Alu repeat: matches 308. .1 of consensus"
     repeat_region   28615. .28653
                     /note="MADE1 element fragment"
     repeat_region   30327. .30473
                     /note="MER25 element fragment"
     repeat_region   31929. .32039
                     /note="L1 element fragment"
     repeat_region   32095. .32644
                     /note="L1 element fragment"
     repeat_region   33098. .33513
                     /note="L1 element fragment"
     repeat_region   33657. .33778
                     /note="MLT1A element fragment"
     repeat_region   34094. .34225
                     /note="L1 element fragment"
     repeat_region   35068. .35434
                     /note="L1 element fragment"
     repeat_region   35491. .35822
                     /note="L1 element fragment"
     repeat_region   36004. .36090
                     /note="L1 element fragment"
     repeat_region   36197. .36271
                     /note="L1 element fragment"
     repeat_region   36755. .36859
                     /note="MIR2 element fragment"
     repeat_region   37196. .37303
                     /note="L1 element fragment"
     repeat_region   37822. .38320
                     /note="MER42A element fragment"
     repeat_region   38408. .38500
                     /note="MER42A element fragment"
     repeat_region   38918. .39342
                     /note="L1 element fragment"
     repeat_region   41937. .42224
                     /note="Alu repeat: matches 1. .308 of consensus"
     repeat_region   42225. .42252
                     /note="14 copies of 2 mer 100 % conserved"
     repeat_region   42972. .43286
                     /note="L1 element fragment"
     repeat_region   43296. .43587
```



```
                     /partial
                     /note="Alu repeat: matches 306. .1 of consensus"
     repeat_region   43833. .44015
                     /partial
                     /note="Alu repeat: matches 115. .308 of consensus"
     repeat_region   44244. .44297
                     /note="MER44C element fragment"
     repeat_region   44419. .44578
                     /note="MER44A element fragment"
     repeat_region   45464. .45683
                     /note="110 copies of 2 mer 83 % conserved"
     repeat_region   45467. .45682
                     /note="18 copies of 12 mer 83 % conserved"
     repeat_region   45487. .45694
                     /note="8 copies of 26 mer 81 % conserved"
     repeat_region   46907. .47113
                     /note="MIR element fragment"
     repeat_region   48665. .48796
                     /note="MLT1A element fragment"
     repeat_region   48884. .49142
                     /note="THE1B element fragment"
     repeat_region   49174. .49245
                     /note="36 copies of 2 mer 82 % conserved"
     repeat_region   49183. .49242
                     /note="5 copies of 12 mer 85 % conserved"
     repeat_region   49321. .49422
                     /note="MLT1A element fragment"
     repeat_region   49424. .49610
                     /note="L1 element fragment"
     repeat_region   50879. .51237
                     /note="L1 element fragment"
     repeat_region   52212. .52506
                     /note="Alu repeat: matches 1. .308 of consensus"
     repeat_region   52522. .52692
                     /note="MLT2B2 element fragment"
     repeat_region   52830. .53902
                     /note="MER7B element fragment"
     repeat_region   53982. .54050
                     /note="MER7B element fragment"
     repeat_region   54873. .55033
                     /note="L1 element fragment"
     repeat_region   55170. .56020
                     /note="L1 element fragment"
     repeat_region   56491. .56952
                     /note="MLT1D element fragment"
     repeat_region   57593. .57622
                     /note="15 copies of 2 mer 87 % conserved"
     repeat_region   61852. .62191
                     /note="L1 element fragment"
     repeat_region   62856. .63162
                     /note="L1 element fragment"
```



```
     repeat_region   63212. .63548
                     /note="L1 element fragment"
     repeat_region   64146. .64292
                     /note="L1 element fragment"
     repeat_region   64336. .64935
                     /note="L1 element fragment"
     repeat_region   65038. .65445
                     /note="L1 element fragment"
...
Note: remainder of annotations omitted.

  Query Match          14.1%;  Score 18;  DB 25;  Length 174045;
  Best Local Similarity 100.0%;  Pred. No. 7.65e+00;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db 135976 TTAAAAGAATAAAGAATA 135993
          ||||||||||||||||||
Qy     97 TTAAAAGAATAAAGAATA 114


RESULT   26
LOCUS       AC004925   179793 bp    DNA             HTG       12-JUN-1998
DEFINITION  *** SEQUENCING IN PROGRESS *** Homo sapiens clone DJ0907C10: HTGS
            phase 1, 13 unordered pieces.
ACCESSION   AC004925
NID         g3213072
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 179793)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 179793)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-1998) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     ***                                                          ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                          ***
            * This sequence is unfinished. It consists of 13 contigs for
            * which the order is not known; their order in this record is
            * arbitrary. In some cases, the exact lengths of the gaps
            * between the contigs are also unknown; these gaps are  presented
            * as runs of N as a convenience only. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
```

```
              *        1     1458: contig of 1458 bp in length
              *     1459     1476: gap of unknown length
              *     1477     3405: contig of 1929 bp in length
              *     3406     3423: gap of unknown length
              *     3424     5125: contig of 1702 bp in length
              *     5126     5143: gap of unknown length
              *     5144     6732: contig of 1589 bp in length
              *     6733     6750: gap of unknown length
              *     6751     8670: contig of 1920 bp in length
              *     8671     8688: gap of unknown length
              *     8689    11246: contig of 2558 bp in length
              *    11247    11264: gap of unknown length
              *    11265    12818: contig of 1554 bp in length
              *    12819    12836: gap of unknown length
              *    12837    15625: contig of 2789 bp in length
              *    15626    15643: gap of unknown length
              *    15644    18376: contig of 2733 bp in length
              *    18377    18394: gap of unknown length
              *    18395    22426: contig of 4032 bp in length
              *    22427    22444: gap of unknown length
              *    22445    24291: contig of 1847 bp in length
              *    24292    24309: gap of unknown length
              *    24310    30187: contig of 5878 bp in length
              *    30188    30205: gap of unknown length
              *    30206   179793: contig of 149588 bp in length.
FEATURES             Location/Qualifiers
     source          1. .179793
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="DJ0907C10"
BASE COUNT    57058 a  33993 c  33707 g  54819 t    216 others
ORIGIN

  Query Match           14.1%;  Score 18;  DB 17;  Length 179793;
  Best Local Similarity 100.0%;  Pred. No. 7.65e+00;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db 104944 TCAAAAGGGTCATTTTTT 104961
          ||||||||||||||||||
Qy     28 TCAAAAGGGTCATTTTTT 45


RESULT   27
LOCUS       AC002461   197273 bp     DNA              PRI       20-AUG-1997
DEFINITION  Human BAC clone RG204I16 from 7q31, complete sequence.
ACCESSION   AC002461
NID         g2337868
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
```



```
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 197273)
  AUTHORS   Blanchard,M, Johnson,D and O'Brien,D.
  TITLE     The sequence of H. sapiens BAC clone RG204I16
  JOURNAL   Unpublished (1997)
REFERENCE   2  (bases 1 to 197273)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-AUG-1997) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     SUBMITTED BY:
            Genome Sequencing Center
            Department of Genetics
            Washington University
            St. Louis MO 63108, USA
            http://genome.wustl.edu/gsc
            mailto:sapiens@watson.wustl.edu

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded or sequenced with an alternate
            chemistry; an attempt was made to resolve all sequencing problems,
            such as compressions and repeats; all regions were covered by
            sequence from more than one subclone; and the assembly was
            confirmed by restriction digest.

            MAPPING INFORMATION:
            The sequence of this clone was established as part of a mapping and
            sequencing collaboration between the NHGRI Chromosome 7 Mapping
            Project (Eric D. Green, Director), John D. McPherson in the
            Department of Genetics (Washington University), and the Washington
            University Genome Sequencing Center.  For additional information
            about the map position of this sequence, see
            http://www.nhgri.nih.gov/DIR/GTB/CHR7 or send
            mailto:egreen@nhgri.nih.gov

            SOURCE INFORMATION:
            This clone is from a release of the human BAC library. The library
            contains cloned DNA from human sperm.  See:  Shizuya et al., Proc.
            Natl. Acad. Sci. USA 89:8794-7 (1992); Kim et al., Genomics
            34:213-8 (1996).  The clone is available from Research Genetics,
            Inc. (http://www.resgen.com).
            VECTOR:  pBeloBAC11
            Selection:  chloramphenicol

            NEIGHBORING SEQUENCE INFORMATION:
```



```
            The clone sequenced to the right is RG208003, 200 bp overlap.
            Actual start of this clone is at base position 1 of RG204I16;
            actual end is at 197273 of RG204I16.  This clone is part of an
            unanchored island, orientation is unknown.

            This clone contains STS's sWSS2434 (NID:g1113491), sWSS3360
            (NID:g1185343).
FEATURES             Location/Qualifiers
     source          1. .197273
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /clone="RG204I16"
                     /clone_lib="CITB-HS-A"
                     /map="7q31"
     repeat_region   782. .1072
                     /rpt_family="ALU"
     repeat_region   complement(1809. .1855)
                     /rpt_family="L1"
     repeat_region   2112. .2377
                     /rpt_family="L1"
     repeat_region   2420. .3533
                     /rpt_family="L1"
     repeat_region   complement(3023. .3050)
                     /rpt_family="L1"
     repeat_region   3551. .3611
                     /rpt_family="L1"
     repeat_region   complement(3587. .3631)
                     /rpt_family="L1"
     repeat_region   3630. .3990
                     /rpt_family="L1"
     repeat_region   complement(3991. .4323)
                     /rpt_family="MER"
     repeat_region   4344. .6304
                     /rpt_family="L1"
     repeat_region   complement(4888. .5307)
                     /rpt_family="L1"
     repeat_region   6543. .6580
                     /rpt_family="L1"
     repeat_region   complement(8349. .8477)
                     /rpt_family="L1"
     repeat_region   13289. .13328
                     /rpt_family="L1"
     repeat_region   complement(13581. .13599)
                     /rpt_family="L1"
     repeat_region   complement(13600. .13889)
                     /rpt_family="ALU"
     repeat_region   complement(13973. .13990)
                     /rpt_family="L1"
     repeat_region   16029. .16103
                     /rpt_family="L1"
```



```
     repeat_region   18606. .18767
                     /rpt_family="L1"
     repeat_region   21348. .21367
                     /rpt_family="L1"
     repeat_region   21760. .21814
                     /rpt_family="L1"
     repeat_region   22363. .22387
                     /rpt_family="L1"
     repeat_region   24746. .24764
                     /rpt_family="L1"
     repeat_region   26901. .27053
                     /rpt_family="L1"
     repeat_region   32206. .32233
                     /rpt_family="L1"
     repeat_region   complement(37294. .37970)
                     /rpt_family="L1"
     repeat_region   complement(37991. .38289)
                     /rpt_family="L1"
     repeat_region   complement(40832. .40865)
                     /rpt_family="L1"
     repeat_region   complement(42572. .42593)
                     /rpt_family="L1"
     repeat_region   complement(44907. .44940)
                     /rpt_family="L1"
     repeat_region   45050. .45078
                     /rpt_family="L1"
     repeat_region   48153. .48326
                     /rpt_family="L1"
     repeat_region   48642. .49026
                     /rpt_family="ALU"
     repeat_region   49584. .49611
                     /rpt_family="L1"
     repeat_region   complement(51821. .51844)
                     /rpt_family="L1"
     repeat_region   complement(51845. .52136)
                     /rpt_family="ALU"
     repeat_region   54122. .54332
                     /rpt_family="ALU"
     repeat_region   54784. .54809
                     /rpt_family="L1"
     repeat_region   complement(56195. .56342)
                     /rpt_family="ALU"
     repeat_region   57519. .57770
                     /rpt_family="ALU"
     repeat_region   complement(59150. .59252)
                     /rpt_family="L1"
     repeat_region   61543. .61833
                     /rpt_family="ALU"
     repeat_region   complement(62467. .62588)
                     /rpt_family="L1"
     repeat_region   complement(62804. .64357)
```

```
                     /rpt_family="THR"
     repeat_region   complement(64511. .64660)
                     /rpt_family="L1"
     repeat_region   65462. .65506
                     /rpt_family="L1"
     repeat_region   complement(66053. .66107)
                     /rpt_family="L1"
     repeat_region   67125. .67151
                     /rpt_family="L1"
     repeat_region   72066. .72152
                     /rpt_family="L1"
     repeat_region   complement(72153. .72944)
                     /rpt_family="L1"
     repeat_region   72912. .72968
                     /rpt_family="L1"
     repeat_region   73953. .73970
                     /rpt_family="L1"
     repeat_region   74040. .74108
                     /rpt_family="L1"
     repeat_region   complement(78047. .78109)
                     /rpt_family="ALU"
     repeat_region   78110. .78372
                     /rpt_family="THE"
     repeat_region   80202. .80233
                     /rpt_family="L1"
     repeat_region   complement(82619. .82731)
                     /rpt_family="L1"
     repeat_region   complement(83668. .83694)
                     /rpt_family="L1"
     repeat_region   85909. .87532
                     /rpt_family="L1"
     repeat_region   complement(86124. .86541)
                     /rpt_family="L1"
     repeat_region   87632. .87700
                     /rpt_family="L1"
     repeat_region   88488. .88528
                     /rpt_family="L1"
     repeat_region   88613. .88909
                     /rpt_family="ALU"
...
Note: remainder of annotations omitted.

  Query Match           14.1%;  Score 18;  DB 26;  Length 197273;
  Best Local Similarity 100.0%;  Pred. No. 7.65e+00;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  80489 TCTCAACTTTTAAAAGAA 80506
          ||||||||||||||||||
Qy     88 TCTCAACTTTTAAAAGAA 105
```



```
RESULT   28
LOCUS       AC005703   236157 bp    DNA             HTG       06-OCT-1998
DEFINITION  *** SEQUENCING IN PROGRESS *** Homo sapiens chromosome 17, clone
            hRPK.849_N_15; HTGS phase 1, 4 unordered pieces.
ACCESSION   AC005703
NID         g3687274
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 236157)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone hRPK.849_N_15
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 236157)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boutwell,C.,
            Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., Corliss,D., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R.,
            Gage,D., Gardyna,S., Geraigery,K., Grant,G., Hagos,B., Heaford,A.,
            Herena,L., Horton,L., Howland,J.C., Jacotot,L., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Macdonald,P., Marquis,N., McEwan,P.,
            McGurk,A., McKernan,K., Meldrim,J., Molla,M., Morris,W., Morrow,J.,
            Mychaleckyj,J., Nahf,R., Naylor,J., Niloff,M., O'Connor,T.,
            O'Donnell,P., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A.,
            Severy,P., Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C.,
            Subramanian,A., Tesfaye,S., Tichovolsky,N., Torruella-Miller,I.,
            Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D.,
            Ye,W.J., Zhao,J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-SEP-1998) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Oct 1, 1998 this sequence version replaced gi:3668098.
            All repeats were identified using RepeatMasker: Smit, A.F.A. &
            Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html.
            ***                                                          ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                          ***
            * This sequence is unfinished. It consists of 4 contigs for
            * which the order is not known; their order in this record is
            * arbitrary. In some cases, the exact lengths of the gaps
            * between the contigs are also unknown; these gaps are  presented
            * as runs of N as a convenience only. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
            *       1    25656: contig of 25656 bp in length
            *                   gap of unknown length
            *   25657   101180: contig of 75524 bp in length
```



```
            *                   gap of unknown length
            *  101181   199751: contig of 98571 bp in length
            *                   gap of unknown length
            *  199752   236157: contig of 36406 bp in length.
FEATURES             Location/Qualifiers
     source          1. .236157
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="hRPK.849_N_15"
                     /clone_lib="RPCI-11 human BAC library"
                     /map="17"
                     /chromosome="17"
BASE COUNT    66393 a  50700 c  49942 g  69122 t
ORIGIN

  Query Match           14.1%;  Score 18;  DB 17;  Length 236157;
  Best Local Similarity 100.0%;  Pred. No. 7.65e+00;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db 161024 TCAAAAGGGTCATTTTTT 161041
          ||||||||||||||||||
Qy     28 TCAAAAGGGTCATTTTTT 45


RESULT   29
LOCUS       LETXETB2     1279 bp    RNA             PLN       24-AUG-1995
DEFINITION  L.esculentum tXET-B2 mRNA for xyloglucan endo-transglycosylase.
ACCESSION   X82684
NID         g577067
KEYWORDS    xyloglucan endo-transglycosylase.
SOURCE      tomato.
  ORGANISM  Solanum lycopersicum
            Eukaryotae; mitochondrial eukaryotes; Viridiplantae;
            Charophyta/Embryophyta group; Embryophyta; Magnoliophyta;
            Magnoliopsida; Solananae; Solanales; Solanaceae; Solanum.
REFERENCE   1  (bases 1 to 1279)
  AUTHORS   Arrowsmith,D.A. and de Silva,J.
  TITLE     Characterisation of two tomato fruit-expressed cDNAs encoding
            xyloglucan endo-transglycosylase
  JOURNAL   Plant Mol. Biol. 28 (3), 391-403 (1995)
  MEDLINE   95359399
REFERENCE   2  (bases 1 to 1279)
  AUTHORS   Arrowsmith,D.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-1994) D.A. Arrowsmith, Unilever Research,
            Colworth House, Sharnbrook, Bedford, MK44 1LQ, UK
FEATURES             Location/Qualifiers
     source          1. .1279
                     /organism="Solanum lycopersicum"
                     /cultivar="VFN8"
                     /sub_species="Ailsa Craig"
```



```
                     /db_xref="taxon:4081"
                     /dev_stage="breaker"
                     /tissue_type="whole fruit"
                     /clone_lib="lambda gt11"
                     /clone="pTXET-B2"
     sig_peptide     232. .285
                     /gene="tXET-B2"
     gene            232. .1095
                     /gene="tXET-B2"
     CDS             232. .1095
                     /gene="tXET-B2"
                     /codon_start=1
                     /product="xyloglycan endo-transglycosylase"
                     /db_xref="PID:g577068"
                     /translation="MLLQLSLLTLVLLSPVSADNFYQDAAVTPGDQRAQIQDGGRLLT
                     LSLDKISGSGPQSKNEYLFGRFDMQLKLVPGNSAGTVTTFYLSSQSGAGHDEIDFEFLG
                     NSSGLPYTVHTNVYSQGKGNKEQQFRLWFDPTSSFHTYSIVWNSQRIIFLVDNIPIRV
                     FNNHEALGVAYPKNQAMRVYASLWNADDWATQGGRVKTDWSMAPFTASYRNFNTNACV
                     WSAATSTSSCGGSKTESVNNDETWQTQQLNANGRNRIRWVQQKYMIYNYCADANRFSQ
                     GFSPECKRSRF"
BASE COUNT      369 a    236 c    269 g    405 t
ORIGIN

  Query Match           13.3%;  Score 17;  DB 23;  Length 1279;
  Best Local Similarity 100.0%;  Pred. No. 3.26e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1211 ATATTCTTTATTCTTTT 1227
          |||||||||||||||||
Cp    115 ATATTCTTTATTCTTTT 99


RESULT   30
LOCUS       A44631       1289 bp    DNA             PAT       07-MAR-1997
DEFINITION  Sequence 7 from Patent WO9513384.
ACCESSION   A44631
NID         g2299404
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1289)
  AUTHORS   De,S.J., Whiteman,S., Hellyer,S.A. and Arrowsmith,D.A.
  TITLE     TOMATO XYLOGLUCAN ENDO-TRANSGLYCOSYLASE
  JOURNAL   Patent: WO 9513384-A 7 18-MAY-1995;
            UNILEVER PLC (GB)
COMMENT     Other publication CA 2176133 950518
            Other publication AU 8112994 950529.
FEATURES             Location/Qualifiers
     source          1. .1289
                     /organism="unidentified"
```

```
                     /db_xref="taxon:32644"
     CDS             232. .1095
                     /note="unnamed protein product"
                     /codon_start=1
                     /db_xref="PID:e306063"
                     /db_xref="PID:g2299405"
                     /translation="MLLQLSLLTLVLLSPVSADNFYQDAAVTFGDQRAQIQDGGRLLT
                     LSLDKISGSGFQSKNEYLFGRFDMQLKLVPGNSAGTVTTFYLSSQGAGHDEIDFEFLG
                     NSSGLPYTVHTNVYSQGKGNKEQQFRLWFDPTSSFHTYSIVWNSQRIIFLVDNIPIRV
                     FNNHEALGVAYPKNQAMRVYASLWNADDWATQGGRVKTDWSMAPFTASYRNFNTNACV
                     WSAATSTSSCGGSKTESVNNDETWQTQQLNANGRNRIRWVQQKYMIYNYCADNRFSQ
                     GFSPECKRSRF"
BASE COUNT      371 a    240 c    271 g    407 t
ORIGIN

  Query Match           13.3%;  Score 17;  DB 21;  Length 1289;
  Best Local Similarity 100.0%;  Pred. No. 3.26e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1211 ATATTCTTTATTCTTTT 1227
         |||||||||||||||||
Cp    115 ATATTCTTTATTCTTTT 99


RESULT   31
LOCUS       DMFUGENE     1794 bp    DNA             INV       06-APR-1993
DEFINITION  D.melanogaster fused gene.
ACCESSION   X55759
NID         g288918
KEYWORDS    segmentation gene; serine/threonine protein kinase.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Arthropoda;
            Tracheata; Insecta; Pterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1794)
  AUTHORS   Preat,T., Therond,P., Lamour-Isnard,C., Limbourg-Bouchon,B.,
            Tricoire,H., Erk,I., Mariol,M.C. and Busson,D.
  TITLE     A putative serine/threonine protein kinase encoded by the
            segment-polarity fused gene of Drosophila
  JOURNAL   Nature 347 (6288), 87-89 (1990)
  MEDLINE   90370097
FEATURES             Location/Qualifiers
     source          1. .1794
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="X"
                     /map="17C-D"
     gene            904. .1794
                     /note="fused"
                     /gene="fu"
```



```
                     /allele=""
                     /db_xref="FlyBase:FBgn0001079"
     CDS             904. .>1794
                     /gene="fu"
                     /codon_start=1
                     /product="serine/threonine protein kinase"
                     /db_xref="PID:g288919"
                     /db_xref="FLYBASE:FBgn0001079"
                     /db_xref="SWISS-PROT:P23647"
                     /translation="MNRYAVSSLVQGGSPGCVYKATRKDDSKVVAIKVISKVSGAGQV
                     IKQQVHTTSSHHIHVLQRGRATKELKNLRRECDIQARLKHPHVIEMIESFESKTDLFV
                     VTEFALMDLHRYISYNGAMGEEPARRVTGHLVSALYYLHSNRILHRDLKPQNVLLDKN
                     MHAKLCDFGLARNMTLGTHVLTSIKGTPLYMAPELLADEPYDHHADMWSLGCIAYESM
                     AGQPPFCASSILHLVKMIKHEDVKWPSTLTCECRSFLQGLLEKDPGLRISWTQLLCHP
                     FVEGRIFIAETQAEAAKESPF"
BASE COUNT      462 a    446 c    451 g    435 t
ORIGIN

  Query Match           13.3%;  Score 17;  DB 18;  Length 1794;
  Best Local Similarity 100.0%;  Pred. No. 3.26e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    367 TGTCAAAAGGGTCATTT 383
          |||||||||||||||||
Qy     26 TGTCAAAAGGGTCATTT 42


RESULT   32
LOCUS       MIU09180     2908 bp    DNA             INV       29-DEC-1994
DEFINITION  Meloidogyne incognita esophageal gland protein SEC-1 (sec-1) gene,
            complete cds.
ACCESSION   U09180
NID         g558752
KEYWORDS    .
SOURCE      southern root-knot nematode.
  ORGANISM  Meloidogyne incognita
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Nematoda;
            Secernentea; Diplogasteria; Tylenchida; Tylenchina; Tylenchoidea;
            Heteroderidae; Meloidogyninae; Meloidogyne.
REFERENCE   1  (bases 1 to 2908)
  AUTHORS   Ray,C., Abbott,A.G. and Hussey,R.S.
  TITLE     Trans-splicing of a Meloidogyne incognita mRNA encoding a putative
            esophageal gland protein
  JOURNAL   Mol. Biochem. Parasitol. 68 (1), 93-101 (1994)
  MEDLINE   95198777
REFERENCE   2  (bases 1 to 2908)
  AUTHORS   Ray,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-APR-1994) Celeste Ray, Plant Pathology, University of
            Georgia, Athens, GA 30602-7274, USA
FEATURES             Location/Qualifiers
```



```
     source          1. .2908
                     /organism="Meloidogyne incognita"
                     /strain="Race 3"
                     /specific_host="Glycine max"
                     /db_xref="taxon:6306"
                     /lab_host="Lycopersicon esculentum"
                     /clone="pCRH10 and pCRH15"
                     /clone_lib="M. incognita female ZAP-cDNA library and M.
                     incognita EMBL"
     TATA_signal     60. .64
     misc_feature    76
                     /function="trans-spliced leader acceptor site"
                     /evidence=experimental
     gene            join(84. .188,230. .413,456. .613,904. .1092,1130. .1390,
                     1639. .1785,1827. .2045,2092. .2187,2233. .2379,2590. .2694)
                     /gene="sec-1"
     CDS             join(84. .188,230. .413,456. .613,904. .1092,1130. .1390,
                     1639. .1785,1827. .2045,2092. .2187,2233. .2379,2590. .2694)
                     /gene="sec-1"
                     /codon_start=1
                     /function="esophageal gland protein"
                     /product="SEC-1"
                     /db_xref="PID:g558753"
                     /translation="MDIITIGGLVAVVAIVIGSCFTIYKLLTRETTFEEVYGNDMAKP
                     LVAEKGKAKGKRNHSKRQSPKKETDNKEVHVSEDDEASGDFVVVKQMEEATKELNLVS
                     SSTSEVNDELENLRIENAELKKEKSEQAEKLVAFEEMIKNKTIEPKQMDDENKKLRSK
                     NDQFSSQLEKQLKDLEHLRAENDQLKREKNEQDERHVIFEQTVKHLLIETHQLGDENK
                     KLRLKCEQFCEQLEKKKTEIAKFCEDRDVAMKEVLKLREERDLAKTFMEKAAADIHSQ
                     DIMIRNMETNFEHQRQALQIEFNNERGHFKKVEQEWLNANGQLANELTNANEFVRDLE
                     YKKSQVEQKLVDEQKIRSDAYAQIEKLKVEMSHASELYNNLQAKFEAVQAENEKLVSV
                     NTELKKEATHFQSQNKAMSQSSAELDQFKVTIAELNTKLDQQRSRNQELRDANYKLMD
                     TVKLLQSSKSLVNGTPPIANGGISGSHVGQDSAVVEKLALENKQFKSTLDSLNKQLEE
                     VSKIALEVEIDRDAKVAQIEKMLNEAKS"
     3'UTR           2695. .2803
     polyA_signal    2786. .2791
     polyA_site      2803
BASE COUNT      967 a    417 c    549 g    975 t
ORIGIN

  Query Match           13.3%;  Score 17;  DB 18;  Length 2908;
  Best Local Similarity 100.0%;  Pred. No. 3.26e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1863 TTAAAAGTTGAGATGAG 1879
         |||||||||||||||||
Cp    100 TTAAAAGTTGAGATGAG 84


RESULT   33
LOCUS       AB014563     4365 bp    mRNA            PRI       15-JUL-1998
DEFINITION  Homo sapiens mRNA for KIAA0663 protein, complete cds.
```



```
ACCESSION   AB014563
NID         g3327139
KEYWORDS    .
SOURCE      Homo sapiens adult male brain cDNA to mRNA, clone_lib:pBluescriptII
            SK plus clone:HK02006.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Ishikawa,K., Nagase,T., Suyama,M., Miyajima,N., Tanaka,A.,
            Kotani,H., Nomura,N. and Ohara,O.
  TITLE     Prediction of the coding sequences of unidentified human genes. X.
            The complete sequences of 100 new cDNA clones from brain which can
            code for large proteins in vitro
  JOURNAL   DNA Res. 5 (3), 169-176 (1998)
  MEDLINE   98403880
REFERENCE   2  (bases 1 to 4365)
  AUTHORS   Ohara,O., Suyama,M., Nagase,T. and Ishikawa,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-MAY-1998) to the DDBJ/EMBL/GenBank databases. Osamu
            Ohara, Kazusa DNA Research Institute, Laboratory of DNA Technology;
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            (E-mail:cdnainfo@kazusa.or.jp, Tel:+81-438-52-3913,
            Fax:+81-438-52-3914)
FEATURES             Location/Qualifiers
     source          1. .4365
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="HK02006"
                     /clone_lib="pBluescriptII SK plus"
                     /dev_stage="adult"
                     /sex="male"
                     /tissue_type="brain"
     gene            214. .2646
                     /gene="KIAA0663"
     CDS             214. .2646
                     /gene="KIAA0663"
                     /codon_start=1
                     /product="KIAA0663 protein"
                     /db_xref="PID:d1032599"
                     /db_xref="PID:g3327140"
                     /translation="MPNQGEDCYFFFYSTCTKGDSCPFRHCEAAIGNETVCTLWQEGR
                     CFRQVCRFRHMEIDKKRSEIPCYWEMQPTGCQKLNCAFHHNRGRVVDGLFLPPSKTVL
                     PTVPESPEEEVKASQLSVQQNKLSVQSNPSPQLRSVMKVESSENVPSPTHPPVVINAA
                     DDDEDDDDQFSEEGDETKTPTLQPTPEVHNGLRVTSVRKPAVNIKQGECLNFGIKTLE
                     EIKSKKMKEKSKKQGEGSSGVSSLLLHPEPVPGPEKENVRTVVRTVTLSTKQGEEPLV
                     RLSLTERLGKRKFSAGGDSDPPLKRSLAQRLGKKVEAPETNIDKTPKKAQVSKSLKER
                     LGMSADPDNEDATDKVNKVGEIHVKTLEEILLERASQKRGELQTKLKTEGPSKTDDST
                     SGARSSSTIRIKTPSEVLAEKKHRQQEAERQKSKKDTTCIKLKIDSEIKKTVVLPPIV
                     ASRGQSEEPAGKTKSMQEVHIKTLEEIKLEKALRVQQSSESSTSSPSQHEAIPGARRL
                     LRITKRTGMKEEKNLQEGNEVDSQSSIRTEAKEASGETTGVDITKIQVKRCETMREKH
```

```
                     MQKQQEREKSVLTPLRGDVASCNTQVAEKPVLTAVPGITRHLTKRLPTKSSQKVEVET
                     SGIGDSLLNVKCAAQTLERRGKAKPKVNVKPSVVKVVSSPKLAPKRKAVEMHAAVIAA
                     VKPLSSSSVLQEPPAKKAAVAVVPLVSEDKSVTVPEAENPRDSLVLPPTQSSSDSSPP
                     EVSGPSSSQMSMKTRRLSSASTGKPPLSVEDDFEKLIWEISGGKLEAEIDLDPGKDED
                     DLLLELSEMIDS"
BASE COUNT     1320 a    878 c    976 g   1191 t
ORIGIN

  Query Match           13.3%;  Score 17;  DB 26;  Length 4365;
  Best Local Similarity 100.0%;  Pred. No. 3.26e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   4213 ATATTCTTTATTCTTTT 4229
            |||||||||||||||||
Cp    115 ATATTCTTTATTCTTTT 99


RESULT   34
LOCUS       ATCYP450D    4937 bp    DNA             PLN       07-OCT-1996
DEFINITION  A.thaliana gene cytochrome P450.
ACCESSION   X87368
NID         g871987
KEYWORDS    Cytochrome P450.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryotae; mitochondrial eukaryotes; Viridiplantae;
            Charophyta/Embryophyta group; Embryophyta; Magnoliophyta;
            Magnoliopsida; Capparales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 4937)
  AUTHORS   Szekeres,M., Nemeth,K., Koncz-Kalman,Z., Mathur,J., Kauschmann,A.,
            Altmann,T., Redei,G.P., Nagy,F., Schell,J. and Koncz,C.
  TITLE     Brassinosteroids rescue the deficiency of CYP90, a cytochrome P450,
            controlling cell elongation and de-etiolation in Arabidopsis
  JOURNAL   Cell 85, 171-182 (1996)
  MEDLINE   96200769
REFERENCE   2  (bases 1 to 4937)
  AUTHORS   Szekeres,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAY-1995) M. Szekeres, Inst. of Plant Biology,
            Biological Research Center, Hungarian Academy of Sciences, PO Box
            521, 6701 Szeged, HUNGARY
REFERENCE   3  (bases 1 to 4937)
  AUTHORS   Szekeres,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JUN-1995) M. Szekeres, Inst. of Plant Biology,
            Biological Research Center, Hungarian Academy of Sciences, PO Box
            521, 6701 Szeged, HUNGARY
REFERENCE   4  (bases 1 to 4937)
  AUTHORS   Szekeres,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-1995) M. Szekeres, Inst. of Plant Biology,
```



```
            Biological Research Center, Hungarian Academy of Sciences, PO Box
            521, 6701 Szeged, HUNGARY
COMMENT     On Jun 28, 1995 this sequence version replaced gi:853720.
FEATURES             Location/Qualifiers
     source          1. .4937
                     /organism="Arabidopsis thaliana"
                     /cultivar="columbia"
                     /db_xref="taxon:3702"
                     /clone_lib="lambda EMBL4"
                     /clone="G206"
                     /chromosome="5"
                     /map="14.3"
     TATA_signal     755. .802
     mRNA            join(792. .1483,1680. .1829,1917. .2165,3903. .3989,
                     4084. .4162,4248. .4354,4446. .4576,4674. .4916)
     exon            792. .1483
                     /number=1
     gene            968. .4773
                     /gene="CPD"
     CDS             join(968. .1483,1680. .1829,1917. .2165,3903. .3989,
                     4084. .4162,4248. .4354,4446. .4576,4674. .4773)
                     /gene="CPD"
                     /function="loss of gene function results in dwarf
                     phenotype"
                     /note="cytochrome P450"
                     /codon_start=1
                     /product="CYP90 protein"
                     /db_xref="PID:e264618"
                     /db_xref="PID:g871988"
                     /translation="MAFTAFLLLLSSIAAGFLLLLRRTRYRRMGLPPGSLGLPLIGET
                     FQLIGAYKTENPEPFIDERVARYGSVFMTHLFGEPTIFSADPETNRFVLQNEGKLFEC
                     SYPASICNLLGKHSLLLMKGSLHKRMHSLTMSFANSSTIKDHLMLDIDRLVRPNLDSW
                     SSRVLLMEEAKKITFELTVKQLMSFDPGEWSESLRKEYLLVIEGFFSLPLPLPSTTYR
                     KAIQARRKVAEALTVVVMKRRREEEEEGAERKKDMLAALLAADDGFSDEEIVDPLVALL
                     VAGYETTSTIMTLAVKFLTETPLALAQLKEEHEKIRAMKSDSYSLEWSDYKSMPFTQC
                     VVNETLRVANIIGGVFRRAMTDVEIKGYKIPKGWKVFSSFRAVHLDPNHFKDARTFNP
                     WRWQSNSVTTGPSNVFTPFGGGPRLCPGYELARVALSVFLHRLVTGFSWVPAEQDKLV
                     FFPTTRTQKRYPIFVKRRDFAT"
     intron          1484. .1679
                     /gene="CPD"
                     /number=1
     exon            1680. .1829
                     /gene="CPD"
                     /number=2
     intron          1830. .1916
                     /gene="CPD"
                     /number=2
     exon            1917. .2165
                     /gene="CPD"
                     /number=3
     intron          2166. .3902
```



```
                     /gene="CPD"
                     /number=3
     exon            3903. .3989
                     /gene="CPD"
                     /number=4
     intron          3990. .4083
                     /gene="CPD"
                     /number=4
     exon            4084. .4162
                     /gene="CPD"
                     /number=5
     intron          4163. .4247
                     /gene="CPD"
                     /number=5
     exon            4248. .4354
                     /gene="CPD"
                     /number=6
     intron          4355. .4445
                     /gene="CPD"
                     /number=6
     exon            4446. .4576
                     /gene="CPD"
                     /number=7
     intron          4577. .4673
                     /gene="CPD"
                     /number=7
     exon            4674. .4915
                     /number=8
BASE COUNT     1526 a    864 c    870 g   1677 t
ORIGIN

  Query Match           13.3%;  Score 17;  DB 23;  Length 4937;
  Best Local Similarity 100.0%;  Pred. No. 3.26e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   3323 CTGAAAAGTGTAAAAGG 3339
            |||||||||||||||||
Cp     70 CTGAAAAGTGTAAAAGG 54


RESULT   35
LOCUS       DMFUSED      5025 bp    DNA             INV       14-OCT-1994
DEFINITION  D.melanogaster fused gene sequence.
ACCESSION   X80468
NID         g516300
KEYWORDS    fused gene; fused protein; serine/threonine kinase.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Arthropoda;
            Tracheata; Insecta; Pterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
```



```
REFERENCE   1  (bases 1 to 5025)
  AUTHORS   Therond,P., Busson,D., Guillemet,E., Limbourg-Bouchon,B., Preat,T.,
            Terracol,R., Tricoire,H. and Lamour-Isnard,C.
  TITLE     Molecular organisation and expression pattern of the segment
            polarity gene fused of Drosophila melanogaster
  JOURNAL   Mech. Dev. 44 (1), 65-80 (1993)
  MEDLINE   94206843
REFERENCE   2  (bases 1 to 5025)
  AUTHORS   Lamour-Isnard,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JUL-1994) C. Lamour-Isnard, Inst Jacques Monod,
            Universite Paris 7 CNRS, 2 Place Jussieu, 75251 Paris Cedex 05,
            FRANCE
FEATURES             Location/Qualifiers
     source          1. .5025
                     /organism="Drosophila melanogaster"
                     /strain="oregon"
                     /db_xref="taxon:7227"
                     /clone_lib="oregon r cosmid library"
                     /clone="cos c4'"
                     /chromosome="x"
     mRNA            join(818. .1014,1087. .2456,2517. .3091,3160. .4295)
                     /gene="fu"
                     /db_xref="FlyBase:FBgn0001079"
                     /product="serine threonine kinase"
     gene            818. .4295
                     /note="fused"
                     /gene="fu"
                     /allele=""
                     /db_xref="FlyBase:FBgn0001079"
     exon            904. .1014
                     /gene="fu"
                     /db_xref="FlyBase:FBgn0001079"
                     /number=1
     CDS             join(904. .1014,1087. .2456,2517. .3091,3160. .3521)
                     /gene="fu"
                     /codon_start=1
                     /product="serine threonine kinase"
                     /db_xref="PID:g516301"
                     /db_xref="FLYBASE:FBgn0001079"
                     /translation="MNRYAVSSLVGQGSLGCVYKATRKDDSKVVAIKVISKRGRATKE
                     LKNLRRECDIQARLKHPHVIEMIESFESKTDLFVVTEFALMDLHRYLSYNGAMGEEPA
                     RRVTGHLVSALYYLHSNRILHRDLKPQNVLLDKNMHAKLCDPGLARNMTLGTHVLTSI
                     KGTPLYMAPELLADEPYDHHADMWSLGCIAYEMMAGQPPFCASSILHLVKMIKHEDVK
                     WPSTLTCECRSFLQGLLEKDPGLRISWTQLLCHPFVEGRIFIAETQAEAAKESPFTNP
                     EAKVKSSKQSDPEVGDLDEALAALDFGESRQENLTTSRDSINAIAPSDVDHLETDVED
                     NMQRVVVPFADLSYRDLSGVRAMPMVHQPVINSHTCFVSGNSNMILNHMNDNFDPQAS
                     LRGGVVAAKPIVAPTVRQSRSKDLEKRKLSQNLDNFSVRLGHSVDHEAQRKATEIATQ
                     EKHNQENKPPAEAISTANSQPPQQQPQQLKHSMHSTNEEKLSSDNTPPCLLPGWDSCD
                     ESQSPPIENDEWLAFLNRSVQELLDGELDSLKQHNLVSIIVAPLRNSKAIPRVLKSVA
                     QLLSLPFVLVDPVLIVDLELIRNVYVDVKLVPNLMYACKLLLSHKQLSDSAASAPLTT
```

```
                     GSLSRTLRSIPELTVEELETACSLYELVCHLVHLQQQFLTQFCDAVAILAASDLFLNF
                     LTHDFRQSDSDAASVRLAGCMLALMSCVLRELPENAELVERIVFNPRLNFVSLLQSRH
                     HLLGQRSCQLLRLLARFSLRGVQRIWNGELRFALQQLSEHHSYPALRGEAAQTLDEIS
                     HFTFFVT"
     intron          1015. .1086
                     /gene="fu"
                     /db_xref="FlyBase:FBgn0001079"
                     /number=1
     exon            1087. .2456
                     /gene="fu"
                     /db_xref="FlyBase:FBgn0001079"
                     /number=2
     intron          2457. .2516
                     /gene="fu"
                     /db_xref="FlyBase:FBgn0001079"
                     /number=2
     exon            2517. .3091
                     /gene="fu"
                     /db_xref="FlyBase:FBgn0001079"
                     /number=3
     intron          3092. .3159
                     /gene="fu"
                     /db_xref="FlyBase:FBgn0001079"
                     /number=3
     exon            3160. .3521
                     /gene="fu"
                     /db_xref="FlyBase:FBgn0001079"
                     /number=4
BASE COUNT     1190 a    1305 c    1286 g    1244 t
ORIGIN

  Query Match           13.3%;  Score 17;  DB 18;  Length 5025;
  Best Local Similarity 100.0%;  Pred. No. 3.26e+01;
  Matches    17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Db     367 TGTCAAAAGGGTCATTT 383
           |||||||||||||||||
Qy      26 TGTCAAAAGGGTCATTT 42


RESULT   36
LOCUS       DROFUSEDST    5025 bp     DNA              INV       21-JUL-1994
DEFINITION  Drosophila melanogaster serine/threonine kinase (fused) gene exons
            1-4, 5' end.
ACCESSION   L34782
NID         g514355
KEYWORDS    fused gene; serine/threonine kinase.
SOURCE      Drosophila melanogaster (strain Oregon R) DNA.
  ORGANISM  Drosophila melanogaster
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Arthropoda;
            Tracheata; Insecta; Pterygota; Diptera; Brachycera; Muscomorpha;
```



```
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 5025)
  AUTHORS   Tricoire,H. and Lamour-Isnard,C.
  TITLE     Drosophila melanogaster
  JOURNAL   Mech. Dev. 44, 65-80 (1993)
  MEDLINE   94206843
FEATURES             Location/Qualifiers
     source          1. .5025
                     /organism="Drosophila melanogaster"
                     /strain="Oregon R"
                     /db_xref="taxon:7227"
                     /map="x 17c-d x, 1-59,5"
     misc_feature    1. .5025
                     /note="fragment rescuing the fused mutant phenotype"
     mRNA            818. .4259
                     /gene="fused"
     gene            join(904. .1014,1087. .2456,2517. .3091,3160. .3518)
                     /note="fused"
                     /gene="fu"
                     /allele=""
                     /db_xref="FlyBase:FBgn0001079"
     CDS             join(904. .1014,1087. .2456,2517. .3091,3160. .3521)
                     /partial
                     /gene="fu"
                     /codon_start=1
                     /function="segment polarity gene"
                     /product="serine/threonine kinase"
                     /db_xref="PID:g514356"
                     /translation="MNRYAVSSLVGQGSLGCVYKATRKDDSKVVAIKVISKRGRATKE
                     LKNLRRECDIQARLKHPHVIEMIESFESKTDLFVVTEFALMDLHRYLSYNGAMGEEPA
                     RRVTGHLVSALYYLHSNRILHRDLKPQNVLLDKNMHAKLCDFGLARNMTLGTHVLTSI
                     KGTPLYMAPELLADEPYDHHADMWSLGCIAYESMAGQPPFCASSILHLVKMIKHEDVK
                     WPSTLTCECRSFLQGLLEKDPGLRISWTQLLCHPFVEGRIFIAETQAEAAKESPPTNP
                     EAKVKSSKQSDPEVGDLDEALAALDFGESRQENLTTSRDSINAIAPSDVDHLETDVED
                     NMQRVVVPFADLSYRDLSGVRAMPMVHQPVINSHTCPVSGHSNMILNHMNDNFDPQAS
                     LRGGVVAAKPIVAPTVRQSRSKDLEKRRLLSQNLDNFSVRLGHSVDHEAQRKATEIATQ
                     EKHNQENKPPAEAISYANSQPPQQQPQQLKHSMHSTNEEKLSSDNTPPCLLPGWDSCD
                     ESQSPPIENDEWLAFLNRSVQELLDGELDSLKQHNLVSIIVAPLRNSKAIPRVLKSVA
                     QLLSLPFVLVDPVLIVDLELIRNVYVDVKLVPNLMYACKLLLSHKQLSDSAASAPLTT
                     GSLSRTLRSIPELTVEELETACSLYELVCHLVHLQQQFLTQFCDAVAILAASDLFLNF
                     LTHDFRQSDSDAASVRLAGCMLALMSCVLRELPENAELVERIVFNPRLNFVSLLQSRH
                     HLLGQRSCQLLRLLARFSLRGVQRIWNGELRFALQQLSEHHSYPALRGEAAQTLDEIS
                     HFTFFVT"
     exon            <904. .1014
                     /gene="fu"
                     /db_xref="FlyBase:FBgn0001079"
                     /number=1
     intron          1015. .1086
                     /gene="fused"
                     /number=1
     exon            1087. .2456
```



```
                     /gene="fu"
                     /db_xref="FlyBase:FBgn0001079"
                     /number=2
     intron          2457. .2516
                     /gene="fused"
                     /number=2
     exon            2517. .3091
                     /gene="fu"
                     /db_xref="FlyBase:FBgn0001079"
                     /number=3
     intron          3092. .3159
                     /gene="fused"
                     /number=3
     exon            3160. .3518
                     /gene="fu"
                     /db_xref="FlyBase:FBgn0001079"
                     /number=4
BASE COUNT     1190 a    1305 c    1286 g    1244 t
ORIGIN

  Query Match           13.3%;  Score 17;  DB 18;  Length 5025;
  Best Local Similarity 100.0%;  Pred. No. 3.26e+01;
  Matches    17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Db     367 TGTCAAAAGGGTCATTT 383
           |||||||||||||||||
Qy      26 TGTCAAAAGGGTCATTT 42


RESULT   37
LOCUS       BSUEROP      10118 bp     DNA              BCT       19-MAY-1997
DEFINITION  B.subtilis urease operon and downstream DNA.
ACCESSION   Y08559
NID         g1592693
KEYWORDS    urease.
SOURCE      Bacillus subtilis.
  ORGANISM  Bacillus subtilis
            Eubacteria; Firmicutes; Low G+C gram-positive bacteria;
            Bacillaceae; Bacillus.
REFERENCE   1  (bases 1 to 10118)
  AUTHORS   Cruz-Ramos,H., Glaser,P., Wray,L.V. Jr. and Fisher,S.H.
  TITLE     The Bacillus subtilis ureABC operon
  JOURNAL   J. Bacteriol. 179 (10), 3371-3373 (1997)
  MEDLINE   97294484
REFERENCE   2  (bases 1 to 10118)
  AUTHORS   Glaser,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-SEP-1996) Philippe Glaser, Regulation de l'Expression
            Genetique, Institut, Pasteur, 28 Rue du Dr Roux, Paris, Paris,
            75724, France
FEATURES             Location/Qualifiers
```



```
     source          1. .10118
                     /organism="Bacillus subtilis"
                     /strain="168"
                     /db_xref="taxon:1423"
     gene            145. .476
                     /gene="ureA"
     RBS             145. .151
                     /gene="ureA"
     CDS             159. .476
                     /gene="ureA"
                     /codon_start=1
                     /transl_table=11
                     /product="Urease"
                     /db_xref="PID:e269484"
                     /db_xref="PID:g1592694"
                     /db_xref="SPTREMBL:P75030"
                     /translation="MKLTPVEQEKLLIFAAGELAKQRKARGVLLNYPEAAAYITCFIM
                     EGARDGKGVAELMEAGRHVLTEKDVMEGVPEMLDSIQVEATFPDGVKLVTVHQPISAE
                     VKS"
     gene            461. .847
                     /gene="ureB"
     RBS             461. .467
                     /gene="ureB"
     CDS             473. .847
                     /gene="ureB"
                     /codon_start=1
                     /transl_table=11
                     /product="urease"
                     /db_xref="PID:e269485"
                     /db_xref="PID:g1592695"
                     /db_xref="SPTREMBL:P71035"
                     /translation="MKPGAFQIAEGTITINEGREIREVTVKNTGSRSIQVGSHFHFAE
                     ANGALLFDRELAIGMRLDVPSGTSVRFEPGEQKTVSLVEIRGRKTIRGLNGMADTFID
                     ERGKEKTLANLKQAGWMEGVIR"
     RBS             829. .836
                     /gene="ureC"
     gene            829. .2553
                     /gene="ureC"
     CDS             844. .2553
                     /gene="ureC"
                     /codon_start=1
                     /transl_table=11
                     /product="urease"
                     /db_xref="PID:e269547"
                     /db_xref="PID:g1592696"
                     /db_xref="SPTREMBL:P77837"
                     /translation="MKMSREEYAELPGPTTGDKIRLGDTDLWIEVEKDFTVYGEEMIF
                     GGGKTIRDGMGQNGRITGKDGALDLVITNVVLLDYTGIVKADVGVKDGRIVGVGKSGN
                     PDIMDGVDPHMVIGAGTEVISGEGKILTAGGVDTHIHFICPQQMEVALSSGVTTLLGG
                     GTGPATGSKATTCTSGAWYMARMLEAAEEFPINVGFLGKGNASDKAPLIEQVEAGAIG
                     LKLHEDWGTTPSAIKTCMEVVDEADIQVAIHTDTINEAGFLENTLDAIGDRVIHTYHI
```

```
                     EGAGGGHAPDIMKLASYANILPSSTTPTIPTTVNTMDEHLDMMMVCHHLDAKVPEDVA
                     FSHSRIRAATIAAEDILHDIGAISMTSSDSQAMGRVGEVIIRTWQVADKMKKQRGALA
                     GENGNDNVRAKRYIAKYTINPAITHGLSHEVGSVEKGKLADLVLWDPVFFGVKPELVL
                     KGGMIARAQMGDPNASIPTPEPVPMRQMYASYGKANRSTSITFMSQASIERGVAESLG
                     LEKRISPVKNIRKLSKLDMKLNSALPKIEIDPKTYQVFADGEELSCQPVDYVPLGQRY
                     FLF"
     RBS             2674. .2680
                     /gene="ywnA"
     gene            2674. .3089
                     /gene="ywnA"
     CDS             2688. .3089
                     /gene="ywnA"
                     /codon_start=1
                     /transl_table=11
                     /product="Unknown"
                     /db_xref="PID:e269486"
                     /db_xref="PID:g1592697"
                     /db_xref="SPTREMBL:P71036"
                     /translation="MINSRLAVAIHILSLISMDEKTSSEIIADSVNTNPVVVRRMISL
                     LKRADILTSRAGVPGASLKKDPADISLLEVTRAVQKQEELFAVHENPNPKCPVGKKIQ
                     NALDETFESVQRAMENELASKSLKDVMNHLF"
     gene            3141. .3798
                     /gene="ywnB"
     RBS             3141. .3148
                     /gene="ywnB"
     CDS             3157. .3798
                     /gene="ywnB"
                     /codon_start=1
                     /transl_table=11
                     /product="Unknown"
                     /db_xref="PID:e269548"
                     /db_xref="PID:g1592698"
                     /db_xref="SPTREMBL:P71037"
                     /translation="MKIGIIGASGKAGNEILKEAKKRGHEVTAIVRNASKVQEQDVAI
                     LEKDVFELTAEDIKPFDAVVNAPGAAPGQEHLHVEAGRALISILKDAKHTRLLVVGGA
                     GSLFVDEARTTRLMDTPEFPKEYLPTASNQGENLKDLQQTDSISWTFLSPAAFFDPAG
                     KRTGSYQKGKDNVIVNAKGDSYISYADYAIAVLDELEHPEHKNERPTVVSEAE"
     terminator      3800. .3828
     gene            complement(3833. .4233)
                     /gene="ywnC"
     CDS             complement(3833. .4216)
                     /gene="ywnC"
                     /codon_start=1
                     /transl_table=11
                     /product="Unknown"
                     /db_xref="PID:e269487"
                     /db_xref="PID:g1592699"
                     /db_xref="SPTREMBL:P71038"
                     /translation="MNPETMNKTLISISKWGKATGILPIIMGAITALSGAFFFLIGAV
                     PGVLQIISGIFLMRSAREAGQMAEHNSGQSEDLMLENYAKFVKMQGIYLIVSIAVSIL
                     AIIAFFIFLMLGIADGLFSDTYSTY"
```



```
     RBS             complement(4226. .4233)
                     /gene="ywnC"
     RBS             4347. .4354
                     /gene="ywnD"
     gene            4347. .5134
                     /gene="ywnD"
     CDS             4361. .5134
                     /gene="ywnD"
                     /note="Product similar to Streptomyces lividans
                     transcriptional activator TipA"
                     /codon_start=1
                     /transl_table=11
                     /product="Unknown"
                     /db_xref="PID:e269488"
                     /db_xref="PID:g1592700"
                     /db_xref="SPTREMBL:P71039"
                     /translation="MKYQVKQVAEISGVSIRTLHHYDNIELLNPSALTDAGYRLYSDA
                     DLERLQQILFFKEIGFRLDEIKEMLDHPNFDRKAALQSQKEILMKKKQRMDEMIQTID
                     RTLLSVDGGETMNKRDLFAGLSMKDIEEHQQTYADEVRKLYGKEIAEETEKRTSAYSA
                     DDWRTIMAEFDSIYRRIAARMKHGPDDAEIQAAVGAFRDHICQYHYDCTLDIFRGLGE
                     VYITDERFTDSINQYGEGLAAFLREAIIIYCDHQENPRP"
     terminator      5118. .5148
     gene            complement(5155. .6614)
                     /gene="ywnE"
     CDS             complement(5155. .6603)
                     /gene="ywnE"
                     /note="Product similar to Escherichia coli cardiolipin
                     synthetase"
                     /codon_start=1
                     /transl_table=11
                     /product="Unknown"
                     /db_xref="PID:e269549"
                     /db_xref="PID:g1592701"
                     /db_xref="SPTREMBL:P71040"
                     /translation="MSISILLSLFFILNILLAIIVIFKERRDASASWAWLLVLFFIP
                     VLGFILYLLFGHNLRRKHLFQWEDRKKIGIERLLKHQLEDLETKQFQFNNRATFDNKD
                     LIYMLIMNNHAVFTEDNSVDVITDGRDKPQRLLSDISKAKDHIHLQYYIYKGDELGKK
                     LRDALIQKAKEGVQVRVLYDELGSRTLRKKFPKELREAGGHVEVFPPSKLRPINLRLN
                     YRNHRKLVIIDGMTGYVGGFNVGDEYLGLNPKFGYWRDTHIRLQGTAVHAIQTRFILD
                     WNQASHHHTLTYIPNHPPDYGPKGNVGMQIVTSGPDSEWEQIKNGYIKMISNARKSIL
                     IQTPYFIPDASLLDALRIACLSGIDVNIMIPNKPDHAFVYWATLSYIGDLLKAGATVY
                     IYDNGFIHAKTIVVDDEIASVGTANIDVRSFRLNFEVNAFIYDITIAKKLVSTFKEDL
                     LVSRKFTYEEYLQRPLWIRIKESVSRLLSPIL"
     RBS             complement(6612. .6614)
                     /gene="ywnE"
     RBS             6831. .6837
                     /gene="ywnF"
     gene            6831. .7280
                     /gene="ywnF"
     CDS             6846. .7280
                     /gene="ywnF"
```



```
                     /codon_start=1
                     /transl_table=11
                     /product="Unknown"
                     /db_xref="PID:e269489"
                     /db_xref="PID:g1592702"
                     /db_xref="SPTREMBL:P71041"
                     /translation="MNFYRVEQMPGFIKTEMQKIQKAVQPPMKKTVIYRPLAIPLAAF
                     SLFNLAAFLFHASADRESLISAGIFALLAALGLAFFKEAGYQHKQIQKTVHIYMLNRI
                     KKSEILSEERKSSYARQIKEEPFAMRSFVEFLTEEDRRKKMY"
     terminator      7291. .7336
     gene            complement(7567. .8100)
                     /gene="ywnG"
     CDS             complement(7567. .8085)
                     /gene="ywnG"
                     /codon_start=1
                     /transl_table=11
                     /product="Unknown"
                     /db_xref="PID:e269490"
                     /db_xref="PID:g1592703"
                     /db_xref="SPTREMBL:P71042"
                     /translation="MISLDKDENEIEHHNEENSLVEQETAPVGQESRQLSASAVKSLS
                     DIAKWGKISGILLIIMGSLVTLSVLMTVIGAIPGVLLIISGVFLMRSAKAAAEAEGNL
                     TGSAGESMLENYGTFIKMQLFYAASSIVTVLIGIIVAIFVLVVIGIAAFENTPSYDDP
                     DSYTYEDDPVFE"
     gene            complement(8082. .8598)
                     /gene="ywnH"
...
Note: remainder of annotations omitted.

  Query Match           13.3%;  Score 17;  DB 15;  Length 10118;
  Best Local Similarity 100.0%;  Pred. No. 3.26e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    6701 TTTTAAAAGAATAAAGA 6717
           |||||||||||||||||
Qy      95 TTTTAAAAGAATAAAGA 111


RESULT   38
LOCUS       CEF29D10    23952 bp    DNA             INV       21-SEP-1998
DEFINITION  Caenorhabditis elegans cosmid F29D10, complete sequence.
ACCESSION   Z75952
NID         g1429273
KEYWORDS    HTG; C2H2; Myosin; Zinc finger.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 23952)
  AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
```



```
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
  TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
  JOURNAL   Nature 368 (6466), 32-38 (1994)
  MEDLINE   94150718
REFERENCE   2  (bases 1 to 23952)
  AUTHORS   White,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-NOV-1996) Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     Coding sequences below are predicted from computer analysis, using
            predictions from Genefinder (P. Green, U. Washington), and other
            available information.
            For a graphical representation of this sequence and its analysis
            see:-
            http://webace.sanger.ac.uk/cgi-
            bin/display?db=wormace&class=Sequence &object=F29D10
            Current sequence finishing criteria for the C. elegans genome
            sequencing consortium are that all bases are either sequenced
            unambiguously on both strands, or on a single strand with both a
            dye primer and dye terminator reaction, from distinct subclones.
            Exceptions are indicated by an explicit note.
            IMPORTANT:  This sequence is NOT necessarily the entire insert of
            the specified clone.  It may be shorter because we only sequence
            overlapping sections once, or longer because we arrange for a small
            overlap between neighbouring submissions.
            IMPORTANT: This sequence is not the entire insert of clone F29D10.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we arrange for a small overlap between
            neighbouring submissions.
            The true left end of clone F29D10 is at 1 in this sequence. The
            true left end of clone F55H12 is at 23844 in this sequence. The
            true right end of clone F36A2 is at 5793 in this sequence. The
            start of this sequence (1. .114) overlaps with the end of sequence
            Z81077.
            The end of this sequence (23844. .23952) overlaps with the start of
            sequence Z81091.
FEATURES                Location/Qualifiers
     source             1. .23952
                        /organism="Caenorhabditis elegans"
                        /db_xref="taxon:6239"
                        /chromosome="I"
                        /clone="F29D10"
```

```
gene            6362. .6856
                /gene="F29D10.1"
CDS             join(6362. .6640,6689. .6856)
                /gene="F29D10.1"
                /codon_start=1
                /db_xref="PID:e252945"
                /db_xref="PID:g1429274"
                /db_xref="SPTREMBL:Q19898"
                /translation="MAQMIEHFEIFVLCKITFTLFRITFAIIVLLSFSFSLLVICVKK
                KTKKEKHPDSPPTLVTPPPQGAKLSWKGRQKPLVVFDEKLSDQSKKEKSVKIPKTVEE
                SEKIVKVKKVPSSQKVCDSSASSKKRKSPKKKSEDVDELKTETERE"
gene            7686. .8080
                /gene="F29D10.2"
CDS             join(7686. .7859,7904. .8080)
                /gene="F29D10.2"
                /note="cDNA EST yk309a8.3 comes from this gene; cDNA EST
                yk309a8.5 comes from this gene"
                /codon_start=1
                /db_xref="PID:e253176"
                /db_xref="PID:g1429275"
                /db_xref="SPTREMBL:Q19899"
                /translation="MEAVVDVVDVEEVVEDVDVVVVDQDAVAVVDLNVVVLVVELVAV
                LDVVLAADHAVVDVVADVVVDAVDVEVEVVSVVPFKNFELMKPIVHLGLKKDHLMTNV
                SPIIYSPFIIYHDS"
gene            complement(8625. .9710)
                /gene="F29D10.3"
CDS             complement(join(8625. .8845,8901. .9240,9298. .9396,
                9666. .9710))
                /gene="F29D10.3"
                /note="Weak similarity with botulinum neurotoxin type A
                (TREMBL ID GS07071)"
                /codon_start=1
                /db_xref="PID:e252946"
                /db_xref="PID:g1429276"
                /db_xref="SPTREMBL:Q19900"
                /translation="MPKHRPFERVRQIAEVVQKVNDRCRLFSFGSLTINRLDSSSQSI
                LAFKRNETNDTCPAKSFQLTFESSSSEWIDEKGFIYNLTFTSFNNQWIFDYTVTRCPD
                NSALWTRNTSKGIFHVCVETKLFPNNTNNVYEEAHNLCMEPNGMGLTAPPDSTELKSF
                IATKKDLFAAGSNNGALDSTMWIDGRSYESKNYIYDDISHVGTASYLFGGNEPDGVGP
                NYCLFLEYNNGIGDKL"
gene            complement(11750. .16023)
                /gene="hum-4"
CDS             complement(join(11750. .11823,11874. .12186,12237. .12367,
                13078. .13289,13345. .13482,13531. .13637,13687. .13863,
                13915. .14022,14071. .14421,14468. .14549,14613. .15014,
                15128. .15248,15296. .15405,15451. .15532,15579. .15711,
                15801. .16013,16063. .16206,16254. .16383,16438. .16580,
                16692. .16823))
                /gene="hum-4"
                /note="similar to myosin; cDNA EST EMBL:D27611 comes from
                this gene; cDNA EST EMBL:D32618 comes from this gene; cDNA
```



```
                EST EMBL:D34964 comes from this gene; cDNA EST EMBL:D35303
                comes from this gene; cDNA EST EMBL:D36573 comes from this
                gene; cDNA EST EMBL:D64509 comes from this gene; cDNA EST
                EMBL:D67523 comes from this gene; cDNA EST EMBL:C12624
                comes from this gene; cDNA EST EMBL:C12810 comes from this
                gene; cDNA EST EMBL:C12944 comes from this gene; cDNA EST
                EMBL:C13741 comes from this gene; cDNA EST EMBL:C10784
                comes from this gene; cDNA EST EMBL:C10903 comes from this
                gene; cDNA EST EMBL:C10950 comes from this gene; cDNA EST
                EMBL:C11512 comes from this gene; cDNA EST yk219h3.3 comes
                from this gene; cDNA EST yk219h3.5 comes from this gene;
                cDNA EST yk212c10.3 comes from this gene; cDNA EST
                yk212c10.5 comes from this gene; cDNA EST yk205a5.3 comes
                from this gene; cDNA EST yk205a5.5 comes from this gene;
                cDNA EST yk194g5.3 comes from this gene; cDNA EST
                yk194g5.5 comes from this gene; cDNA EST yk192h5.3 comes
                from this gene; cDNA EST yk192h5.5 comes from this gene"
                /codon_start=1
                /db_xref="PID:e252947"
                /db_xref="PID:g1429277"
                /db_xref="SPTREMBL:Q19901"
                /translation="MAFHWQSKVNVQHVGVDDMVLLPKLTEQSIVENLKKRLQANSIF
                TYIGPVLISVNPFKQMPYFTEKEMLLYQGAAQYENAPHIYALADNMYRNMLIDNESQC
                VIISGESGAGKTVNAKFIMNYISRISGGGQKVQHIKDVILQSNPLLEAFGNSATVRNW
                NSSRFGKYVEIVFSRGGEPIGGKLSNFLLEKSRVVHQNEGDRNFHVFYQLCAGADKNL
                RSTFGIGELQYYNYLNMSGVFKADDTDDGKEFESTLHAMKVVGVNDQDQLEVLRIVAT
                VLHIGNITFTEENNFAAVSGKDYLEYFAFLLGLTSADIEAKLTGRKMESKWGTQKEEI
                DMKLNVEQASYTRDAMVKAIYARLFDYLVKKVNDAMNITSQSTSDNFSVGILDIYGFE
                IFNNNGFEQFCINFVNEKLQQIFIELTLKAEQEEYVREGIKWTEIDYFDNKIVCDLIE
                TKRPPGIMSLLDDTCAQNHGQREGVDRQLLTTLSKSFAGHPHPGPGSDSFVIKHYAGD
                VTYNVDGFCDRNRDVLYPDLILLLMQKSSRPFIQALFPENVAASAGKRPTTFSTKIRTQ
                ANTLVESLMKCSPHYVRCIKPNETKRPNDWEESRVKHQVEYLGLRENIRVRRAGFAYR
                RAFDKFAQRYAIVSPQTWPCFQGDQQRACEIICDSVHMEKNQYQMGKTKIFVKNPESL
                FLLEETRERKFDGYARVIQRAWRQFSARKQHIKQKEQAAADLMTGKKERRRYSLNRNFV
                GDYIGLEHHPTLQSLVGKRQRVLFACTANNYDRKFRVTKLDLLLLTVNHLTLIGKEKVK
                NGPEKGKIVEVIKRQPDLPCIKSIGLSPYQDDFVILYLGNDDYSSLLETPFKTEFCTA
                LSKAYEERTNGTLHLDFRSSHVVSYKKMKFDFSDGKRTVQFGNDGTSSAEKTLKPNGK
                VLNVSIGTGLPNTTRPSTERPQGGGTTPRRDQLRTSTRRTKQNNQSTGQNGQSQAMRAP
                VPAHGMNNNYNQTPAPVSTNHQYSQEPARIPVMGNVINQLNNMNLSGMGNSPAGRGPP
                PARGPKPPPPAKPKLNPVVIAVYPYEAQDVDELSFEAGAEIELMNKDASGWWQGKVNN
                RVGLFPGNYVKE"
gene            complement(19561. .21173)
                /gene="F29D10.5"
CDS             complement(join(19561. .19802,19877. .20029,20235. .20372,
                20846. .21173))
                /gene="F29D10.5"
                /note="similar to Zinc finger, C2H2 type (2 domains)"
                /codon_start=1
                /db_xref="PID:e253177"
                /db_xref="PID:g1429278"
                /db_xref="SPTREMBL:Q19902"
```



```
                /translation="MSHINLYSLPTDHKNMIRCRYCEKIYKIDNGSSGFIRHLRCKHP
                NVLYQMGAISNGKEYNLITKQETLSIDTNNEINLRDDDDDFMMMKEFSSNQILVNFLSQ
                LDLPTNDVNRVLGKRHRISPTPTIHSSVTPNEEHEPDLKMFNSGDTLQNFLIDIDNHI
                RQKNQESQEAIGTGEESTSSVSALDMPSSQKESREPSPNSSMLDETNGGTETGDSSPTP
                FWPARPFGCVICPCSFPRVKDLTIHVTSVHRTRPRCNTCLAEFVQPQHLEHHYSISHA
                GTSTGTVVNT"
gene            join(complement(4191. .4321),complement(3782. .4132),
                complement(3470. .3735),complement(3299. .3424),
                complement(3041. .3245),complement(2572. .2994),
                complement(2284. .2512),complement(1805. .2233),
                complement(1406. .1756),complement(1198. .1356),
                complement(996. .1155),complement(385. .948),
                complement(115. .340),complement(Z81077:33191. .33428),
                complement(Z81077:33109. .33145),
                complement(Z81077:32779. .32926),
                complement(Z81077:32039. .32735),
                complement(Z81077:31316. .31988),
                complement(Z81077:31149. .31270),
                complement(Z81077:30651. .31079),
                complement(Z81077:30295. .30603),
                complement(Z81077:29928. .30242),
                complement(Z81077:29099. .29882),
                complement(Z81077:28799. .29042),
                complement(Z81077:28181. .28745),
                complement(Z81077:27908. .28129),
                complement(Z81077:27729. .27857),
                complement(Z81077:27460. .27675))
                /gene="F36A2.13"
CDS             join(complement(4191. .4321),complement(3782. .4132),
                complement(3470. .3735),complement(3299. .3424),
                complement(3041. .3245),complement(2572. .2994),
                complement(2284. .2512),complement(1805. .2233),
                complement(1406. .1756),complement(1198. .1356),
                complement(996. .1155),complement(385. .948),
                complement(115. .340),complement(Z81077:33191. .33428),
...
Note: remainder of annotations omitted.

  Query Match          13.3%;  Score 17;  DB 18;  Length 23952;
  Best Local Similarity 100.0%;  Pred. No. 3.26e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   6411 CTTTTACACTTTTCAGA 6427
         |||||||||||||||||
Qy     55 CTTTTACACTTTTCAGA 71


RESULT   39
LOCUS       CEC54C8     23960 bp    DNA             INV       21-SEP-1998
DEFINITION  Caenorhabditis elegans cosmid C54C8, complete sequence.
ACCESSION   Z83102
```



```
NID             g1752853
KEYWORDS        HTG; C-type lectin domain; Steroid hormone receptor.
SOURCE          Caenorhabditis elegans.
  ORGANISM      Caenorhabditis elegans
                Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
                Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE       1  (bases 1 to 23960)
  AUTHORS       Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
                Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
                Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
                Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
                Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
                Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
                O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
                Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
                Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
                Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
                Wilkinson-Sproat,J. and Wohldman,P.
  TITLE         2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
                elegans
  JOURNAL       Nature 368 (6466), 32-38 (1994)
  MEDLINE       94150718
REFERENCE       2  (bases 1 to 23960)
  AUTHORS       Dobson,R.
  TITLE         Direct Submission
  JOURNAL       Submitted (06-NOV-1996) Louis, MO 63110, USA. E-mail:
                jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT         On Dec 25, 1996 this sequence version replaced gi:1695005.
                Coding sequences below are predicted from computer analysis, using
                predictions from Genefinder (P. Green, U. Washington), and other
                available information.
                For a graphical representation of this sequence and its analysis
                see:-
                http://webace.sanger.ac.uk/cgi-
                bin/display?db=wormace&class=Sequence &object=C54C8
                Current sequence finishing criteria for the C. elegans genome
                sequencing consortium are that all bases are either sequenced
                unambiguously on both strands, or on a single strand with both a
                dye primer and dye terminator reaction, from distinct subclones.
                Exceptions are indicated by an explicit note.
                IMPORTANT:  This sequence is NOT necessarily the entire insert of
                the specified clone.  It may be shorter because we only sequence
                overlapping sections once, or longer because we arrange for a small
                overlap between neighbouring submissions.
                IMPORTANT: This sequence is not the entire insert of clone C54C8.
                It may be shorter because we only sequence overlapping sections
                once, or longer because we arrange for a small overlap between
                neighbouring submissions.
                The true left end of clone C54C8 is at 21084 in
                sequence Z81492.
                The true right end of clone C54C8 is at 23957 in this sequence. The
```

true right end of clone E03H4 is at 106 in this sequence. The start of this sequence (1. .106) overlaps with the end of sequence Z81492.

```
FEATURES             Location/Qualifiers
     source          1. .23960
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /chromosome="I"
                     /clone="C54C8"
     gene            1306. .2735
                     /gene="C54C8.1"
     CDS             join(1306. .1345,1392. .1610,1656. .1742,1803. .1888,
                     2018. .2241,2297. .2498,2545. .2619,2682. .2735)
                     /gene="C54C8.1"
                     /note="predicted using Genefinder; Similarity to zinc
                     finger proteins"
                     /codon_start=1
                     /db_xref="PID:e348350"
                     /db_xref="PID:g1752857"
                     /db_xref="SPTREMBL:O17706"
                     /translation="MKLSTSLEMPFRESTDPICSVCNFSSLIAPHFGGLVCSACASFF
                     RRTVALNIHYLCKKDNQCKGMRKNCRACRFESCVKIAGMKRSLVKERKNLNNIPLYIL
                     NRRNETGNQGVVRAFVSSNQNRLENRLSLSPLTELPINDEMDNTLNIRTKEELPEIVS
                     YQSKVAAETCRTCPGVDLLDNRDIILILRKIFQPFSNIWIESTWKTWFSNDFSNDTEKFD
                     IKLMEPIGQVKSTLLISLSRLKFNIFEPTAFKAICIWKLVYHETSRAMKIVAQEHYEG
                     VTKALISYYQTTTLLDSMEIAIRVGEITLLVSSVFQMYHDMAKLYLRMGLPF"
     gene            3784. .5475
                     /gene="C54C8.2"
     CDS             join(3784. .3836,4022. .4214,4264. .4568,4623. .4694,
                     4745. .4861,5057. .5150,5218. .5354,5397. .5475)
                     /gene="C54C8.2"
                     /codon_start=1
                     /db_xref="PID:e348999"
                     /db_xref="PID:g1877071"
                     /db_xref="SPTREMBL:O02230"
                     /translation="MHDEQFCVGYNFLEAEDTFREDGLEPVTLAIHGTPEVLQLLGNK
                     PLNWDGPISPGLFVDFHSQKALNYISMLHKCDAAFKRQMTVHFAPRISPSQSECPMIQ
                     VLGYQDCATFLQKSKQLLEEIEDSFQIYPINLMRNIARRGAKSDLHLIIDTDMMMSTN
                     FAKMVKPIANRMIDGKNKQVLVVRRFETNENELPMSFGDLEEGIENHKTFQFHHKFFF
                     VGHQIPNLMEWFERSHASNDVVAWEIPYTGNDWEVQIILHRNDPTNVEYFPSRVKDMQ
                     SLIYKLCRANYTFNLLSHVFNVHKGIKEDDTMYSKVVTAHTKRQGRLRTLSRYVTEID
                     RKYPDTMKRCGQFLL"
     gene            complement(6289. .7268)
                     /gene="C54C8.3"
     CDS             complement(join(6289. .6399,6453. .6579,6626. .6825,
                     6892. .7186,7213. .7268))
                     /gene="C54C8.3"
                     /codon_start=1
                     /db_xref="PID:e348352"
                     /db_xref="PID:g1877072"
                     /db_xref="SPTREMBL:O02231"
                     /translation="MGGW1TQHHLKLKILARNLNYSWLYLPEIEETSQEKDILIIVAS
```



```
                     RTDSYARRNILRQTWMSKANSEIVANGRMKPLFLVGLTPGEYKMKKMVMQEAKLYGDI
                     IVVDMNDNYEELTYKSLAILLLYGVSKAPRYQMIGKIDEDVMFFPDKLTELYDQGFIDA
                     TPLRIYGLKMQSGANIFRDKTHRWTVPESSYSCSKFPEYVSGMLYMVTWEAAQQIIKS
                     TKYRDFIQVEDVFLTGILAEDLGISVKNLPEFYKYPSDVGFFSF"
     gene            9017. .11004
                     /gene="C54C8.4"
     CDS             join(9017. .9275,9336. .9421,9475. .9694,9981. .10045,
                     10103. .10270,10315. .10508,10797. .11004)
                     /gene="C54C8.4"
                     /note="similar to glycosyltransferase; cDNA EST yk314f3.3
                     comes from this gene; cDNA EST yk314f3.5 comes from this
                     gene"
                     /codon_start=1
                     /db_xref="PID:e348353"
                     /db_xref="PID:g1877073"
                     /db_xref="SPTREMBL:O02232"
                     /translation="MSNPHCKLTIKLLLSCFVIYSTFSILNYLLFTGSNIKNLLNERR
                     PQKYQPYPIIEAGGGLNISSENIAIVIILSKSHDRSYYQIAIDSVECYAKAHGYEPVL
                     TNDSNFGCDHIKINIFRRHCVLAKILVDYEAVVHIDGDMGIVNPKRRIQEYMTESSDI
                     VFYDRHFAPEVAMCYIVKNTEYAITLLNEFAEYERKLPNSSHGSENGALHIFLAEKLF
                     PHNQIDIDVCRIAWENSWNIVDQFAYTHCIRAIPGASIDFDKIRILRKGTGFARDDWI
                     TGGVWSPERDFMLHGWKMRQLAPTPAGVVKAGAMSELIWYSPFLGSFDMEKCSMGNTT
                     WYYNPSLIASRETVEVVLQQEDVISAARKVNYFYRLYKVITNSFWWYIYQSHVWKSDY
                     PLYFLNE"
     gene            12819. .13771
                     /gene="C54C8.5"
     CDS             join(12819. .12916,12963. .13285,13365. .13621,13670. .13771)
                     /gene="C54C8.5"
                     /codon_start=1
                     /db_xref="PID:e349000"
                     /db_xref="PID:g1877074"
                     /db_xref="SPTREMBL:O17708"
                     /translation="MEPSSKNSDASNRTIIVVTPTYKRITRLPDMIRLAMTLAHVKNL
                     HWIVVEDGYGIVPAVRELLERTNLSTTYMAHKTARGYPTRGWYQRTMALRTIRSSWSK
                     ILREHDAVVYFADDDNAYDVRLFTDYIRNVNTLGVWAVGLVGGAIVEAPKVVNHKVTG
                     FNVRWSPNRRFAVDMAGFAINLKYILNSDAVFGKDCKRGEGAPETCLIEDIGLDLEDI
                     KPFGYDATKVQDIMVWHTKTSPPEIEQRDEPIDSLGYFVEY"
     gene            complement(13860. .14736)
                     /gene="C54C8.6"
     CDS             complement(join(13860. .14117,14455. .14736))
                     /gene="C54C8.6"
                     /codon_start=1
                     /db_xref="PID:e348354"
                     /db_xref="PID:g1877075"
                     /db_xref="SPTREMBL:O02233"
                     /translation="MRATVSGKTTTILPLLSKLIVIVIKSTSIHCSTLTSSGLFDSTA
                     PRVFKNLVNYGGDMWFYMKPYPSNGKSGSQQMENVYQGLKSAQIKFSHITQHGISVQF
                     FTSQSNWHSITNNNRSFSSYKLWWAHTNGAGNGGATSPNFEGFVSFGGWTSSSLVGKE
                     YARQVQECGVTVNKYVYVK"
     gene            complement(18506. .19910)
                     /gene="C54C8.7"
```



```
     CDS             complement(join(18506. .19281,19326. .19457,19503. .19644,
                     19707. .19910))
                     /gene="C54C8.7"
                     /note="predicted using Genefinder; similar to C-type
                     lectin domain"
                     /codon_start=1
                     /db_xref="PID:e348998"
                     /db_xref="PID:g1752854"
                     /db_xref="SPTREMBL:O17705"
                     /translation="MRTVAALLFLFTCLESFESSSATPVCTNGFTLINNKCLKLLNTP
                     VNHKAAEISCSSFGATLVTVKNDHDNQAIATIVGSSTPLVWLGLYCFSSNSSQCLWDD
                     ESGSARSYNNFSSGFPLVALGQCVTYSTQGALTGKWLSADCESQRHAPVCELPTTFAD
                     NCLNNYNGYCTTFSSSPQTFIGAQSTCAQTCGNLASVHSPHENRYITTPAPQDYTYIG
                     AIWKTDYSLRWMDGSAWDYNNIDPIYPNRFNYCLHMSTITTSGSGFWYGDDCSLSREY
                     VCKREAGVPCTTVPPPVTVRPSSTNPSNCNAGLLMSPGVPTSPNYPQNYFNNENCTYQ
                     LSTLGSYRITLKFSNFRTESNYDFVSVTDGPTTSSPSLGRYSGNIGSFYVSSSENNML
                     VTFTSDAGVVFQGFTARFYSVVHRS"
     gene            20663. .20902
                     /gene="C54C8.9"
     CDS             20663. .20902
                     /gene="C54C8.9"
                     /note="predicted using Genefinder"
                     /codon_start=1
                     /db_xref="PID:e348351"
                     /db_xref="PID:g1752860"
                     /db_xref="SPTREMBL:O17707"
                     /translation="MKLLILFSLFAIFFGVIALDSPIPEFYSTGRSTRVPSHRHRIPR
                     LGKREVPNFQADNVPEAGGRVRRYVGPPLKLVILI"
     gene            join(complement(245. .260),complement(50. .193),
                     complement(Z81492:39726. .39791),
                     complement(Z81492:39516. .39679),
                     complement(Z81492:39024. .39443),
                     complement(Z81492:38897. .38965),
                     complement(Z81492:38799. .38852))
                     /gene="E03H4.13"
     CDS             join(complement(245. .260),complement(50. .193),
                     complement(Z81492:39726. .39791),
                     complement(Z81492:39516. .39679),
                     complement(Z81492:39024. .39443),
                     complement(Z81492:38897. .38965),
                     complement(Z81492:38799. .38852))
                     /gene="E03H4.13"
                     /note="Similarity to zinc finger proteins"
                     /codon_start=1
...
Note: remainder of annotations omitted.

  Query Match          13.3%;  Score 17;  DB 18;  Length 23960;
  Best Local Similarity 100.0%;  Pred. No. 3.26e+01;
  Matches    17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```



```
Db  22541 CTTTGAGAATCTGAAAA 22557
          |||||||||||||||||
Cp     80 CTTTGAGAATCTGAAAA 64


RESULT   40
LOCUS       CEF14F11    25350 bp    DNA              INV       21-SEP-1998
DEFINITION  Caenorhabditis elegans cosmid F14F11, complete sequence.
ACCESSION   Z54307
NID         g1008949
KEYWORDS    HTG; Potassium channel.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 25350)
  AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
  TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
  JOURNAL   Nature 368 (6466), 32-38 (1994)
  MEDLINE   94150718
REFERENCE   2  (bases 1 to 25350)
  AUTHORS   Lightning,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-NOV-1996) Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     Coding sequences below are predicted from computer analysis, using
            predictions from Genefinder (P. Green, U. Washington), and other
            available information.
            For a graphical representation of this sequence and its analysis
            see:-
            http://webace.sanger.ac.uk/cgi-
            bin/display?db=wormace&class=Sequence &object=F14F11
            Current sequence finishing criteria for the C. elegans genome
            sequencing consortium are that all bases are either sequenced
            unambiguously on both strands, or on a single strand with both a
            dye primer and dye terminator reaction, from distinct subclones.
            Exceptions are indicated by an explicit note.
            IMPORTANT:  This sequence is NOT necessarily the entire insert of
            the specified clone.  It may be shorter because we only sequence
            overlapping sections once, or longer because we arrange for a small
```

```
            overlap between neighbouring submissions.
            IMPORTANT: This sequence is not the entire insert of clone F14F11.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we arrange for a small overlap between
            neighbouring submissions.
            The true left end of clone F14F11 is at 1 in this sequence. The
            true right end of clone F14F11 is at 12071 in
            sequence Z54327.
            The true left end of clone C26D10 is at 25243 in this sequence. The
            start of this sequence (1. .100) overlaps with the end of sequence
            AL023831.
            The end of this sequence (25243. .25350) overlaps with the start of
            sequence Z54327.
FEATURES    Location/Qualifiers
     source 1. .25350
            /organism="Caenorhabditis elegans"
            /db_xref="taxon:6239"
            /chromosome="II"
            /clone="F14F11"
     gene   complement(1146. .1435)
            /gene="F14F11.2"
     CDS    complement(join(1146. .1320,1377. .1435))
            /gene="F14F11.2"
            /codon_start=1
            /db_xref="PID:g1008951"
            /db_xref="SPTREMBL:Q19465"
            /translation="MSCISMLFILLVACLLVSNAMYINPDYYYVEQLPTMKKSGQLRA
            LAGSRNCFFSPVNCIITHDINSYRRRLAKGSSYA"
     gene   complement(11966. .16681)
            /gene="F14F11.1"
     CDS    complement(join(11966. .12114,12258. .12400,12613. .12740,
            12846. .13260,13353. .13528,14437. .14641,14740. .14879,
            14924. .15027,15088. .15214,15262. .15404,15450. .15543,
            15809. .15928,16016. .16123,16258. .16337,16483. .16681))
            /gene="F14F11.1"
            /note="similar to potassium channel"
            /codon_start=1
            /db_xref="PID:g1008950"
            /db_xref="SPTREMBL:Q19464"
            /translation="MDTEHRVILNVGGIRHETYSHVLKKIPATRLSRLTPNLANYDPV
            LNEYFFDRHPGVFSMILNYYRTGKLHYPTNVCGPLFEEELEPWGLDANQVEPCCWMTY
            TQHRDTQDTLAVIESLDLDGDPPTQEEIAKKFGWEDDYTYTGNMSQWQRLKPRVWALFD
            EPWSSKYARVISSLSVAFILASTCSPILKTDPSPQIPDIDVPYSLRVVDEGGFKNYHK
            TIGTDKPVTSPHPNFFYVDLICNIWFTIELLIRSLPCPSPHKFVRSPLTIIDVISTGA
            FFPESLLHAILIQTGSLVTLDFLSMICVLRLFKLTQHFSGLKILIQTFKASAQELFLL
            VFPVILAIVIFAALVYYAERSQLNKDNQFTSIPLGLWWSLVTISTVGFGDMVPKTYLG
            MLVGSLCALMGVLTIALPVPVIVSNFSNLYSHSQARAKLPKKRRRVLQVISVVNIIFV
            IISIVCFCLETHPSFRIPDIDISTTMGNFSAPLNGVQMNNHIVVGATATRAHPSFTVE
            LLSNIWFSGEFLTRMVFCPNVATFLRTPVNLIDFIATVSFTIDWALDRTLSGSNRDSV
            EFFSIIRILRLFKLTQHSTGLKILIQTFKASAQELFLLVFPVLLGIVIFAALVYYAER
            VEHNEENQFSSIPVPQTSLGRLVGSVCALMGVLTIALPVPVIVSNFAMFYSHAQARSK
```



```
                        LPKKRRRVLQPHEIKPVVGRSTTAVMISALTQRGPVSGGGSNPSDGGPLSAFSNNTPL
                        LVCPPDGSTKPKKLSQNKNGGTTPIQNNGSKASKLI"
BASE COUNT     9128 a    3934 c    4460 g    7828 t
ORIGIN

  Query Match           13.3%;  Score 17;  DB 18;  Length 25350;
  Best Local Similarity 100.0%;  Pred. No. 3.26e+01;
  Matches    17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Db  23935 TTTTAAAAGAATAAAGA 23951
          |||||||||||||||||
Qy     95 TTTTAAAAGAATAAAGA 111


RESULT   41
LOCUS       CEF22B8     27102 bp    DNA             INV       21-SEP-1998
DEFINITION  Caenorhabditis elegans cosmid F22B8, complete sequence.
ACCESSION   Z83106
NID         g1752899
KEYWORDS    HTG.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 27102)
  AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
  TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
  JOURNAL   Nature 368 (6466), 32-38 (1994)
  MEDLINE   94150718
REFERENCE   2  (bases 1 to 27102)
  AUTHORS   McMurray,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-NOV-1996) Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     On Dec 25, 1996 this sequence version replaced gi:1695022.
            Coding sequences below are predicted from computer analysis, using
            predictions from Genefinder (P. Green, U. Washington), and other
            available information.
            For a graphical representation of this sequence and its analysis
            see:-
```



```
            http://webace.sanger.ac.uk/cgi-
            bin/display?db=wormace&class=Sequence &object=F22B8
            Current sequence finishing criteria for the C. elegans genome
            sequencing consortium are that all bases are either sequenced
            unambiguously on both strands, or on a single strand with both a
            dye primer and dye terminator reaction, from distinct subclones.
            Exceptions are indicated by an explicit note.
            IMPORTANT:  This sequence is NOT necessarily the entire insert of
            the specified clone.  It may be shorter because we only sequence
            overlapping sections once, or longer because we arrange for a small
            overlap between neighbouring submissions.
            IMPORTANT: This sequence is not the entire insert of clone F22B8.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we arrange for a small overlap between
            neighbouring submissions.
            The true left end of clone F22B8 is at 1 in this sequence. The true
            right end of clone F22B8 is at 9256 in
            sequence Z83109.
            The true left end of clone F44G3 is at 27005 in this sequence. The
            true right end of clone F14H3 is at 12320 in this sequence. The
            start of this sequence (1. .104) overlaps with the end of sequence
            Z83105.
            The end of this sequence (27005. .27102) overlaps with the start of
            sequence Z83109.
FEATURES    Location/Qualifiers
     source 1. .27102
            /organism="Caenorhabditis elegans"
            /db_xref="taxon:6239"
            /chromosome="V"
            /clone="F22B8"
     gene   48. .1697
            /gene="F22B8.1"
     CDS    join(48. .335,775. .881,928. .1015,1061. .1177,1218. .1697)
            /gene="F22B8.1"
            /note="similar to 7TM receptor"
            /codon_start=1
            /db_xref="PID:e1187925"
            /db_xref="PID:g2653190"
            /db_xref="SPTREMBL:O17830"
            /translation="MRGHRLLCDHTFADKNNIYNILNNSLSFFKFWKMRINWIHHYAP
            KISGSCSFVINLLSIYIAHDERKTHLGNYRILLIFFSIYNILNSTIDVFMPMKSLFRT
            IALCLRGSFIFATYPILHAHFLYRYMVLFRSTFLSSCFIPYGIICSVSYCLTIIGFTF
            TMAYFWMVADTERLTYMQNIPLEEFNQNVDDLNIFIVIYHETSPAVLWKSWSAILILT
            TLSVISLVINSIFSILIITKLKSENLSMSNTTKRLQKQLIKALITQSTIPILVCFTPC
            LISWYLPVFGIDIGNGIHWALSVALSFFPVLDPLSLFPFLPIFGVRLREIPGIKRRVA
            IRTTDGTLHEESRKNNSLNAIKVNN"
     gene   10961. .12606
            /gene="F22B8.4"
     CDS    join(10961. .11044,11306. .11366,11495. .11613,11701. .12045,
            12100. .12160,12205. .12328,12402. .12606)
            /gene="F22B8.4"
```



```
            /codon_start=1
            /db_xref="PID:e1246773"
            /db_xref="PID:g2814399"
            /db_xref="SPTREMBL:O45392"
            /translation="MLTRKKPVVFYVTMSFLALLLILHNSDRNVSLVGKRFNCTWLSI
            LEDLKLVGSQEQIMVDLTIPVKNDSDIMPLSATSKNHLFNFNEMYNIVHRNWPNQNII
            LYSLDLTDTQILDIEKRPNVKVRKFDYSKYPKYVENWMEYRFKALILAVSFGLKMVTE
            LVLRKRFEIMQMPGGLMPIQNGFIRIHWMPFMSIWQNAPTMLTNSSIQTFINSTHSNY
            AVLTKGLLDYFPTFDTDVLKTNEKGLQVSAMLVYLARTPFTLEILKWHALCALEEKCM
            NPPTAKLKCDIIPTWNSYAGCFRYDQSSINILLFNQFRDHNSYFMAKNVGVVNRTT"
     gene   15374. .16506
            /gene="F22B8.3"
     CDS    join(15374. .15592,15713. .16014,16062. .16506)
            /gene="F22B8.3"
            /codon_start=1
            /db_xref="PID:e1187927"
            /db_xref="PID:g2653191"
            /db_xref="SPTREMBL:O17831"
            /translation="MDHGDNSDSYSHYVTFALRVAQFGFFATTTSCTMLILLTLYGVK
            RDPGSYKNLLLMFSGLGIVFATTEVLLFPNLHSYKAGYFFYSMERPPGLDTKNVSRFL
            VFYTGLYGSTICLLSVQPVYRYWAVFSESKLKFFKGWRLIFCIMYCAAFGADWGLSIY
            YFDEMDAYADHYFRAEMHKRYRLNISEIAGFSLVAYNFDNTPRWRNISCTISLFCIML
            IQYTIIIYCATIMYLKMEEKLKLLSFAIRNLHRQFYKTLVIQIFTPTIFLFSPLIFII
            FHPLFNVEIVVPTGMFLCAITIYPAADAIVIMYVVTDYRNTIKSK"
     gene   18635. .21655
            /gene="F22B8.7"
     CDS    join(18635. .18660,18866. .19017,19613. .19698,19747. .19857,
            19906. .19994,20249. .20395,20441. .20624,21189. .21269,
            21317. .21655)
            /gene="F22B8.7"
            /note="cDNA EST EMBL:C11005 comes from this gene; cDNA EST
            EMBL:C12970 comes from this gene; cDNA EST yk409a12.5
            comes from this gene; cDNA EST yk417e8.5 comes from this
            gene; cDNA EST yk438f6.3 comes from this gene; cDNA EST
            yk438f6.5 comes from this gene"
            /codon_start=1
            /db_xref="PID:e1246775"
            /db_xref="PID:g2814401"
            /db_xref="SPTREMBL:O45394"
            /translation="MNLIPQDVLKKYLLPIRDLATSTFVFKKKYTRECQRSTPRNFIN
            QFLLLFLKVPFSLLQELKLSKMLEIDKDKKVLIGCLAGSFLFYHTARQIYSYMSVKPR
            EWVPVGVVKSLHIFPIKSCKSVDVFAFKCTKLGPVMGDLEDRAFMLVEESTGKPITAR
            QKPKLVHVENYMTDGMLEVTVPGQPKLSVDLRKVLQNKRTIRATLFKNLKQDGYDCGD
            EVAKLLSDYIEEPNYRLIFYKEGLYTERTVIPDDQWWNNPVPKRNDDSGFTDLAPFHI
            ATDASLKVLNEKLDNKITMRYFRPSIYIEGCAAWDEDKWAEIRIGDAHLECFAPCTRC
            VLTTVDPVKGEMSKEMQPLKKLREFRLVPDGKMRKVHMESPVFGVYAGLVNEGYIHTG
            QTVYVRYKPTAF"
     gene   complement(22386. .23923)
            /gene="F22B8.5"
     CDS    complement(join(22386. .22471,22532. .22709,22756. .23313,
            23714. .23923))
```

```
                     /gene="F22B8.5"
                     /note="similar to 7TM receptor"
                     /codon_start=1
                     /db_xref="PID:e1246774"
                     /db_xref="PID:g2814400"
                     /db_xref="SPTREMBL:045393"
                     /translation="MWGRPSYLPHARPCIFMGELGFPSHAFPGFILVYLTLFHIKRQF
                     GSYKYLLVNFQLVGFCFASCETYLFHTFLHLYNASLIYFSLSHPLGFSNLTMEWHQAVT
                     TGLYSSTICLLAIQSIYRYWALFDTPKLEYFNGIYYFIWVGYITTFPGILWSLTVGYFP
                     AMDEPSMEYLSKEISLRYERNITDTPILDLIAYQGNDFRWKNLTGALMMTDISTFQYT
                     IIIVCGYNMYHGMQTKLSLLSAQHRRLHRQFFRALTIQVVSPTIILFCPVFFMIYVPF
                     ADLEMSFPSCIIQSGFTVYPALDSVIMMCCVSEYFRIFKKLFVSMKQPKSSQDDSEKK
                     RIQMVNPKI"
          gene       25087. .27040
                     /gene="F22B8.6"
          CDS        join(25087. .25161,25447. .26068,26559. .27040)
                     /gene="F22B8.6"
                     /note="predicted using Genefinder; similar to cystathione
                     gamma-lyase like"
                     /codon_start=1
                     /db_xref="PID:e1187924"
                     /db_xref="PID:g1752902"
                     /db_xref="SPTREMBL:045391"
                     /translation="MATFPHPGTDAVHVGQEPEQWDINQVVPPISMSTTYKQDNPGEP
                     KRYDYARGGMPTRDVLQRNLAALEGAKHCQAPSSGLAASSAVINLLNHGDHIVCSDDV
                     YGGTIRYIRQIAVKKYGMEVDSVDLTDVQNLEKAIKPNTKMVWFESPSNPLLKVVDIA
                     AVVQTAKRANPEIVVVVDNTFMTPYFQRPLSLGADIAVHSITKYINGHSDIIMGAAIT
                     NNDEFQQHLHFMQRAIGGVPSPFDCFLVNRGLKTLHVRMKAHFENALAIAKFLEAHDC
                     IESVLYPALPSHPQHTVHVKQTKGMSGMISFTLKGELCQSRAFLSALKVFTLAESLGG
                     CESLAELPSIMTHASVPSATRAELGITDNLIRISVGIEDLDDLIADLDQALKIAIPNV
                     *
BASE COUNT     8651 a    4766 c    4704 g    8981 t
ORIGIN

  Query Match           13.3%;  Score 17;  DB 18;  Length 27102;
  Best Local Similarity 100.0%;  Pred. No. 3.26e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  21118 ATCTCAACTTTTAAAAG 21134
          |||||||||||||||||
Qy      87 ATCTCAACTTTTAAAAG 103


RESULT   42
LOCUS       CELF56B3     38574 bp    DNA             INV       11-FEB-1998
DEFINITION  Caenorhabditis elegans cosmid F56B3.
ACCESSION   AF045646
NID         g2854196
KEYWORDS    .
SOURCE      Caenorhabditis elegans strain=Bristol N2.
  ORGANISM  Caenorhabditis elegans
```



```
            Eukaryotae; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
  REFERENCE 1  (bases 1 to 38574)
    AUTHORS Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
    TITLE   2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
    JOURNAL Nature 368 (6466), 32-38 (1994)
    MEDLINE 94150718
  REFERENCE 2  (bases 1 to 38574)
    AUTHORS Graves,T., Bradshaw,H. and Pronick,W.
    TITLE   The sequence of C. elegans cosmid F56B3
    JOURNAL Unpublished (1998)
  REFERENCE 3  (bases 1 to 38574)
    AUTHORS Waterston,R.
    TITLE   Direct Submission
    JOURNAL Submitted (02-FEB-1998) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
  COMMENT   Submitted by:
                  Genome Sequencing Center
                  Department of Genetics, Washington University,
                  St. Louis, MO  63110, USA, and
                  Sanger Centre, Hinxton Hall
                  Cambridge CB10 IRQ, England
                  e-mail: rw@nematode.wustl.edu and jes@sanger.ac.uk


            NOTICE:  This sequence may not be the entire insert of this clone.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we provide a small overlap between
            neighboring submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded or sequenced with an alternate
            chemistry; an attempt was made to resolve all sequencing problems,
            such as compressions and repeats; all regions were covered by
            sequence from more than one subclone

                        NEIGHBORING COSMID INFORMATION:
```



```
            The 5' cosmid is Y55F3C;3' cosmid is Y55F3B.  Actual start of this
            cosmid is at base position 1 of CELF56B3; actual end is at 38574 of
            CELF56B3

            NOTES:

            Coding sequences below are predicted from computer analysis, using
            the program Genefinder(P. Green and L. Hillier, ms in preparation).
FEATURES              Location/Qualifiers
     source           1. .38574
                      /organism="Caenorhabditis elegans"
                      /strain="Bristol N2"
                      /db_xref="taxon:6239"
                      /chromosome="IV"
                      /clone="F56B3"
          gene        complement(<2. .2277)
                      /gene="F56B3.5"
          CDS         complement(join(<2. .387,1647. .1743,1827. .1997,2044. .2111,
                      2157. .2277))
                      /gene="F56B3.5"
                      /note="contains similarity to enoly-CoA hydratases"
                      /codon_start=1
                      /evidence=not_experimental
                      /db_xref="PID:g2854202"
                      /translation="MLSQKILLNLRAPAFHPLATYSTTTNPANEVFIERLTGKDEGITI
                      LMMNRPAKHNSLGRVFMDQFREVLDELKYDPKTRVVIFNSKCDNVFCSGADLKERKTM
                      SQQEATRFVNGLRDSPTDVERLPQPVIAAIDGFALGGGLELALACDIRVASQKAKMGL
                      VETKWALIPGAGGSQRLYRIVGVAKAKELIYTAEVLNGADAAKLGVVNHVVEANPIEK
                      SLEIARKIIPRGPIAVKLAKLAINLGSQTDITSALSVEQQCTAQIVHSKDRLEGMAAF
                      AEKRD"
          gene        2739. .7489
                      /gene="F56B3.4"
          CDS         join(2739. .2794,2842. .3135,3192. .3419,3870. .3938,
                      4042. .4420,4684. .4765,5026. .5858,5901. .6103,6584. .6662,
                      6880. .7388,7441. .7489)
                      /gene="F56B3.4"
                      /note="contains similarity to human cyclin
                      A/CDK2-associated protein p19, an RNA polymerase II
                      elongation factor-like protein (GB:U33760)"
                      /codon_start=1
                      /evidence=not_experimental
                      /db_xref="PID:g2854201"
                      /translation="MTPLSKLCAGAVTQNVLKSYFLTPSQPRSFSRTITALEKSQKSP
                      SKQEHALQRSILEEVLVHEKAKPTTFGGKVAEKASNTPMYTAVVAGIGLIGAFIYVLA
                      GEFFAQDSPQTIFNKALALVRDDGRCQEIFGASIAGFGEETSRGRRRHVAHHKYEKDG
                      MQRIRVLPHVKGDRDEGIAQAEMEQRDGDWQWRHQIWARKQMPKNKKNNKKRAPASAP
                      KPAPPSAPPRLLQLASSNGSVLQADIRALQLSSTISTIKELGTDEEDCAEIKPIPVD
                      EHETYLDLLIKWCDQHGDDPEIAKAEKGKKKVVIPSWDQHFPSVLPMGNLLAVIKAA
                      YDLDITGLLILSPIQCFGVIFEFPFFIEKSLRFRFSKMSEHVKRPTATETDDDLLKMQE
                      EWEAANKKPSVEIHKMKNRQKVEEVASLPSKVVRDQAARFQIDFDEEQAASGVLFPVK
                      ERNFAERSEILAENLQKFQLYSKDDGFPEPLDLSAYFSAENPKTLTNLGGKSFFAAEF
```



```
                      DRINGNLAENLDFKQETTENPSGSDDFQAENEKYLKSLDFEKIQEMKLEITERFDPKL
                      LEFLKNRGNKASEETQKKPKKISKFKASRQQKEPENEQKGAPEPPAQAPPTKVVEEMM
                      NELEVLEEYGNREDQEKYNRLATVFWKIIDIFEIFGLNFFSVKNGKSEKFQDAVQLDL
                      TAKFGRNLVSRQQKNAVKLFDNCKLRPETSAGSTPDPIELARSSIDQIKTLYLEEIA
                      TDSSSKSIIYEFGKGLNPILDNCWTLIPVRRVLDAVERRNGTVTPDDVEIVRLSLLWT
                      LLLPSERKSAFFAPAEPNEFYVHIAEVFLIGAEILADDVIAECARRILNGYILAAAKD
                      AKIGLRMSGKVAGLDAFMPFYENLLKNYEQYSMGDEHFAMTILIAAYLNSPVGDSSKS
                      NFSLNLSRNSQN"
          gene        complement(8566. .11175)
                      /gene="F56B3.6"
          CDS         complement(join(8566. .8829,9647. .9850,11026. .11175))
                      /gene="F56B3.6"
                      /note="contains similarity to the human polyposis
                      locus-encoded protein (SW:Q00765)"
                      /codon_start=1
                      /evidence=not_experimental
                      /db_xref="PID:g2854203"
                      /translation="MSARGNTSAQNRAASALGAAASSTSGQPAETPRAALKRHVAPTE
                      GLAVAQNDIHNFLPESHGQIFDMAMMYAESAGFKRDHLSYAAFGLIAFFLVFGSVARL
                      LCNLIGPGYPTYASVKAIRSPGGDDDTVWLIYWTCFAVLYLVDFFSEAILSWFPFYYI
                      AKACFLVYLYLPQTQGSVMFYETIVDPLVIFVDKNLDKFYYNRSN"
          gene        complement(13598. .20517)
                      /gene="F56B3.7"
          CDS         complement(join(13598. .13766,15235. .15471,15901. .16120,
                      16170. .16344,18622. .18974,19197. .19397,19449. .19560,
                      20197. .20280,20329. .20517))
                      /gene="F56B3.7"
                      /note="Similar to UDP-glucuronosyltransferase"
                      /codon_start=1
                      /evidence=not_experimental
                      /db_xref="PID:g2854204"
                      /translation="MYLFFQIFFTPMTSPIRCLVFLFTLPRVIQTLHVLFYILLLGKS
                      HIDFTDSLIESLVENGHTVDLIIIARMNSHVKSNGTNMVTRKYTYGFQEESPWNKAPHL
                      LNPFKGRERTWNEHTIYMDIVTDLCEIGLSDAGLHAFISSQKYDIGIATEYENCGVAL
                      MKTYKIPSIVSVSSLPILDRQSLAGAGMPSPAVTQAIFEAEDLTTWGKFRNFLWAAH
                      ITLLINPYCEKHQSEVIKKHLGENFDTSNLIESLDLAFVNSNELIETPRVSSHKIKYI
                      GGINLKKSKDRLLDEEVEKVITQKPIGNGIVVFCPGTQVPSSLFPIEVRRAFAQAFRHF
                      PDFTFVWKYEMQDGDEQIFANTTNLRLLKWLPQTDLLNDARTKAFISHVGLNSTLESS
                      YAGVPILAVPLFADQPHNAFSGMSIGTTYMLDKTRLTTPNIVKGLEAVLYDSSYTLNA
                      KRISKMLHERPNPPKKIFVEWIEFASRNPLLHRNLNLIGQNMSVFEYYCLDIISFCIL
                      LLSIVISVFYLGTGWVFLLFLESHQGCAANLPNLPSSANFELRRTHQKFGITLLTKTI
                      GFCLQKFSKIAQN"
          gene        complement(22758. .25939)
                      /gene="F56B3.8"
          CDS         complement(join(22758. .22983,23560. .23825,24655. .24754,
                      25385. .25648,25722. .25939))
                      /gene="F56B3.8"
                      /note="similar to the L2P family of ribosomal proteins"
                      /codon_start=1
                      /evidence=not_experimental
                      /db_xref="PID:g2854205"
```

```
                     /translation="MATRNMPARGTKPLPRHLWDMSDLEEKTSGQYTHKPLRINRLGG
                     RDPETGRKVNQHIGGGVKFDYFMIDFHRRGPADQGATYDERILEIRRDPNRTCHIALC
                     AGINGKRWILATENMKFVEENRFFNRFWLENPENSPPSTPLSQKYLKIGRKIAPLAQN
                     RAGDVISTSGHLSENPVIGVEGNAYPIGSLAAGTKNSQFVLLEKSEISPLSSIERYPT
                     MDSETFVKAAGTSATIVRHQGDFTVVKLPHKHEFSLHRTCMATVGRLSHADIDGKIFG
                     SAQMHRRFGYKMSSGLFHKKDGYFGRKVRALPPVRILDTPPQAPPPNQRPTLSKKDLS
                     GLHGHAKVHNLIPAGYSTRDYPE"
     gene            complement(26504. .27940)
                     /gene="F56B3.9"
     CDS             complement(join(26504. .26708,26755. .26856,26902. .27212,
                     27259. .27477,27566. .27940))
                     /gene="F56B3.9"
                     /codon_start=1
                     /evidence=not_experimental
                     /db_xref="PID:g2854197"
                     /translation="MSTETHYKRQLFVILKLFIIMSFMTKCPKCPNYESKKTTNVYRH
                     LREVHNASEAEVDQFRIQRKEGKSLTTKGSGAWKCPVCELVIPTEKGLKNHMNQKHRK
                     KPIDPSTITLATTSPTMVSPPPKRIQLPKSTEKSLVSSVATVRAQPTQKPGGRYKCPV
                     ENCTPFKTRKELALHFSEKHDPGSELETTMSFETTEPKIWLPCRQEETCSSMQISLSS
                     HDGNRITHFRLCRHEGSYQSRAKVKNVGESKKKTGDNVCPAPINTEKHALTGEIGVVG
                     YFSHYGHSLEHARRALSDTETEYLIGLMKEGFNNRQIIRKLDQKYDENSRLGPVLSKD
                     LSYLRKTYCLNEGQLDDDDMKSIRRRVERDDPEDGIYHPSQPDSNGDNFILGTAFTKR
                     QNLKFNIPQCL"
     gene            27870. .29176
                     /gene="F56B3.3"
     CDS             join(27870. .27909,27964. .28036,28120. .28225,28445. .28532,
                     28598. .28749,28907. .28963,29015. .29176)
                     /gene="F56B3.3"
                     /codon_start=1
                     /evidence=not_experimental
                     /db_xref="PID:g2854200"
                     /translation="MNDMMIKSFKITKRLFSNYLNFQLSLVQPPITFPRFEMPINGKI
                     SDRPPRPPSLSASSSPAPGSARKPYFFLYIPHHHQHLHQLLHLELLDFLQKSDFFLFV
                     NSPPPPPSSSTSSSRTPGFCRKGRPTSTAQNASALLITLLICIPLKFNYNKLFPRSAL
                     MLFCSLICDFWNHFWLKIFFQNLQIQCETINFPYKNIGGMKKLNLNAREKSVQWSAFT
                     RRAFRER"
     gene            32349. .35123
                     /gene="F56B3.2"
     CDS             join(32349. .32615,32672. .32831,33506. .33966,34458. .34614,
                     35017. .35123)
...
Note: remainder of annotations omitted.

  Query Match          13.3%;  Score 17;  DB 18;  Length 38574;
  Best Local Similarity 100.0%;  Pred. No. 3.26e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  33152 TTTTAAAAGAATAAAGA 33168
          |||||||||||||||||
Qy     95 TTTTAAAAGAATAAAGA 111
```



```
RESULT   43
LOCUS       CELD1009    40475 bp    DNA             INV       25-NOV-1995
DEFINITION  Caenorhabditis elegans cosmid D1009.
ACCESSION   U40938
NID         g1072169
KEYWORDS    .
SOURCE      Caenorhabditis elegans strain=Bristol N2.
  ORGANISM  Caenorhabditis elegans
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Nematoda;
            Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea;
            Rhabditidae; Caenorhabditis.
REFERENCE   1  (bases 1 to 40475)
  AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
  TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
  JOURNAL   Nature 368 (6466), 32-38 (1994)
  MEDLINE   94150718
REFERENCE   2  (bases 1 to 40475)
  AUTHORS   Favello,T.
  TITLE     The sequence of C. elegans cosmid D1009
  JOURNAL   Unpublished (1995)
REFERENCE   3  (bases 1 to 40475)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-1995) Robert Waterston
COMMENT             Submitted by:
                    Genome Sequencing Center
                    Department of Genetics, Washington University,
                    St. Louis, MO  63110, USA, and
                    Sanger Centre, Hinxton Hall
                    Cambridge CB10 IRQ, England
                    e-mail: rw@nematode.wustl.edu and jes@sanger.ac.uk
                    NEIGHBORING COSMID INFORMATION:
          The 5' cosmid is C13E3, 200 bp overlap;3' end lies in a gap
          followed by the cosmid C48C5.  Actual start of this cosmid is at
          base position 197 of CELD1009; actual end is at 40475 of CELD1009

          NOTES:
          Coding sequences below are predicted from computer analysis, using
          the program Genefinder(P. Green and L. Hillier, ms in preparation).
```



```
FEATURES             Location/Qualifiers
     source          1. .40475
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /map="X"
     gene            7152. .9535
                     /gene="D1009.3"
     CDS             join(7152. .7184,7234. .7370,7794. .7881,7969. .8370,
                     8419. .8540,8762. .9160,9325. .9535)
                     /gene="D1009.3"
                     /note="coded for by C. elegans cDNA yk9a2.5; coded for by
                     C. elegans cDNA yk9a2.3;  similar to fibrinogen-like
                     protein A precursor"
                     /codon_start=1
                     /db_xref="PID:g1072170"
                     /translation="MMQDTKDEHLTSFAENPTKSVERDALVEPQFTRPEQVEKTHVEE
                     AKWPWGKKNEKMPRFLVTILLICLLVAALLIALIAIVVALNSNSASVREPEIVYTI
                     PSRDNQRDNVVHARSNIPKTLKKSLDLLEDTTTITTTSQTPDDTDTTEPFVEEEEY
                     HLEEATEHHEIATIEATTPEPSEIATKKPAIIYTSPPTTPKTTTSTGYLQFYDSIDGDA
                     SDNCLERLALGSPSGVYSIQSVEKFQAPCDMDTTTGGWTVIQRRVDGDGSFHRGTMKK
                     FVEGPGNLQGSHWLGLEKLHNLAPIGATPAILRIEIQGETCDLTCSKRPANTWVGEWK
                     VNFGKKQDGYKIQITEEGVGNLTTNGVDPFFGANGRRFSTNDNDQDENTFMNCAQFRM
                     VGPWWHPKSCSDVGLNGYFQTTSEKYDVKDRNQNAKRYFVWAFDKQIYNGYPYTIHVR
                     KSTMLLKTGDPQR"
     gene            10453. .12901
                     /gene="D1009.2"
     CDS             join(10453. .10552,10912. .11389,11434. .11734,11890. .11946,
                     12389. .12490,12539. .12901)
                     /gene="D1009.2"
                     /note="coded for by C. elegans cDNA CEESF41R;  contains a
                     cyclophilin-like PPIase domain"
                     /codon_start=1
                     /db_xref="PID:g1072171"
                     /translation="MPPEVRGNKRAFFDISINGEPAGRIVFSLWNHCCPRTVENFRAF
                     CTGELGKMNGHYASTQGSVFHRVIKGFMIQGGDITHGNGTGGYSIYGRTFDDENLALK
                     HKKPYLLSMANSGPDTNGSQFFITSEEVPHLDGKHCVFGEVIKGVFVVKAIENLETGN
                     EDKPVCKVEITHCGEMVRKGDVVGNAGAAASEPEPTGVKMEDEPKTRNWLMRYSKSPE
                     SREEKKKDKHGREEKRDRRRRSNDRHGRDRRSRSRSQSRDRNRRRDDRSGRDGRVGRN
                     ERDDRSGRDERRDDRRSSAPVNVKKKTQMKSPNFVHAPTKDGIKVRGRGKLTFLGGRT
                     RSTTPPHWKREESKKLTLEAHQKRQEELEERQKRAAQREKEEAERNARLEKERVERQN
                     QRELERQKQEAERELKRQQENEESGNRRRRRSSRSGSENDESELEEPKNRSASPRKRS
                     SSSSSSSSGHSSDSD"
     gene            complement(16809. .18460)
                     /gene="D1009.4"
     CDS             complement(join(16809. .17103,17539. .17673,17723. .17857,
                     18007. .18064,18409. .18460))
                     /gene="D1009.4"
                     /codon_start=1
                     /evidence=not_experimental
                     /db_xref="PID:g1072172"
                     /translation="MLHLIVLLVALSSAVTAGRPRRALDGLDGSGFGFDKRALNSLDG
```



```
                     AGFGFEKRALNSLDGQGFGFEKRALDGLDGAGFGFDKRALNSLDGAGFGFEKRALDGL
                     DGSGFGFDKRALNSLDGAGFGFEKRALNSLDGAGFGFEKRALDGLDGAGFGFDKRALN
                     SLDGAGFGFEKRALDGLDGAGFGFDKRALNSLDGNGFGFDKRTFKHSSNKLRSVFRNL
                     KGFKQH"
     tRNA            24748. .24819
                     /note="Gln;  codon recognized: CAA"
     gene            30386. .37985
                     /gene="D1009.1"
     CDS             join(30386. .30390,33624. .33797,34402. .34489,34620. .34727,
                     35690. .35888,35971. .36427,36472. .36635,37023. .37468,
                     37555. .37620,37675. .37771,37822. .37985)
                     /gene="D1009.1"
                     /note="coded for by C. elegans cDNA yk83h3.3; coded for by
                     C. elegans cDNA yk36c4.5; coded for by C. elegans cDNA
                     yk83h3.5;  strong similarity to mouse fatty acod
                     transporter protein precursor (PIR:A55093) containg and
                     AMP binding pattern (PROSITE:PS00455)"
                     /codon_start=1
                     /db_xref="PID:g1072173"
                     /translation="MREMPDSPKFALVTFVVYAVVLYNVNSVPWKPVPIGYVVPRLLR
                     TDFGRRALATLPRDFAGLKLLISVKSTIRGLFKKDRPIHEIFLNQVKQHPNKVAIIEI
                     ESGRQLTYQELNALANQYANLYVSEGYKMGDVVALFMENSIDFFAIWLGLSKIGVVSA
                     FINSNLKLEPLAHSINVSKCKSCITNINLLPMFKAARENKLISDEIHVPLAGTQVDGR
                     HRSLQQDLHLFSEDEPPVIDGLNFRSVLCYIYTSGTTGNPKPAVIKHFRYFWIAMGAG
                     KAFGINKSDVVYITMPMYHSAAGIMGIGSLIAFGSTAVIRKKFSASNFWKDCVKYNVT
                     ATQYIGEICRYLLAANPCPEEKQHNVRLMWGNGLRQDIWKEFVGRPGIKKIGELYGST
                     EGNSLVNVDNHVGACGFMPIYPHIGSLYPVRLIKVDRATGELERDKNGLCVPCVPGE
                     TGEMVGVIKEKDILLKFPEGYVSEGDTAKKIYRDVFKHGDKVFASGDILHWDDLGYLYF
                     VDRCGDTFRWKGENVSTTEVEGILQPVMDVEDATVYGVTVGKMEGRAGMAGIVVKDGT
                     DVEKFIADITSRLTENLASYAIPVFIRLCKEVDRTGTFKLKKTDLQKQGYDLVACKGD
                     PIYYWSAAEKSYKPLTDKMQQDIDTGVYDRI"
     gene            complement(38456. .39210)
                     /gene="D1009.5"
     CDS             complement(join(38456. .38503,38746. .38809,38855. .38924,
                     38974. .39110,39158. .39210))
                     /gene="D1009.5"
                     /note="similar to t complex testis-specific protein
                     (GB:U21673)"
                     /codon_start=1
                     /evidence=not_experimental
                     /db_xref="PID:g1072174"
                     /translation="MEDANDRNFVLRPTPGQKFRPKAVAGMIQEILGEKLGALTIYNV
                     DEAELVSKDISASIRERLKGLQLPRYKYVIQTMIAEQCGNGATTAVQCVWDEDCDGYL
                     TQRYVTGSIWCEVLVFAIFHY"
     tRNA            38656. .38737
                     /gene="D1009.5"
                     /note="Leu;  codon recognized: CTT"
BASE COUNT    13436 a   6790 c   6976 g  13273 t
ORIGIN

  Query Match          13.3%;  Score 17;  DB 18;  Length 40475;
```

```
  Best Local Similarity 100.0%;  Pred. No. 3.26e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  27471 GATATTCTTTATTCTTT 27487
          |||||||||||||||||
Cp    116 GATATTCTTTATTCTTT 100


RESULT  44
LOCUS       AF039905    40697 bp    DNA             PRI       11-FEB-1998
DEFINITION  Homo sapiens cosmid Q95D4, chromosome 21 5' of IFNAR2.
ACCESSION   AF039905
NID         g2766549
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 40697)
  AUTHORS   Reboul,J. and Lutfalla,G.
  TITLE     Comparing interferon receptors gene clusters in vertebrates
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 40697)
  AUTHORS   Reboul,J. and Lutfalla,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-DEC-1997) IGM, CNRS, 1919 route de Mende, Montpellier
            cedex 5 34293, France
FEATURES             Location/Qualifiers
     source          1. .40697
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /clone="cosmid Q95D4"
     repeat_region   65. .220
                     /rpt_family="unidentified"
                     /evidence=not_experimental
     exon            complement(215. .445)
                     /note="GRAIL, score = 65.000%, comment = good"
                     /evidence=not_experimental
     exon            224. .387
                     /note="GRAIL, score = 70.000%, comment = goodshadow"
                     /evidence=not_experimental
     exon            complement(241. .418)
                     /note="Xpound exon prediction, score = 86% (0%)"
                     /evidence=not_experimental
     repeat_region   complement(740. .901)
                     /rpt_family="MIR"
                     /evidence=not_experimental
     repeat_region   2137. .2166
                     /note="1I1 with 83% homology to 1I2"
                     /rpt_type=inverted
```



```
                     /rpt_unit=tatcgtgtagcaaaccctaaccacagcagc
                     /evidence=not_experimental
     repeat_region   2404. .2433
                     /note="1I2 with 83% homology to 1I1"
                     /rpt_type=inverted
                     /rpt_unit=gctgctgtggttagcgcttgctgcgtgata
                     /evidence=not_experimental
     repeat_region   complement(3247. .3578)
                     /rpt_family="MLT1B"
                     /evidence=not_experimental
     repeat_region   4351. .4388
                     /note="2I1 with 92% homology to 2I2"
                     /rpt_type=inverted
                     /rpt_unit=aataattatc-aatgatgcttgtttgcaataaagtgatt
                     /evidence=not_experimental
     repeat_region   4401. .4439
                     /note="2I2 with 92% homology to 2I1"
                     /rpt_type=inverted
                     /rpt_unit=aatcactttattgcaaacaaacatcattagaaaaatatt
                     /evidence=not_experimental
     repeat_region   complement(4712. .5017)
                     /rpt_family="AluJo"
                     /evidence=not_experimental
     repeat_region   5338. .5357
                     /note="3I1 with 95% homology to 3I2"
                     /rpt_type=inverted
                     /rpt_unit=agaggcagggtggggtgtgg
                     /evidence=not_experimental
     repeat_region   5547. .5566
                     /note="3I2 with 95% homology to 3I1"
                     /rpt_type=inverted
                     /rpt_unit=ccacacccccacgctgcctct
                     /evidence=not_experimental
     repeat_region   5784. .6028
                     /rpt_family="MIR"
                     /evidence=not_experimental
     repeat_region   6609. .6664
                     /note="4I1 with 73% homology to 4I2"
                     /rpt_type=inverted
                     /evidence=not_experimental
     repeat_region   6910. .6966
                     /note="4I2 with 73% homology to 4I1"
                     /rpt_type=inverted
                     /evidence=not_experimental
     exon            complement(7497. .7664)
                     /note="GRAIL, score = 78.000%, comment = excellent"
                     /evidence=not_experimental
     repeat_region   complement(7864. .8138)
                     /rpt_family="AluJb"
                     /evidence=not_experimental
     repeat_region   8335. .8459
```



```
                     /rpt_family="L2"
                     /evidence=not_experimental
     repeat_region   complement(8510. .8742)
                     /rpt_family="MER30"
                     /evidence=not_experimental
     exon            8800. .9061
                     /note="GRAIL, score = 54.000%, comment = good"
                     /evidence=not_experimental
     repeat_region   9071. .9168
                     /rpt_family="L2"
                     /evidence=not_experimental
     repeat_region   9234. .9306
                     /rpt_family="MER2"
                     /evidence=not_experimental
     repeat_region   9307. .9386
                     /rpt_family="MADE1"
                     /evidence=not_experimental
     repeat_region   9387. .9436
                     /rpt_family="MER2"
                     /evidence=not_experimental
     repeat_region   complement(9437. .9730)
                     /rpt_family="AluSc"
                     /evidence=not_experimental
     repeat_region   9733. .9930
                     /rpt_family="MER2"
                     /evidence=not_experimental
     repeat_region   9934. .10404
                     /rpt_family="L2"
                     /evidence=not_experimental
     repeat_region   10608. .10726
                     /rpt_family="L2"
                     /evidence=not_experimental
     exon            10940. .11034
                     /note="GRAIL, score = 68.000%, comment = good"
                     /evidence=not_experimental
     repeat_region   11400. .11700
                     /rpt_family="unidentified"
                     /evidence=not_experimental
     repeat_region   11820. .12120
                     /rpt_family="L1M4"
                     /evidence=not_experimental
     repeat_region   12120. .12220
                     /rpt_family="unidentified"
                     /evidence=not_experimental
     repeat_region   12226. .12522
                     /rpt_family="AluSx"
                     /evidence=not_experimental
     repeat_region   12828. .12963
                     /rpt_family="L1M4"
                     /evidence=not_experimental
     repeat_region   13020. .13319
```



```
                     /rpt_family="AluY"
                     /evidence=not_experimental
     misc_feature    13154. .13295
                     /note="GC score = 6.80 (142bp)"
                     /note="Region: GC content"
                     /evidence=not_experimental
     repeat_region   13360. .13480
                     /rpt_family="L1ME"
                     /evidence=not_experimental
     repeat_region   complement(13487. .13851)
                     /rpt_family="MLT1E"
                     /evidence=not_experimental
     repeat_region   13945. .14002
                     /note="homology = 94.80%, score = 23, counts = 2"
                     /rpt_type=tandem
                     /rpt_unit=atgaagggcattaattatgtctgcaaaac
                     /evidence=not_experimental
     repeat_region   14514. .14813
                     /rpt_family="AluJb"
                     /evidence=not_experimental
     repeat_region   15160. .15257
                     /rpt_family="FLAM_C"
                     /evidence=not_experimental
     repeat_region   15320. .15635
                     /rpt_family="L2"
                     /evidence=not_experimental
     exon            complement(15716. .15811)
                     /note="Xpound exon prediction, score = 76% (0%)"
                     /evidence=not_experimental
     repeat_region   complement(15906. .16328)
                     /rpt_family="MLT2FB"
                     /evidence=not_experimental
     repeat_region   16333. .16632
                     /rpt_family="L2"
                     /evidence=not_experimental
     repeat_region   complement(16858. .17200)
                     /rpt_family="MLT1A2"
                     /evidence=not_experimental
     repeat_region   complement(17346. .17473)
                     /rpt_family="MER20"
                     /evidence=not_experimental
     repeat_region   complement(17709. .18006)
                     /rpt_family="AluSx"
                     /evidence=not_experimental
     repeat_region   complement(19099. .19219)
                     /rpt_family="MER46"
                     /evidence=not_experimental
     exon            complement(19330. .19431)
                     /note="GRAIL, score = 42.000%, comment = marginal"
                     /evidence=not_experimental
     repeat_region   complement(19809. .20415)
```

```
                     /rpt_family="MER4A"
                     /evidence=not_experimental
     exon            19969. .20066
                     /note="MZEF, score = 50.8%"
                     /evidence=not_experimental
...
Note: remainder of annotations omitted.

  Query Match           13.3%;  Score 17;  DB 26;  Length 40697;
  Best Local Similarity 100.0%;  Pred. No. 3.26e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  23493 GATGAGACACTTTGAGA 23509
          |||||||||||||||||
Cp     89 GATGAGACACTTTGAGA 73


RESULT   45
LOCUS       AF020803     45034 bp    DNA             PRI       22-JAN-1998
DEFINITION  Homo sapiens cosmid Q37D2 from the Down Syndrome critical region on
            chromosome 21q22.2, complete sequence.
ACCESSION   AF020803
NID         g2801437
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 45034)
  AUTHORS   Blechschmidt,K., Rump,A. and Rosenthal,A.
  TITLE     Sequencing and Analysis of 900 kb from the Down Syndrome Critical
            Region on Chromosome 21q22.2
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 45034)
  AUTHORS   Blechschmidt,K., Nordsiek,G., Drescher,B., Weber,J., Schattevoy,R.,
            Rosenthal,A., Yaspo,M.-L., Osoegawa,K. and Soeda,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-AUG-1997) Genome Analysis, Institute for Molecular
            Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
FEATURES             Location/Qualifiers
     source          1. .45034
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.2"
                     /clone="cosmid Q37D2"
     repeat_region   complement(1136. .1421)
                     /rpt_family="ALU"
                     /evidence=not_experimental
     repeat_region   complement(1899. .1957)
                     /rpt_family="MLT2A1"
```



```
                     /evidence=not_experimental
     repeat_region   complement(1998. .2037)
                     /rpt_family="MLT2A1"
                     /evidence=not_experimental
     repeat_region   complement(2058. .2140)
                     /rpt_family="MLT2A1"
                     /evidence=not_experimental
     repeat_region   complement(2158. .2232)
                     /rpt_family="MLT2A1"
                     /evidence=not_experimental
     repeat_region   complement(2233. .2287)
                     /rpt_family="MLT2A1"
                     /evidence=not_experimental
     repeat_region   complement(2288. .2345)
                     /rpt_family="MLT2A1"
                     /evidence=not_experimental
     exon            complement(2755. .2819)
                     /note="MZEF, score = 66.6%"
                     /evidence=not_experimental
     repeat_region   complement(3151. .3235)
                     /rpt_family="MIR"
                     /evidence=not_experimental
     repeat_region   complement(3448. .3528)
                     /rpt_family="MLT1E"
                     /evidence=not_experimental
     exon            complement(3525. .3668)
                     /note="GRAIL, score = 43.000%, comment = marginal"
                     /evidence=not_experimental
     repeat_region   complement(3551. .3646)
                     /rpt_family="MLT1D"
                     /evidence=not_experimental
     repeat_region   complement(3650. .3701)
                     /rpt_family="MLT1D"
                     /evidence=not_experimental
     exon            complement(3743. .3935)
                     /note="GRAIL, score = 63.000%, comment = good"
                     /evidence=not_experimental
     repeat_region   complement(3773. .3863)
                     /rpt_family="MLT1D"
                     /evidence=not_experimental
     repeat_region   complement(3865. .3930)
                     /rpt_family="MLT1D"
                     /evidence=not_experimental
     exon            4261. .4366
                     /note="GRAIL, score = 47.000%, comment = marginal shadow"
                     /evidence=not_experimental
     repeat_region   4684. .4770
                     /note="homology = 78.20%, score = 20, counts = 3"
                     /rpt_type=tandem
                     /rpt_unit=attttttttattcatttagataattttaa
                     /evidence=not_experimental
```



```
     repeat_region   complement(5507. .5558)
                     /rpt_family="MLT1A"
                     /evidence=not_experimental
     repeat_region   complement(5562. .5666)
                     /rpt_family="MLT1A"
                     /evidence=not_experimental
     repeat_region   complement(5676. .5749)
                     /rpt_family="MLT1A"
                     /evidence=not_experimental
     repeat_region   complement(5780. .5826)
                     /rpt_family="MLT1A"
                     /evidence=not_experimental
     repeat_region   complement(5947. .5988)
                     /rpt_family="ALU"
                     /evidence=not_experimental
     repeat_region   complement(5993. .6148)
                     /rpt_family="ALU"
                     /evidence=not_experimental
     repeat_region   complement(6149. .6228)
                     /rpt_family="ALU"
                     /evidence=not_experimental
     repeat_region   complement(6231. .6275)
                     /rpt_family="MLT1A"
                     /evidence=not_experimental
     repeat_region   complement(7132. .7159)
                     /rpt_family="MLT2C2"
                     /evidence=not_experimental
     repeat_region   7187. .7218
                     /note="homology = 96.90%, score = 28, counts = 16"
                     /rpt_type=tandem
                     /rpt_unit=ac
                     /evidence=not_experimental
     repeat_region   complement(7263. .7365)
                     /rpt_family="MLT2B2"
                     /evidence=not_experimental
     repeat_region   complement(7451. .7520)
                     /rpt_family="MLT2B2"
                     /evidence=not_experimental
     exon            7467. .7554
                     /note="GRAIL, score = 61.000%, comment = good"
                     /evidence=not_experimental
     repeat_region   complement(7553. .7590)
                     /rpt_family="MLT2B2"
                     /evidence=not_experimental
     repeat_region   8177. .8462
                     /rpt_family="ALU"
                     /evidence=not_experimental
     exon            10819. .10844
                     /note="Xpound exon prediction, score = 78% (0%)"
                     /evidence=not_experimental
     repeat_region   11244. .11375
```



```
                     /rpt_family="ALU"
                     /evidence=not_experimental
     repeat_region   11376. .11507
                     /rpt_family="ALU"
                     /evidence=not_experimental
     repeat_region   11617. .11785
                     /rpt_family="ALU"
                     /evidence=not_experimental
     exon            12645. .12722
                     /note="GRAIL, score = 52.000%, comment = good"
                     /evidence=not_experimental
     repeat_region   13374. .13436
                     /note="homology = 71.40%, score = 20, counts = 9"
                     /rpt_type=tandem
                     /rpt_unit=tcttttt
                     /evidence=not_experimental
     exon            complement(13548. .13683)
                     /note="MZEF, score = 92.1%"
                     /evidence=not_experimental
     repeat_region   complement(14814. .14865)
                     /rpt_family="L1MA2"
                     /evidence=not_experimental
     repeat_region   complement(15112. .15335)
                     /rpt_family="ALU"
                     /evidence=not_experimental
     repeat_region   complement(15336. .15399)
                     /rpt_family="ALU"
                     /evidence=not_experimental
     repeat_region   complement(15715. .16001)
                     /rpt_family="ALU"
                     /evidence=not_experimental
     exon            18241. .18336
                     /note="GRAIL, score = 75.000%, comment = excellent"
                     /evidence=not_experimental
     repeat_region   complement(18649. .18714)
                     /rpt_family="L1PA15"
                     /evidence=not_experimental
     repeat_region   complement(18716. .19241)
                     /rpt_family="L1PA15"
                     /evidence=not_experimental
     exon            complement(18743. .18794)
                     /note="GRAIL, score = 50.000%, comment = good shadow"
                     /evidence=not_experimental
     repeat_region   complement(19254. .19319)
                     /rpt_family="L1PA11"
                     /evidence=not_experimental
     repeat_region   19617. .19691
                     /rpt_family="MIR"
                     /evidence=not_experimental
     repeat_region   19810. .19863
                     /rpt_family="MSTA"
```

```
                     /evidence=not_experimental
     repeat_region   19921. .20060
                     /rpt_family="MSTA"
                     /evidence=not_experimental
     repeat_region   20063. .20108
                     /rpt_family="MSTA"
                     /evidence=not_experimental
     repeat_region   20109. .20165
                     /rpt_family="MS"
                     /evidence=not_experimental
     repeat_region   complement(20323. .20410)
                     /rpt_family="MLTIC"
                     /evidence=not_experimental
     repeat_region   complement(20531. .20588)
...
Note: remainder of annotations omitted.

  Query Match           13.3%;  Score 17;  DB 26;  Length 45034;
  Best Local Similarity 100.0%;  Pred. No. 3.26e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  17799 ACTTTGAGAATCTGAAA 17815
          |||||||||||||||||
Cp     81 ACTTTGAGAATCTGAAA 65


RESULT   46
LOCUS       AB017070    45911 bp    DNA             PLN       02-SEP-1998
DEFINITION  Arabidopsis thaliana genomic DNA, chromosome 5, P1 clone: MNL12,
            complete sequence.
ACCESSION   AB017070
NID         g3510346
KEYWORDS    HTG.
SOURCE      Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui P1
            clone:MNL12.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Charophyta/Embryophyta group;
            Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta;
            Magnoliophyta; eudicotyledons; Rosidae; Capparales; Brassicaceae;
            Arabidopsis.
REFERENCE   1  (sites)
  AUTHORS   Nakamura,Y.
  TITLE     Structural Analysis of Arabidopsis thaliana Chromosome 5. IX
  JOURNAL   Unpublished (1998)
REFERENCE   2  (bases 1 to 45911)
  AUTHORS   Nakamura,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-AUG-1998) to the DDBJ/EMBL/GenBank databases.
            Yasukazu Nakamura, Kazusa DNA Research Institute, Laboratory of
            Gene Structure 2; 1532-3, Yana, Kisarazu, Chiba 292, Japan
            (E-mail:ynakamu@kazusa.or.jp, Tel:+81-438-52-3935,
```



```
            Fax:+81-438-52-3934)
  FEATURES             Location/Qualifiers
       source          1. .45911
                       /organism="Arabidopsis thaliana"
                       /strain="Columbia"
                       /db_xref="taxon:3702"
                       /chromosome="5"
                       /clone="MNL12"
                       /clone_lib="Mitsui P1"
  BASE COUNT    15086 a    7821 c    7863 g   15141 t
  ORIGIN

    Query Match           13.3%;  Score 17;  DB 24;  Length 45911;
    Best Local Similarity 100.0%;  Pred. No. 3.26e+01;
    Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

  Db  34165 ATTCTTTATTCTTTTAA 34181
            |||||||||||||||||
  Cp    113 ATTCTTTATTCTTTTAA 97


  RESULT   47
  LOCUS       CELD1007    47554 bp    DNA             INV       14-MAY-1997
  DEFINITION  Caenorhabditis elegans cosmid D1007.
  ACCESSION   AF003151
  NID         g2088797
  KEYWORDS    .
  SOURCE      Caenorhabditis elegans strain=Bristol N2.
    ORGANISM  Caenorhabditis elegans
              Eukaryotae; mitochondrial eukaryotes; Metazoa; Nematoda;
              Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea;
              Rhabditidae; Peloderinae; Caenorhabditis.
  REFERENCE   1  (bases 1 to 47554)
    AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
              Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
              Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
              Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
              Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
              Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
              O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
              Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
              Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
              Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
              Wilkinson-Sproat,J. and Wohldman,P.
    TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
              elegans
    JOURNAL   Nature 368 (6466), 32-38 (1994)
    MEDLINE   94150718
  REFERENCE   2  (bases 1 to 47554)
    AUTHORS   Davidson,S. and Rohlfing,T.
    TITLE     The sequence of C. elegans cosmid D1007
```



```
    JOURNAL   Unpublished (1997)
  REFERENCE   3  (bases 1 to 47554)
    AUTHORS   Waterston,R.
    TITLE     Direct Submission
    JOURNAL   Submitted (08-MAY-1997)
  COMMENT     Submitted by:
                      Genome Sequencing Center
                      Department of Genetics, Washington University,
                      St. Louis, MO  63110, USA, and
                      Sanger Centre, Hinxton Hall
                      Cambridge CB10 IRQ, England
                      e-mail: rw@nematode.wustl.edu and jes@sanger.ac.uk


              NOTICE:  This sequence may not be the entire insert of this clone.
              It may be shorter because we only sequence overlapping sections
              once, or longer because we provide a small overlap between
              neighboring submissions.

              This sequence was finished as follows unless otherwise noted:
              all regions were double stranded or sequenced with an alternate
              chemistry; an attempt was made to resolve all sequencing problems,
              such as compressions and repeats; all regions were covered by
              sequence from more than one subclone

                           NEIGHBORING COSMID INFORMATION:


              The 5' cosmid is C44E4, 1602 bp overlap;3' end lies in a gap.
              Actual start of this cosmid is at base position 1497 of CELD1007;
              actual end is at 47554 of CELD1007.  This cosmid lies in an
              unanchored cluster, the orientation of which is unknown.

              NOTES:

              Coding sequences below are predicted from computer analysis, using
              the program Genefinder(P. Green and L. Hillier, ms in preparation).
  FEATURES             Location/Qualifiers
       source          1. .47554
                       /organism="Caenorhabditis elegans"
                       /strain="Bristol N2"
                       /db_xref="taxon:6239"
                       /chromosome="I"
                       /clone="D1007"
       gene            complement(1936. .3696)
                       /gene="D1007.8"
       CDS             complement(join(1936. .2294,2351. .2546,2595. .2810,
                       3253. .3696))
                       /gene="D1007.8"
                       /codon_start=1
```



```
                       /evidence=not_experimental
                       /db_xref="PID:g2088806"
                       /translation="MTVYKGSKKEAKPKECLLFPDKKTNTVRLEKITSNIVKKTRDL
                       DQGTELALKRGIERLRTSSNNQRSGPSSPEEKAKIQKQMQRDTSDSSSDSDGSSDSDK
                       SNGDSSDDEESEKILLEEKKKPKPTTERRPAPPSSSNTVNHKSSKETHRYIPEHAV
                       PMPISALFNSTKERIEYGTADVVEVQVMGVIESVKPGDSDVEKMITDKDNHNILVQKI
                       LTASPTEESAKKFNVGDRVEVFGKVRILENGKKFLNAFSMHKASDEQFAAPQIICRPS
                       EKFYKKASPTGASTNVPTRPGLKNKVGSLNRFPMGMLEQSSSCENNMGVWPTEIEQLA
                       AGISTSGLVKEESDVSEDSSDDSDEEDEEEEEEDEEDSFDKGPVSLVPMARKEKPKE
                       EPEPEYDSFENA"
       gene            4574. .8915
                       /gene="D1007.7"
       CDS             join(4574. .4735,4805. .5063,5110. .5629,5669. .6321,
                       6588. .7126,7188. .7322,7642. .8093,8584. .8678,8734. .8915)
                       /gene="D1007.7"
                       /note="contains similarity to an RNA recognition motif;
                       coded for by C. elegans cDNA yk52g12.5; coded for by C.
                       elegans cDNA yk6a7.5; coded for by C. elegans cDNA
                       yk52g12.3; coded for by C. elegans cDNA yk6a7.3"
                       /codon_start=1
                       /db_xref="PID:g2088805"
                       /translation="METEVMKSPNNELSSLFDSKNMSKNKIQDITKAAIKAKSQYKHV
                       VFSVEKLINKCKPDQRLNVLYVIDSIVRASKHQLKEKDTFGPRFMKQFDKFLMPLLKC
                       GQKEKMRTVRTLNLWMSNKVFKESEIQPLREMCKASGLTIDFEEVELAVKGKQADMSI
                       YSGVYKKKPKRSSSSSQPKSRTPTNPRPDDGLLGAGPSSALRSVPDIPNFVLSEDYFL
                       GTISRREMLELVQKPGIDRSGVLSKDKNLLQRALQIFAGSLSQKVEEVLAENNRINGS
                       SIQNVLTKDFEYSDDEEEKEKEPQPEKQKNLPHAQVLLMFLIFFSNSRLAQSLLTQPQ
                       ILAKLAEVLIPQGNPPGLPFPGEHIVPTSSAALTLGAPPPNLMALQQSLPGFPNQQL
                       GLPNLSGLNQAQLMNVQNAQNMLQLQQRAAQLQALQGNPNAQRNLLMLGNPLLNPFAL
                       QHGVNPMLNDLQAAAAAQQQAMLNEAAQSPEKKILELSGGNSGINNSGDVERARLREK
                       EKERESKERRRMGLPPVRIGFTIIASRTLWLKKIPTNIVENDLKQAVESCGEASRVKV
                       IGNRACAYITMENRRSANDVVSKMREVSVAKKMVKVYWARSPGMDSDQFSDLDWDSNRG
                       VLEIPYEKLPLDLVALCEGAMLDIESLPIEKKLLYKETGETVISIPPPNIQPPVPHPP
                       PMGFPPQHQLTQLPGQPRPAGLPPGVPPMFNLNAPPPPGIPGYPPAPPPPGVGPPPPQ
                       GIPPMGFDPNKPPPPMPQQGFNAGAPPPPFGRGAGPMSSFPPPPRGGHKHMPPPSFR
                       GGRGSNGPPPPHFDRRGGGGPPFPPENGNGRLLQSEMWNREQRKDRGGGAGRDGG
                       RERRDYDRDRSQIDRRRQDDMGARRRSRWGDDDRRDDDRRDDRRDDRRESRRRSPRSP
                       RSPDRRTRRSPSYEREEPPVKKTSVEEETVSSTTLDELKPSVEPTPVPAPIPAPAPEL
                       KAAEEPVKIVAEHHEDQTDEVPMDLE"
       gene            complement(15626. .16783)
                       /gene="D1007.9"
       CDS             complement(join(15626. .15850,16471. .16630,16686. .16783))
                       /gene="D1007.9"
                       /codon_start=1
                       /evidence=not_experimental
                       /db_xref="PID:g2088807"
                       /translation="MVDKIKKKDYKKMEYFITEFSENSKLCRIFIEKTNGISFWGGRD
                       HVCSDPEIGILEKGFWAALRENTNVQQKFEITDRKMHNFMIPTAEKLKKKNWATTRDC
                       RQFLAEFCVTLGNNIFEILKKNSATEVLSSMVLYFFTRKICHEYFFEDLKNLISKKSW
                       "
       gene            complement(17857. .19608)
```

```
                     /gene="D1007.10"
     CDS             complement(join(17857. .17973,18285. .18419,18831. .18920,
                     19095. .19268,19318. .19465,19511. .19608))
                     /gene="D1007.10"
                     /codon_start=1
                     /evidence=not_experimental
                     /db_xref="PID:g2088808"
                     /translation="MHEKCANKPISTGGDCVASNTLCNVHFQNLVFPKHYIDFGEYVR
                     DTFALRSRCTDLTLFACWYCFIAICHVATNTIFSRYVKTIHLVQRPLKTRIRGQAVFL
                     VQFFTMNSIAALAAILFGAAGLSFMFIHFNYAAILFLESYRVFVRSLFVLGRLSVSTS
                     TARMNDIEEQKHLSDEEITDLIVNSQKRIGKYLNFRNIYITSTVFVTFLLNILYAITG
                     LLINRKMLAFGFAIGANQVIDDILNGREYGSTELL"
     gene            complement(20281. .20939)
                     /gene="D1007.11"
     CDS             complement(join(20281. .20638,20689. .20939))
                     /gene="D1007.11"
                     /codon_start=1
                     /evidence=not_experimental
                     /db_xref="PID:g2088809"
                     /translation="MRTLRIAQYSVLTVGFAIYMYRLIEEIPIDIRNLNSDSLEGIIN
                     SDELCDVTVSNRNRGLLVRNDSLDLDILKAKFTTFFSKRYLTRFLSEQVPPLHVIDEA
                     LLVKRFVMCACFMVFCLTVIWPLVIRRMGNLIKRLSVLNQLEDAESVEWARCIREFTQ
                     EKLAVLCFCIVPPFAQTDKLVSDKIKLFREHKILRIRSVQHI"
     gene            complement(21075. .22072)
                     /gene="D1007.12"
     CDS             complement(join(21075. .21161,21414. .21623,21669. .21770,
                     21992. .22072))
                     /gene="D1007.12"
                     /note="Similar to 60S ribosomal protein L24; coded for by
                     C. elegans cDNA yk71b8.5; coded for by C. elegans cDNA
                     yk99e8.5; coded for by C. elegans cDNA yk78g2.3; coded for
                     by C. elegans cDNA yk103d8.3; coded for by C. elegans cDNA
                     yk79g4.3; coded for by C. elegans cDNA yk96d9.3; coded for
                     by C. elegans cDNA yk71b8.3; coded for by C. elegans cDNA
                     yk81c9.3; coded for by C. elegans cDNA yk79a4.3; coded for
                     by C. elegans cDNA yk89b7.3; coded for by C. elegans cDNA
                     CEMSC16F; coded for by C. elegans cDNA yk74c6.3; coded for
                     by C. elegans cDNA yk62c11.3; coded for by C. elegans cDNA
                     yk83e5.3; coded for by C. elegans cDNA yk99e8.3; coded for
                     by C. elegans cDNA yk70e3.3; coded for by C. elegans cDNA
                     yk63a1.3; coded for by C. elegans cDNA yk172a2.5; coded
                     for by C. elegans cDNA yk89b7.5; coded for by C. elegans
                     cDNA yk74c6.5; coded for by C. elegans cDNA yk78g2.5;
                     coded for by C. elegans cDNA yk79g4.5; coded for by C.
                     elegans cDNA yk81c9.5; coded for by C. elegans cDNA
                     yk103d8.5; coded for by C. elegans cDNA yk70e3.5; coded
                     for by C. elegans cDNA yk138g11.5; coded for by C. elegans
                     cDNA yk62c11.5; coded for by C. elegans cDNA yk170h5.5;
                     coded for by C. elegans cDNA yk63a1.5; coded for by C.
                     elegans cDNA yk172a2.3; coded for by C. elegans cDNA
                     yk138g11.3; coded for by C. elegans cDNA yk170h5.3; coded
```



```
                     for by C. elegans cDNA yk83e5.5"
                     /codon_start=1
                     /db_xref="PID:g2088798"
                     /translation="MKVETCVYSGYKIHPGHGKRLVRTDGKVQIPLSGKALKGAKLRR
                     NPRDIRWTVLYRIKNKKGTHGQEQVTRKETKKSVQVVNRAVAGLSLDAILAKRNQTED
                     FRRQQREQAAKIAKDANKAVRAAKAAANKEKKASQPKTQQKTAKNVKTAAPRVGGKR"
     misc_feature    complement(22098. .22099)
                     /note="SL1 trans-splice site; see EST yk83e5.5"
     misc_feature    22707. .22708
                     /note="SL1 trans-splice site; see EST yk109e2.5"
     gene            22722. .23271
                     /gene="D1007.6"
     CDS             join(22722. .22865,22918. .23104,23153. .23271)
                     /gene="D1007.6"
                     /note="Similar to 40S ribosomal protein S10; coded for by
                     C. elegans cDNA yk109e2.5; coded for by C. elegans cDNA
                     yk99h9.5; coded for by C. elegans cDNA yk98a8.5; coded for
                     by C. elegans cDNA yk80h1.5; coded for by C. elegans cDNA
...
Note: remainder of annotations omitted.

  Query Match          13.3%;  Score 17;  DB 18;  Length 47554;
  Best Local Similarity 100.0%;  Pred. No. 3.26e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  16838 AAGAATAAAGAATATCT 16854
          |||||||||||||||||
Qy    101 AAGAATAAAGAATATCT 117


RESULT   48
LOCUS       AC002453    70319 bp    DNA             PRI       20-AUG-1997
DEFINITION  Human BAC clone GS440B14 from 7q21-q22, complete sequence.
ACCESSION   AC002453
NID         g2337879
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 70319)
  AUTHORS   Wamsley,P and O'Brien,D.
  TITLE     The sequence of H. sapiens BAC clone GS440B14
  JOURNAL   Unpublished (1997)
REFERENCE   2  (bases 1 to 70319)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-AUG-1997) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     SUBMITTED BY:
```



```
            Genome Sequencing Center
            Department of Genetics
            Washington University
            St. Louis MO 63108, USA
            http://genome.wustl.edu/gsc
            mailto:sapiens@watson.wustl.edu

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded or sequenced with an alternate
            chemistry; an attempt was made to resolve all sequencing problems,
            such as compressions and repeats; all regions were covered by
            sequence from more than one subclone; and the assembly was
            confirmed by restriction digest.

            MAPPING INFORMATION:
            The sequence of this clone was established as part of a mapping and
            sequencing collaboration between the NHGRI Chromosome 7 Mapping
            Project (Eric D. Green, Director), John D. McPherson in the
            Department of Genetics (Washington University), and the Washington
            University Genome Sequencing Center.  For additional information
            about the map position of this sequence, see
            http://www.nhgri.nih.gov/DIR/GTB/CHR7 or send
            mailto:egreen@nhgri.nih.gov

            SOURCE INFORMATION:
            This clone is from the first BAC library from Genome Systems, Inc.
            (http://www.genomesystems.com).
            Cell line:  lymphoblastoid
            Haplotypes:  two
            VECTOR:  pBeloBAC
            Selection:  chloramphenicol

            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is RG104I04, 200 bp overlap;  the
            clone sequenced to the right is GS165I04, 200 bp overlap.  Actual
            start of this clone is at base position 123004 of RG104I04; actual
            end is at 70125 of GS440B14.

            This clone contains STS's sWSS2937 (NID:g1113649), sWSS2945
            (NID:g1113653), and sWSS1091 (NID:g394026).
FEATURES              Location/Qualifiers
     source           1. .70319
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="7"
                      /clone="GS440B14"
```



```
                      /clone_lib="GSBAC1"
                      /map="7q21-q22"
     repeat_region    2234. .2526
                      /rpt_family="ALU"
     misc_feature     5945. .6232
                      /note="match to human EST H68655 (NID:g1027395)
                      yr87h11.s1"
     misc_feature     complement(6236. .6290)
                      /note="match to human EST H68654 (NID:g1027394)
                      yr87h11.r1"
     repeat_region    complement(6714. .7099)
                      /rpt_family="L1"
     repeat_region    7680. .7717
                      /rpt_family="L1"
     misc_feature     complement(12817. .13102)
                      /note="match to human EST H68654 (NID:g1027394)
                      yr87h11.r1"
     repeat_region    13896. .13917
                      /rpt_family="L1"
     misc_feature     complement(16644. .17011)
                      /note="match to human EST H75859 (NID:g1049930)
                      yu60g09.r1"
     repeat_region    21561. .21707
                      /rpt_family="ALU"
     repeat_region    27804. .27839
                      /rpt_family="L1"
     misc_feature     29373. .29741
                      /note="match to human EST AA178987 (NID:g1760539)
                      zp12h04.s1"
     misc_feature     29799. .29918
                      /note="match to human EST AA263087 (NID:g1898885)"
     repeat_region    31805. .31849
                      /rpt_family="L1"
     repeat_region    32356. .32385
                      /rpt_family="L1"
     misc_feature     complement(33143. .33508)
                      /note="match to human EST AA179378 (NID:g1760729)
                      zp12h04.r1"
     repeat_region    complement(34351. .34372)
                      /rpt_family="L1"
     repeat_region    complement(34375. .34671)
                      /rpt_family="ALU"
     repeat_region    complement(35169. .35377)
                      /rpt_family="L1"
     misc_feature     37282. .37348
                      /note="match to human EST AA263087 (NID:g1898885)"
     misc_feature     37531. .37633
                      /note="match to human EST H86687 (NID:g1068266)
                      yt02h06.r1"
     misc_feature     37667. .37808
                      /note="match to human EST H86687 (NID:g1068266)
```

```
                    yt02h06.r1"
     misc_feature   complement(39319. .39486)
                    /note="match to human EST H85981 (NID:g1067560)
                    yt02h06.s1"
     repeat_region  40845. .40869
                    /rpt_family="L1"
     repeat_region  complement(47642. .47669)
                    /rpt_family="L1"
     repeat_region  47817. .49101
                    /rpt_family="L1"
     misc_feature   49833. .50232
                    /note="match to human EST H85294 (NID:g1064235)
                    yv86f06.r1"
     misc_feature   complement(51119. .51329)
                    /note="match to human EST H84830 (NID:g1064180)
                    yv86f06.s1"
     repeat_region  complement(51694. .51982)
                    /rpt_family="ALU"
     repeat_region  52432. .52948
                    /rpt_family="MER"
     repeat_region  54041. .54065
                    /rpt_family="L1"
     repeat_region  complement(58087. .58118)
                    /rpt_family="L1"
     repeat_region  complement(58119. .58337)
                    /rpt_family="ALU"
     repeat_region  58811. .59102
                    /rpt_family="ALU"
     repeat_region  complement(62988. .63281)
                    /rpt_family="ALU"
     repeat_region  complement(65701. .66180)
                    /rpt_family="L1"
     repeat_region  complement(66189. .66286)
                    /rpt_family="L1"
     repeat_region  complement(66428. .66457)
                    /rpt_family="L1"
     repeat_region  complement(68062. .68145)
                    /rpt_family="L1"
     repeat_region  complement(68209. .68241)
                    /rpt_family="L1"
     repeat_region  complement(69895. .69933)
                    /rpt_family="L1"
BASE COUNT    21858 a  11360 c  13552 g  23549 t
ORIGIN

  Query Match          13.3%;  Score 17;  DB 26;  Length 70319;
  Best Local Similarity 100.0%;  Pred. No. 3.26e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  30484 AAATTCTGCTTTAAAAA 30500
          |||||||||||||||||
```



```
Cp     25 AAATTCTGCTTTAAAAA 9


RESULT   49
LOCUS       AC005348    84132 bp    DNA             PRI       31-JUL-1998
DEFINITION  Homo sapiens chromosome 5, BAC clone 118L13 (LBNL H176), complete
            sequence.
ACCESSION   AC005348
NID         g3366564
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 84132)
  AUTHORS   Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
            Kadner,K., Miguel,T., Miller,C., Pitluck,S., Pollard,M.,
            Rojeski,H., Subramanian,S. and Martin,C.H.
  TITLE     Sequencing of human chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 84132)
  AUTHORS   Ricke,D.O.
  TITLE     Large Scale Sequence Analysis and Annotation with the Sequence
            Comparison Analysis (SCAN) System
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 84132)
  AUTHORS   Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
            Davis,C.A., Kadner,K., Miguel,T., Pitluck,S., Pollard,M.,
            Rojeski,H., Subramanian,S. and Martin,C.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JUL-1998) Human Genome Center, DOE Joint Genome
            Institute, Lawrence Berkeley National Laboratory, MS 74-157,
            Berkeley, CA 94720, U.S.A.
COMMENT     Sequence submitted by:
            DOE Joint Genome Institute.
FEATURES             Location/Qualifiers
     source          1. .84132
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="5q"
                     /clone="118L13"
                     /chromosome="5"
                     /note="LBNL H176"
     misc_feature    1254. .1417
                     /note="GRAIL 2 excellent exon, frame 0"
     repeat_region   1457. .1479
                     /note="(T)23"
                     /rpt_type=tandem
                     /rpt_unit=T
     repeat_region   complement(1462. .1756)
                     /rpt_family="Alu"
```



```
     repeat_region   2737. .2997
                     /rpt_family="Alu"
     misc_feature    complement(4317. .4456)
                     /note="97% identity EST99753"
                     /db_xref="dbEST:AA385945"
     repeat_region   4490. .4514
                     /note="(T)25"
                     /rpt_type=tandem
                     /rpt_unit=T
     repeat_region   complement(4493. .4781)
                     /rpt_family="Alu"
     repeat_region   10255. .10492
                     /rpt_family="Alu"
     repeat_region   10344. .10364
                     /note="(A)21"
                     /rpt_type=tandem
                     /rpt_unit=A
     repeat_region   10537. .10818
                     /rpt_family="Alu"
     misc_feature    12104. .12231
                     /note="GRAIL 2 excellent exon, frame 1"
     misc_feature    12513. .12666
                     /note="GRAIL 2 excellent exon, frame 1"
     repeat_region   complement(13850. .15341)
                     /rpt_family="L1"
     repeat_region   complement(14614. .14911)
                     /rpt_family="Alu"
     repeat_region   complement(15386. .15508)
                     /rpt_family="L1"
     repeat_region   complement(15591. .15871)
                     /rpt_family="Alu"
     misc_feature    complement(16865. .16948)
                     /note="GRAIL 2 excellent exon, frame 0"
     misc_feature    17559. .17802
                     /note="98% identity B90029 BAC end sequence"
     misc_feature    complement(17991. .18158)
                     /note="GRAIL 2 excellent exon, frame 2"
     repeat_region   complement(18003. .18178)
                     /rpt_family="LINE2"
     repeat_region   complement(18358. .18631)
                     /rpt_family="Alu"
     repeat_region   complement(20010. .20106)
                     /rpt_family="MIR"
     repeat_region   20159. .20415
                     /rpt_family="MER7"
     repeat_region   20421. .20493
                     /rpt_family="MER7"
     repeat_region   complement(20922. .21053)
                     /rpt_family="MIR"
     misc_feature    21756. .21794
                     /note="GRAIL 2 excellent exon, frame 1"
```



```
     misc_feature    join(28363. .28668,28801. .28886)
                     /standard_name="ah46e07"
                     /note="99% & 96% identity dbEST:AA719053"
     misc_feature    complement(join(28776. .28886,30438. .30498,31997. .32101,
                     44632. .44710,49262. .49330))
                     /standard_name="rhoGap protein"
                     /note="78-96% identity Gallus gallus U36309"
     misc_feature    complement(28777. .28882)
                     /note="GRAIL 2 excellent exon, frame 1"
     repeat_region   30184. .30207
                     /note="(A)24"
                     /rpt_type=tandem
                     /rpt_unit=A
     misc_feature    complement(30438. .30496)
                     /note="GRAIL 2 excellent exon, frame 0"
     misc_feature    complement(31998. .32085)
                     /note="GRAIL 2 excellent exon, frame 2"
     repeat_region   complement(34037. .34329)
                     /rpt_family="Alu"
     misc_feature    37238. .37586
                     /note="97% identity yj36e05.s1"
                     /db_xref="dbEST:H02575"
     repeat_region   complement(37594. .37879)
                     /rpt_family="Alu"
     repeat_region   37949. .38211
                     /rpt_family="Alu"
     repeat_region   38383. .38710
                     /rpt_family="Alu"
     repeat_region   complement(40212. .40803)
                     /rpt_family="Alu"
     repeat_region   47029. .47285
                     /rpt_family="Alu"
     repeat_region   47553. .47827
                     /rpt_family="Alu"
     repeat_region   47810. .47841
                     /note="(A)32"
                     /rpt_type=tandem
                     /rpt_unit=A
     repeat_region   complement(49484. .49628)
                     /rpt_family="MIR"
     misc_feature    50523. .50556
                     /note="GRAIL 2 excellent exon, frame 2"
     repeat_region   50568. .50589
                     /note="(T)22"
                     /rpt_type=tandem
                     /rpt_unit=T
     repeat_region   complement(50572. .50863)
                     /rpt_family="Alu"
     repeat_region   complement(50961. .51266)
                     /rpt_family="L1"
     repeat_region   complement(52103. .52536)
```

```
                     /rpt_family="L1"
     repeat_region   54677. .54776
                     /rpt_family="Alu"
     repeat_region   56827. .57108
                     /rpt_family="Alu"
     repeat_region   57091. .57119
                     /note="(A)29"
                     /rpt_type=tandem
                     /rpt_unit=A
     misc_feature    complement(58310. .58487)
                     /note="GRAIL 2 excellent exon, frame 2"
     repeat_region   58320. .58346
                     /note="(ATC)9"
                     /rpt_type=tandem
                     /rpt_unit=ATC
     repeat_region   59041. .59344
                     /rpt_family="Alu"
     repeat_region   complement(61440. .61655)
                     /rpt_family="MIR"
     repeat_region   61774. .61902
                     /rpt_family="MIR"
     misc_feature    complement(63990. .64049)
                     /note="GRAIL 2 excellent exon, frame 2"
     repeat_region   64151. .64421
                     /rpt_family="Alu"
     repeat_region   complement(64846. .64892)
                     /rpt_family="MIR"
     repeat_region   65585. .65874
                     /rpt_family="Alu"
     repeat_region   65857. .65879
                     /note="(A)23"
                     /rpt_type=tandem
                     /rpt_unit=A
     repeat_region   66275. .66342
                     /rpt_family="Alu"
     repeat_region   68308. .68556
                     /rpt_family="L1"
     repeat_region   complement(69791. .70079)
                     /rpt_family="Alu"
     misc_feature    71612. .72222
                     /note="99% identity B71939 BAC end sequence"
     repeat_region   complement(72766. .72915)
                     /rpt_family="MIR"
     repeat_region   73526. .73859
                     /rpt_family="Alu"
     repeat_region   73934. .73975
                     /note="(AT)21"
                     /rpt_type=tandem
                     /rpt_unit=AT
     misc_feature    74393. .74523
                     /note="GRAIL 2 excellent exon, frame 1"
```



```
...
Note: remainder of annotations omitted.

  Query Match           13.3%;  Score 17;  DB 26;  Length 84132;
  Best Local Similarity 100.0%;  Pred. No. 3.26e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  65245 AATTTTTAAAGCAGAAT 65261
          |||||||||||||||||
Qy      7 AATTTTTAAAGCAGAAT 23


RESULT   50
LOCUS       AB005237    87835 bp    DNA             PLN       11-DEC-1997
DEFINITION  Arabidopsis thaliana genomic DNA, chromosome 5, P1 clone: MJJ3,
            complete sequence.
ACCESSION   AB005237
NID         g2264309
KEYWORDS    HTG.
SOURCE      Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui P1
            clone:MJJ3.
  ORGANISM  Arabidopsis thaliana
            Eukaryotae; Viridiplantae; Charophyta/Embryophyta group;
            Embryophyta; Tracheophyta; seed plants; Magnoliophyta;
            Magnoliopsida; Capparales; Brassicaceae; Arabidopsis.
REFERENCE   1  (sites)
  AUTHORS   Sato,S., Kotani,H., Nakamura,Y., Kaneko,T., Asamizu,E., Fukami,M.,
            Miyajima,N. and Tabata,S.
  TITLE     Structural analysis of Arabidopsis thaliana chromosome 5. I.
            Sequence features of the 1.6 Mb regions covered by twenty
            physically assigned P1 clones
  JOURNAL   DNA Res. 4 (3), 215-230 (1997)
  MEDLINE   97471969
REFERENCE   2  (bases 1 to 87835)
  AUTHORS   Nakamura,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUL-1997) to the DDBJ/EMBL/GenBank databases.
            Yasukazu Nakamura, Kazusa DNA Research Institute, Laboratory of
            Gene Structure 2; 1532-3, Yana, Kisarazu, Chiba 292, Japan
            (E-mail:ynakamu@kazusa.or.jp, Tel:+81-438-52-3935,
            Fax:+81-438-52-3934)
FEATURES             Location/Qualifiers
     source          1. .87835
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="5"
                     /clone="MJJ3"
                     /clone_lib="Mitsui P1"
BASE COUNT    28070 a  16447 c  15945 g  27373 t
ORIGIN
```



```
  Query Match           13.3%;  Score 17;  DB 23;  Length 87835;
  Best Local Similarity 100.0%;  Pred. No. 3.26e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  25063 CCTTTTACACTTTTCAG 25079
          |||||||||||||||||
Qy     54 CCTTTTACACTTTTCAG 70


Search completed: Tue Dec 22 20:58:33 1998
Job time : 472 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Tue Dec 22 21:35:36 1998; MasPar time 32.08 Seconds
508.929 Million cell updates/sec

Tabular output not generated.

Title: >US-08-530-112A-2058
Description: (1-120) from US08530112A.seq
Perfect Score: 119
N.A. Sequence: 1 GATCTTTACTCTTGAGAAAT.....AAAGCTCCNTAAAAAGCAAA 120
Comp: CTAGAAATGAGAACTCTTTA.....TTTCGAGGNATTTTTCGTTT

Scoring table: TABLE jmetric
Gap 60

Nmatch STD : Dbase 0; Query 0

Searched: 188442 seqs, 68026449 bases x 2

Post-processing: Minimum Match 0%
Listing first 1000 summaries

Database: n-geneseq32
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39 40:part40

Statistics: Mean 6.081; Variance 2.587; scale 2.350

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
	1	119	100.0	120	18	T21058	Human gene signature	3.34e-94
	2	118	99.2	7849	16	Q94109	hML genomic DNA.	3.13e-93
	3	70	58.8	1721	17	Q99554	Thrombopoietin coding	3.11e-47
	4	41	34.5	1795	16	Q94107	hML cDNA.	1.08e-20
	5	41	34.5	1795	28	T64318	Human thrombopoietin	1.08e-20
c	6	25	21.0	25	23	T26842	Human GS HUMGS02342-b	3.59e-07
c	7	25	21.0	37	23	T26844	Human mpl-ligand gene	3.59e-07
c	8	25	21.0	37	23	T26845	Human mpl-ligand gene	3.59e-07
c	9	15	12.6	1350	7	Q43202	Sequence encoding del	7.94e+00
	10	15	12.6	1728	12	Q73481	Low molecular weight	7.94e+00
	11	15	12.6	2000	3	Q21400	Vibrolysin gene seque	7.94e+00
	12	15	12.6	2266	1	N93079	Sequence encoding pro	7.94e+00
	13	15	12.6	2298	1	N80981	Sequence of the singl	7.94e+00
	14	15	12.6	2299	2	N60703	Sequence encoding hum	7.94e+00
	15	15	12.6	2301	2	N60591	Sequence encoding mod	7.94e+00
	16	15	12.6	2301	2	N70390	cDNA encoding human p	7.94e+00
	17	15	12.6	2301	1	N91075	DNA encoding natural	7.94e+00
	18	15	12.6	2303	1	N91740	Sequence of prourokin	7.94e+00
	19	15	12.6	2303	12	Q73483	Full length human uro	7.94e+00
	20	15	12.6	2304	2	N71368	Modified prourokinase	7.94e+00
	21	15	12.6	2304	2	N71330	Sequence encoding mod	7.94e+00
	22	15	12.6	2304	4	N30030	Sequence encoding ful	7.94e+00
	23	15	12.6	2304	2	N71338	Modified prourokinase	7.94e+00
	24	15	12.6	2304	2	N71369	Modified prourokinase	7.94e+00
	25	15	12.6	2377	11	Q68538	Extracellular neutral	7.94e+00
	26	15	12.6	2377	3	Q20360	Human pro-urokinase c	7.94e+00
	27	15	12.6	2377	1	N90726	DNA sequence of extra	7.94e+00
	28	15	12.6	2427	1	Q04107	Human pro-urokinase c	7.94e+00
	29	15	12.6	2919	30	T51124	Human homeoprotein re	7.94e+00
c	30	15	12.6	12412	35	T33537	BCG deletion region 3	7.94e+00
c	31	14	11.8	18	24	T44575	Antisense oligo HUPA-	3.46e+01
c	32	14	11.8	32	18	T02529	PenI gene upstream pr	3.46e+01
c	33	14	11.8	35	21	T34332	3' glaA non-coding se	3.46e+01
	34	14	11.8	62	21	T26260	Human gene signature	3.46e+01
c	35	14	11.8	80	5	Q34036	Downstream sequence o	3.46e+01
c	36	14	11.8	129	18	T21050	Human gene signature	3.46e+01
	37	14	11.8	216	19	T07068	Immunogen DNA from n-	3.46e+01
	38	14	11.8	281	8	Q59885	Human brain Expressed	3.46e+01
	39	14	11.8	291	18	T19568	Human gene signature	3.46e+01
c	40	14	11.8	522	3	Q14619	Fragment S3 from huma	3.46e+01
c	41	14	11.8	528	11	Q64172	Mouse PAP.	3.46e+01
	42	14	11.8	593	31	T43460	ATM gene exon 20.	3.46e+01

c	43	14	11.8	797	11	Q69201	Human Pancreatitis-As	3.46e+01
c	44	14	11.8	798	3	Q14623	Fragment S4 from huma	3.46e+01
	45	14	11.8	1044	22	T29435	Mouse pancreas beta-c	3.46e+01
	46	14	11.8	1044	21	T13905	Mouse pancreas G-prot	3.46e+01
	47	14	11.8	1293	27	T48100	Immunogenic type F bo	3.46e+01
	48	14	11.8	1399	19	T14038	Transcription factor-	3.46e+01
c	49	14	11.8	1420	35	T72167	Alzheimer's disease D	3.46e+01
c	50	14	11.8	1424	29	T39671	D-aminotransferase co	3.46e+01
c	51	14	11.8	1838	1	N90769	cDNA of endothelial c	3.46e+01
c	52	14	11.8	1929	35	T72642	Human testis-specific	3.46e+01
c	53	14	11.8	2136	2	N60755	Sequence encoding hum	3.46e+01
	54	14	11.8	2490	37	T91358	Orf virus strain NZ-7	3.46e+01
c	55	14	11.8	2494	17	T00617	Megakaryocyte kinase	3.46e+01
c	56	14	11.8	2975	11	Q65587	Heat resistant aldehy	3.46e+01
	57	14	11.8	3320	15	Q91200	H-NUC retinoblastoma	3.46e+01
c	58	14	11.8	3561	39	T93874	Lys-gingipain high mo	3.46e+01
c	59	14	11.8	3561	15	Q88141	Porphyromonas gingiva	3.46e+01
c	60	14	11.8	6241	21	T30655	P. gingivalis haemagg	3.46e+01
c	61	14	11.8	6519	23	T37401	Human DOCK180 cDNA.	3.46e+01
	62	14	11.8	24417	38	T97221	Pseudomonas aeruginos	3.46e+01
	63	14	11.8	28994	39	V15826	Genomic DNA for inter	3.46e+01
	64	13	10.9	25	5	Q34384	Downstream PCR primer	1.42e+02
c	65	13	10.9	25	5	Q34386	Downstream PCR primer	1.42e+02
c	66	13	10.9	25	32	T75959	DEN-2 cloning/sequenc	1.42e+02
	67	13	10.9	30	36	T75746	Mouse OB-R construct	1.42e+02
	68	13	10.9	30	23	T34995	Full-length VIGF 3' P	1.42e+02
	69	13	10.9	36	26	T52670	Human ICAM hammerhead	1.42e+02
	70	13	10.9	36	25	T50469	Rabbit CETP HH ribosy	1.42e+02
	71	13	10.9	55	19	T22109	Human gene signature	1.42e+02
	72	13	10.9	91	21	T24338	Human gene signature	1.42e+02
c	73	13	10.9	101	19	T21071	Human gene signature	1.42e+02
	74	13	10.9	104	5	Q34025	Downstream sequence o	1.42e+02
	75	13	10.9	124	22	T18567	Human chondrocyte RNA	1.42e+02
c	76	13	10.9	153	19	T09779	Tomato genomic DNA, D	1.42e+02
	77	13	10.9	186	10	Q77128	Human genome fragment	1.42e+02
c	78	13	10.9	234	14	Q83712	Human beta-globin reg	1.42e+02
	79	13	10.9	252	10	Q76742	Human genome fragment	1.42e+02
	80	13	10.9	258	10	Q76743	Human genome fragment	1.42e+02
c	81	13	10.9	261	10	Q77484	Human genome fragment	1.42e+02
	82	13	10.9	272	20	T22194	Human gene signature	1.42e+02
c	83	13	10.9	291	18	T19341	Human gene signature	1.42e+02
	84	13	10.9	297	18	T20994	Human gene signature	1.42e+02
c	85	13	10.9	306	30	T67440	H. pylori transmembra	1.42e+02
	86	13	10.9	325	8	Q60687	Human brain Expressed	1.42e+02
c	87	13	10.9	326	8	Q59521	Human brain Expressed	1.42e+02
	88	13	10.9	329	8	Q60934	Human brain Expressed	1.42e+02
c	89	13	10.9	331	5	Q31793	Modified endotoxin bi	1.42e+02
c	90	13	10.9	331	29	T62063	Modified Limulus poly	1.42e+02
c	91	13	10.9	331	30	T68929	Endotoxin binding/neu	1.42e+02
c	92	13	10.9	331	30	T69763	Modified horseshoe cr	1.42e+02
	93	13	10.9	333	20	T23648	Human gene signature	1.42e+02

c	94	13	10.9	333	21	T24709	Human gene signature	1.42e+02
c	95	13	10.9	339	30	T68278	H. pylori putative me	1.42e+02
	96	13	10.9	351	19	T21296	Human gene signature	1.42e+02
c	97	13	10.9	359	8	Q60852	Human brain Expressed	1.42e+02
c	98	13	10.9	377	20	T23745	Human gene signature	1.42e+02
	99	13	10.9	387	8	Q59558	Human brain Expressed	1.42e+02
	100	13	10.9	423	33	T80624	Type II topoisomerase	1.42e+02
c	101	13	10.9	426	8	Q60979	Human brain Expressed	1.42e+02
	102	13	10.9	431	19	T19912	Human gene signature	1.42e+02
c	103	13	10.9	457	19	T19899	Human gene signature	1.42e+02
c	104	13	10.9	619	30	T47865	XheI-SpeI fragment en	1.42e+02
	105	13	10.9	740	8	Q48318	3' portion of Drosoph	1.42e+02
	106	13	10.9	740	39	V12061	Alcohol dehydrogenase	1.42e+02
	107	13	10.9	923	35	T85082	Transforming growth f	1.42e+02
c	108	13	10.9	923	22	T28348	Hepatitis C virus cor	1.42e+02
c	109	13	10.9	952	6	Q36437	Sequence obtained usi	1.42e+02
	110	13	10.9	1018	30	T67944	H. pylori cytoplasmic	1.42e+02
c	111	13	10.9	1019	17	T12409	FMDV cDNA coding for	1.42e+02
c	112	13	10.9	1077	39	T89027	S. pneumoniae peptide	1.42e+02
	113	13	10.9	1113	35	T72632	Mangifera indica ACC	1.42e+02
	114	13	10.9	1185	35	T78781	DNA encoding Aquifex	1.42e+02
	115	13	10.9	1247	3	Q21978	Hybrid mini-vitelloge	1.42e+02
	116	13	10.9	1260	3	Q21976	Hybrid mini-vitelloge	1.42e+02
c	117	13	10.9	1271	23	T34991	Vascular IBP-like gro	1.42e+02
	118	13	10.9	1278	13	Q80198	Klebsiella sp. nitril	1.42e+02
	119	13	10.9	1290	3	Q21980	Hybrid mini-vitelloge	1.42e+02
	120	13	10.9	1290	3	Q21981	Hybrid mini-vitelloge	1.42e+02
	121	13	10.9	1290	3	Q21977	Hybrid mini-vitelloge	1.42e+02
	122	13	10.9	1315	32	T72466	Human CII-3 cDNA.	1.42e+02
c	123	13	10.9	1338	29	T67345	H. pylori cytoplasmic	1.42e+02
c	124	13	10.9	1366	32	T72848	Mouse apoptosis induc	1.42e+02
c	125	13	10.9	1443	23	T27745	Neural thread protein	1.42e+02
c	126	13	10.9	1443	13	Q77869	CNS neural thread pro	1.42e+02
	127	13	10.9	1484	34	T61780	Human MAD2 cDNA.	1.42e+02
	128	13	10.9	1512	35	T72172	Alzheimer's disease D	1.42e+02
c	129	13	10.9	1514	8	Q45549	Sequence of gene Y en	1.42e+02
c	130	13	10.9	1556	16	Q95035	Rat hippocampal neuro	1.42e+02
	131	13	10.9	1634	7	Q45653	Human somatostatin re	1.42e+02
c	132	13	10.9	1710	30	T68597	Md-alpha-E7 gene.	1.42e+02
c	133	13	10.9	1782	30	T68273	H. pylori cytoplasmic	1.42e+02
c	134	13	10.9	1785	5	Q28437	Full length islet GAD	1.42e+02
c	135	13	10.9	1785	5	Q28436	Full length brain GAD	1.42e+02
c	136	13	10.9	1827	39	V15514	Hamster oral papillom	1.42e+02
c	137	13	10.9	2015	2	N60107	Sequence encoding pre	1.42e+02
c	138	13	10.9	2022	36	T96728	Human adducin-like 70	1.42e+02
c	139	13	10.9	2055	4	N30110	Sequence encoding pre	1.42e+02
	140	13	10.9	2059	2	N70271	DNA fragment for tran	1.42e+02
c	141	13	10.9	2064	30	T66701	Human serum albumin H	1.42e+02
c	142	13	10.9	2064	3	N92583	Sequence of human ser	1.42e+02
	143	13	10.9	2071	36	T77131	Human RES cDNA.	1.42e+02
c	144	13	10.9	2082	14	Q55135	Staphylococcus aureus	1.42e+02

c	145	13	10.9	2150	35	T91037	Yeast checkpoint cont 1.42e+02
c	146	13	10.9	2251	4	N30050	Sequence encoding hum 1.42e+02
c	147	13	10.9	2258	2	N60078	Recombinant human ser 1.42e+02
	148	13	10.9	2412	6	Q32164	BT toxin 63B 1.42e+02
	149	13	10.9	2412	5	Q31413	Bacillus thuringiensi 1.42e+02
	150	13	10.9	2412	5	Q30940	Encodes BT toxin 63B. 1.42e+02
	151	13	10.9	2412	29	T60072	63B toxin coding sequ 1.42e+02
	152	13	10.9	2414	17	T04895	Rat SCF genomic DNA e 1.42e+02
c	153	13	10.9	2468	4	Q24272	DNA encoding GL-7ACA 1.42e+02
	154	13	10.9	2580	39	V12083	Murine MHC class II i 1.42e+02
	155	13	10.9	2706	37	T92835	Flea saliva protein f 1.42e+02
c	156	13	10.9	2834	28	T61896	Human 14-3-3 zeta or 1.42e+02
	157	13	10.9	2900	9	Q50429	Bovine MTP. 1.42e+02
c	158	13	10.9	2915	3	Q23261	Carrot aspartokinase- 1.42e+02
c	159	13	10.9	2920	36	T96727	Human adducin-like 70 1.42e+02
c	160	13	10.9	3048	40	V15294	Murine AGP-1 encoding 1.42e+02
c	161	13	10.9	3073	3	Q14635	Clone associated with 1.42e+02
c	162	13	10.9	3073	21	T34371	Plasmid pATG29 (ATCC 1.42e+02
	163	13	10.9	3083	2	Q11851	Glutamate receptor 3. 1.42e+02
c	164	13	10.9	3088	11	Q66768	Tomato ADH2 gene. 1.42e+02
	165	13	10.9	3136	2	Q13351	C3 vegetable PEPC gen 1.42e+02
	166	13	10.9	3158	6	Q31880	Cyclin D3 promoter. 1.42e+02
	167	13	10.9	3158	9	Q53212	Human cyclin D3 promo 1.42e+02
	168	13	10.9	3294	8	Q48939	Steroid hormone recep 1.42e+02
	169	13	10.9	3476	37	T95698	Novel human gene, des 1.42e+02
c	170	13	10.9	3500	32	T72782	Mouse interleukin-1 r 1.42e+02
c	171	13	10.9	3519	35	T79251	Mouse beta 2 integrin 1.42e+02
c	172	13	10.9	3519	15	Q91713	Mouse alpha-d subunit 1.42e+02
c	173	13	10.9	3528	35	T79243	Rat beta 2 integrin a 1.42e+02
c	174	13	10.9	3528	15	Q91740	Rat alpha-d composite 1.42e+02
c	175	13	10.9	3597	35	T79257	Rat beta 2 integrin a 1.42e+02
c	176	13	10.9	3597	15	Q91715	Rat alpha-d subunit c 1.42e+02
c	177	13	10.9	3607	31	T43498	ATM gene 3' untransla 1.42e+02
c	178	13	10.9	3607	37	T45420	3' UTR of the ataxia- 1.42e+02
c	179	13	10.9	3642	38	T84934	Human prostate protei 1.42e+02
c	180	13	10.9	3803	15	Q91714	Mouse alpha-d subunit 1.42e+02
c	181	13	10.9	3803	35	T79256	Mouse beta 2 integrin 1.42e+02
	182	13	10.9	4074	38	T73998	B. thuringiensis VIP2 1.42e+02
	183	13	10.9	4074	20	T13950	B. thuringiensis VIP2 1.42e+02
	184	13	10.9	4074	39	V16171	DNA sequence containi 1.42e+02
	185	13	10.9	4328	5	Q28758	Partial sequence of t 1.42e+02
	186	13	10.9	4467	2	N92428	Sequence encoding amy 1.42e+02
	187	13	10.9	4713	39	V12067	Murine IAd alpha chai 1.42e+02
	188	13	10.9	4724	39	V12068	Murine IAd beta chain 1.42e+02
	189	13	10.9	4854	2	Q11538	Rat Stem Cell Factor 1.42e+02
	190	13	10.9	5211	13	Q77789	Pre-pro-cobra C3 codi 1.42e+02
c	191	13	10.9	5394	17	T00872	Murine mC26 structura 1.42e+02
c	192	13	10.9	7647	39	V15519	Hamster oral papillom 1.42e+02
	193	13	10.9	7653	18	T00770	Sirolimus effector pr 1.42e+02
	194	13	10.9	7824	23	T33872	Human RAPT1 cDNA. 1.42e+02
	195	13	10.9	8387	3	Q22489	Human U266 lambda gen 1.42e+02

c	196	13	10.9	8957	17	Q81790	HTLV-II virus NRA str 1.42e+02
	197	13	10.9	8967	21	T31031	Factor-VIII full-leng 1.42e+02
	198	13	10.9	8975	3	N50054	Human factor VIII gen 1.42e+02
	199	13	10.9	9009	33	T61548	Factor VIII:C (Arg168 1.42e+02
	200	13	10.9	9009	40	V12112	Homo sapiens factor V 1.42e+02
	201	13	10.9	9009	9	Q50185	Human Factor VIII cDN 1.42e+02
	202	13	10.9	9071	3	Q23370	U266-Lambda gene and 1.42e+02
	203	13	10.9	9071	3	Q22491	Human U266 lambda gen 1.42e+02
	204	13	10.9	10723	2	Q12787	Dengue 2 virus genome 1.42e+02
	205	13	10.9	10723	31	T49304	cDNA encoding polypro 1.42e+02
	206	13	10.9	10723	31	T49303	cDNA sequence encodin 1.42e+02
c	207	13	10.9	12778	31	T68784	ATM mutant G9170C cod 1.42e+02
c	208	13	10.9	13585	17	T11549	Tumour rejection anti 1.42e+02
c	209	13	10.9	19012	6	Q36952	HSA gene. 1.42e+02
c	210	13	10.9	40875	33	T80043	Insert from cosmid 10 1.42e+02
	211	13	10.9	134525	1	Q04525	Total base sequence o 1.42e+02
c	212	12	10.1	20	31	T73214	PCR primer correspond 5.43e+02
	213	12	10.1	21	2	Q10841	Murine kappa light ch 5.43e+02
	214	12	10.1	23	39	V05013	PCR primer 1 used to 5.43e+02
c	215	12	10.1	24	34	T85739	FMR2 gene intron 5-ex 5.43e+02
c	216	12	10.1	27	15	Q88388	CFTR amplification pr 5.43e+02
	217	12	10.1	33	4	Q27343	Oligomer A for clonin 5.43e+02
c	218	12	10.1	34	40	V10634	A. thaliana PDF1.1 PC 5.43e+02
c	219	12	10.1	35	37	T96361	Primer 78 for B7-1 cD 5.43e+02
	220	12	10.1	36	26	T56520	Mouse TNF-alpha hamme 5.43e+02
	221	12	10.1	36	27	T57657	RSV N hammerhead ribo 5.43e+02
	222	12	10.1	36	25	T50082	Human CETP HH ribozym 5.43e+02
c	223	12	10.1	36	40	V10641	A. thaliana PDF1.2 pr 5.43e+02
	224	12	10.1	39	2	Q11440	Probe #2 complementar 5.43e+02
c	225	12	10.1	39	39	V12103	Human B7.1 cDNA 3' pr 5.43e+02
c	226	12	10.1	40	7	Q42648	PCR primer p53-77 to 5.43e+02
	227	12	10.1	42	21	T34340	3' glaA noncoding pri 5.43e+02
c	228	12	10.1	49	13	Q75472	HPV primer probe HPV1 5.43e+02
c	229	12	10.1	58	9	Q54762	H9A M-CSF coding regi 5.43e+02
c	230	12	10.1	58	9	Q54761	H9A M-CSF coding regi 5.43e+02
c	231	12	10.1	58	9	Q54760	H9A, H15A M-CSF codin 5.43e+02
c	232	12	10.1	72	9	Q54763	Q20A, V78K M-CSF mute 5.43e+02
	233	12	10.1	81	19	T19805	Human gene signature 5.43e+02
c	234	12	10.1	96	18	T00513	Fragment J99 amplific 5.43e+02
c	235	12	10.1	100	12	Q71185	NA2-2 synthetic signa 5.43e+02
c	236	12	10.1	108	21	T23364	Human gene signature 5.43e+02
c	237	12	10.1	119	22	T32383	HIV packaging signal 5.43e+02
c	238	12	10.1	128	19	T21207	Human gene signature 5.43e+02
	239	12	10.1	129	22	T25792	Human gene signature 5.43e+02
	240	12	10.1	139	10	Q77030	Human genome fragment 5.43e+02
c	241	12	10.1	150	19	T23861	Human gene signature 5.43e+02
	242	12	10.1	157	18	T19317	Human gene signature 5.43e+02
	243	12	10.1	159	10	Q77567	Human genome fragment 5.43e+02
	244	12	10.1	161	10	Q77032	Human genome fragment 5.43e+02
	245	12	10.1	162	33	T85631	BAN promoter. 5.43e+02
c	246	12	10.1	175	20	T21437	Human gene signature 5.43e+02

	247	12	10.1	183	5	Q34066	Downstream sequence o 5.43e+02
c	248	12	10.1	185	22	T32384	HIV packaging signal 5.43e+02
c	249	12	10.1	188	10	Q76946	Human genome fragment 5.43e+02
	250	12	10.1	196	18	T19257	Human gene signature 5.43e+02
c	251	12	10.1	200	15	Q93941	Human MSH2 gene exon 5.43e+02
	252	12	10.1	214	19	T21395	Human gene signature 5.43e+02
	253	12	10.1	218	22	T24222	Human gene signature 5.43e+02
	254	12	10.1	228	19	T22706	Human gene signature 5.43e+02
c	255	12	10.1	237	30	T67600	H. pylori surface mem 5.43e+02
c	256	12	10.1	243	19	T09746	Tomato genomic DNA, D 5.43e+02
c	257	12	10.1	256	20	T21672	Human gene signature 5.43e+02
c	258	12	10.1	263	11	Q77248	Human genome fragment 5.43e+02
	259	12	10.1	268	19	T20112	Human gene signature 5.43e+02
	260	12	10.1	273	18	T19188	Human gene signature 5.43e+02
c	261	12	10.1	276	22	T25478	Human gene signature 5.43e+02
c	262	12	10.1	281	4	Q26982	HCV gene 2. 5.43e+02
c	263	12	10.1	284	3	Q20627	C10-22 DNA fragment e 5.43e+02
	264	12	10.1	293	10	Q77581	Human genome fragment 5.43e+02
c	265	12	10.1	299	10	Q77417	Human genome fragment 5.43e+02
	266	12	10.1	309	19	T09810	Tomato genomic DNA, D 5.43e+02
c	267	12	10.1	317	20	T22910	Human gene signature 5.43e+02
c	268	12	10.1	320	8	Q59498	Human brain Expressed 5.43e+02
c	269	12	10.1	321	19	T21641	Human gene signature 5.43e+02
c	270	12	10.1	321	4	Q27004	HK2. 5.43e+02
	271	12	10.1	324	6	Q37078	T cell receptor alpha 5.43e+02
	272	12	10.1	325	8	Q60169	Human brain Expressed 5.43e+02
	273	12	10.1	325	8	Q59675	Human brain Expressed 5.43e+02
	274	12	10.1	325	20	T21871	Human gene signature 5.43e+02
	275	12	10.1	326	18	T19134	Human gene signature 5.43e+02
c	276	12	10.1	328	8	Q59954	Human brain Expressed 5.43e+02
	277	12	10.1	333	22	T26740	Human gene signature 5.43e+02
c	278	12	10.1	335	22	T25317	Human gene signature 5.43e+02
	279	12	10.1	337	8	Q59928	Human brain Expressed 5.43e+02
c	280	12	10.1	338	8	Q59490	Human brain Expressed 5.43e+02
	281	12	10.1	345	8	Q60578	Human brain Expressed 5.43e+02
	282	12	10.1	360	8	Q59228	Human brain Expressed 5.43e+02
	283	12	10.1	360	6	Q39816	Expressed Sequence Ta 5.43e+02
	284	12	10.1	367	8	Q59768	Human brain Expressed 5.43e+02
c	285	12	10.1	369	8	Q61270	Human brain Expressed 5.43e+02
	286	12	10.1	370	38	T93070	Streptococcus pneumon 5.43e+02
	287	12	10.1	376	22	T26635	Human gene signature 5.43e+02
	288	12	10.1	377	8	Q61026	Human brain Expressed 5.43e+02
c	289	12	10.1	377	8	Q60706	Human brain Expressed 5.43e+02
	290	12	10.1	388	16	T04605	CYP19 gene exon I.3 e 5.43e+02
c	291	12	10.1	395	8	Q60531	Human brain Expressed 5.43e+02
c	292	12	10.1	396	33	T80713	Type II topoisomerase 5.43e+02
	293	12	10.1	406	8	Q60024	Human brain Expressed 5.43e+02
c	294	12	10.1	421	31	T43488	ATM gene exon 56. 5.43e+02
	295	12	10.1	424	19	T19917	Human gene signature 5.43e+02
	296	12	10.1	449	16	Q86304	Wilson disease gene e 5.43e+02
	297	12	10.1	460	38	V06108	Viral infection gene 5.43e+02

	298	12	10.1	537	3	Q22579	Sequence of the NH2 r 5.43e+02
c	299	12	10.1	548	38	T95760	Arabidopsis SCARECROW 5.43e+02
	300	12	10.1	576	4	N20046	Bacillus amyloliquefa 5.43e+02
c	301	12	10.1	600	13	Q73726	Retrotransposon probe 5.43e+02
	302	12	10.1	627	1	Q05347	Sequence encoding a p 5.43e+02
c	303	12	10.1	627	29	T40734	DNA encoding HMbk-D a 5.43e+02
	304	12	10.1	639	4	N10027	Sequence which encode 5.43e+02
	305	12	10.1	639	4	N30132	Sequence of VP3 gene 5.43e+02
c	306	12	10.1	654	29	T47675	Human G protein gamma 5.43e+02
	307	12	10.1	663	4	N30133	Sequence of VP3 gene 5.43e+02
	308	12	10.1	680	2	Q12747	5' end of B.burgdorfe 5.43e+02
	309	12	10.1	708	4	N30150	Sequence encoding the 5.43e+02
	310	12	10.1	711	3	N40075	Sequence encoding foo 5.43e+02
c	311	12	10.1	732	30	T68018	H. pylori putative me 5.43e+02
	312	12	10.1	737	1	N90400	Bases 1469-2206 of VP 5.43e+02
	313	12	10.1	743	35	T75597	Canine herpesvirus ge 5.43e+02
c	314	12	10.1	744	16	Q86301	Wilson disease gene e 5.43e+02
	315	12	10.1	774	28	T51058	Human criptin growth 5.43e+02
c	316	12	10.1	789	37	V03609	Bovine BMP-11 coding 5.43e+02
c	317	12	10.1	789	33	T74190	Bovine bone morphogen 5.43e+02
c	318	12	10.1	789	13	Q79444	Partial bovine bone m 5.43e+02
	319	12	10.1	792	18	T15757	Murine interleukin-15 5.43e+02
c	320	12	10.1	800	30	T45981	Human cocaine and amp 5.43e+02
c	321	12	10.1	815	37	T88106	Streptococcus pneumon 5.43e+02
	322	12	10.1	822	20	Q90730	B31/K48/B31/K48 outer 5.43e+02
	323	12	10.1	822	20	Q90729	K48/B31/K48 outer sur 5.43e+02
	324	12	10.1	825	20	Q90725	K48/PGau outer surfac 5.43e+02
	325	12	10.1	825	20	Q90710	B. burgdorferi strain 5.43e+02
	326	12	10.1	825	20	Q90762	PTrob/PBo outer surfa 5.43e+02
	327	12	10.1	825	6	Q38088	ZQ1 Osp A DNA. 5.43e+02
	328	12	10.1	833	4	Q25034	Bovine RSV strain FS- 5.43e+02
c	329	12	10.1	837	20	T14627	Human R33 platelet de 5.43e+02
c	330	12	10.1	840	30	T45980	Rat cocaine and amphe 5.43e+02
	331	12	10.1	861	37	T91798	Dihydropyrimidine deh 5.43e+02
	332	12	10.1	894	26	T40849	Serine protease nfSP7 5.43e+02
	333	12	10.1	912	28	T42082	Human mature LAMP cod 5.43e+02
	334	12	10.1	913	12	Q70048	Dirofilaria immitis p 5.43e+02
	335	12	10.1	937	3	Q22700	Sequence encoding Pla 5.43e+02
	336	12	10.1	939	30	T67457	H. pylori cytoplasmic 5.43e+02
c	337	12	10.1	940	36	T85879	Malassezia fungus MF- 5.43e+02
	338	12	10.1	954	2	N60845	Plasmid pImo336 seque 5.43e+02
	339	12	10.1	957	7	Q40693	B. burgdorferi strain 5.43e+02
	340	12	10.1	961	4	N10025	Sequence of part of t 5.43e+02
	341	12	10.1	972	40	V24793	H. pylori ORF 11gel03 5.43e+02
c	342	12	10.1	972	40	V24793	H. pylori ORF 11gel03 5.43e+02
	343	12	10.1	977	28	T42079	Human LAMP residues 8 5.43e+02
c	344	12	10.1	983	28	T40742	Xylanase coding seque 5.43e+02
c	345	12	10.1	983	25	T43010	Endo-1,4-beta-D-xylan 5.43e+02
	346	12	10.1	1003	30	T40041	Presenilin-1 exon 12. 5.43e+02
c	347	12	10.1	1008	35	T85901	Malassezia fungus MF- 5.43e+02
c	348	12	10.1	1014	16	Q94113	pML2 genomic DNA. 5.43e+02

	349	12	10.1	1017	6	Q38092	NS1-ACA1 fusion codin 5.43e+02
	350	12	10.1	1020	6	Q38090	NS1-ZQ1 fusion coding 5.43e+02
c	351	12	10.1	1026	16	Q94112	pML genomic DNA. 5.43e+02
	352	12	10.1	1029	11	Q65614	Bovine zona pellucida 5.43e+02
c	353	12	10.1	1038	29	T60096	Serpin proteinase inh 5.43e+02
c	354	12	10.1	1044	35	T64686	cDNA for form 2 of pr 5.43e+02
c	355	12	10.1	1044	18	T06881	Prostate/colon tumour 5.43e+02
c	356	12	10.1	1047	18	T06880	Prostate/colon tumour 5.43e+02
c	357	12	10.1	1047	35	T64685	cDNA for form 1 of pr 5.43e+02
c	358	12	10.1	1080	10	Q62344	Influenza type B HA g 5.43e+02
c	359	12	10.1	1080	10	Q62341	Influenza type B HA g 5.43e+02
c	360	12	10.1	1080	10	Q62343	Influenza type B HA g 5.43e+02
c	361	12	10.1	1080	10	Q62345	Influenza type B HA g 5.43e+02
c	362	12	10.1	1080	10	Q62346	Influenza type B HA g 5.43e+02
	363	12	10.1	1107	37	T73178	Mycoplasma arthritidi 5.43e+02
	364	12	10.1	1107	32	T74095	Mycoplasma arthritidi 5.43e+02
	365	12	10.1	1137	29	T42072	Human nuclear prolife 5.43e+02
	366	12	10.1	1144	33	T61047	cDNA encoding human h 5.43e+02
	367	12	10.1	1152	8	Q46074	Partial sequence of t 5.43e+02
	368	12	10.1	1152	3	Q13697	DNA encoding capsid r 5.43e+02
	369	12	10.1	1170	10	Q55272	Jun s I gene. 5.43e+02
	370	12	10.1	1176	23	T14856	Neisseria meningitidi 5.43e+02
c	371	12	10.1	1186	5	Q29184	RuBisCo promoted huSO 5.43e+02
	372	12	10.1	1193	25	T47220	Desulfurococcus strai 5.43e+02
c	373	12	10.1	1197	15	Q89841	Human death associate 5.43e+02
c	374	12	10.1	1207	7	Q40328	Sequence encoding gly 5.43e+02
	375	12	10.1	1278	10	Q55273	Jun v I gene. 5.43e+02
	376	12	10.1	1290	30	T67838	H. pylori secreted or 5.43e+02
c	377	12	10.1	1305	13	Q79603	Human c-erbB-3 5' reg 5.43e+02
c	378	12	10.1	1305	10	Q58031	c-erbB-3 gene 5' regi 5.43e+02
c	379	12	10.1	1308	37	T86547	Streptococcus pneumon 5.43e+02
c	380	12	10.1	1317	1	Q06128	Deduced sequence of m 5.43e+02
	381	12	10.1	1318	17	T00936	Almond N-glycosidase 5.43e+02
c	382	12	10.1	1339	1	N90598	Human lipocortin-III. 5.43e+02
c	383	12	10.1	1342	18	T06878	PTSG clone N31(7). 5.43e+02
c	384	12	10.1	1342	35	T64688	Prostate/colon tumour 5.43e+02
c	385	12	10.1	1342	18	T06879	Reconstructed prostat 5.43e+02
c	386	12	10.1	1342	35	T64687	DNA for prostate/colo 5.43e+02
	387	12	10.1	1358	17	T00938	Almond N-glycosidase 5.43e+02
c	388	12	10.1	1369	38	V04902	Nucleotide sequence o 5.43e+02
	389	12	10.1	1371	34	T78974	Mouse Huntington's di 5.43e+02
	390	12	10.1	1373	17	Q99951	Recombinant high affi 5.43e+02
	391	12	10.1	1395	2	Q11248	Tryptophanase gene. 5.43e+02
	392	12	10.1	1421	12	Q71698	Corticosteroid bindin 5.43e+02
	393	12	10.1	1423	12	Q67210	Human corticosteroid 5.43e+02
	394	12	10.1	1423	12	Q67209	Human corticosteroid 5.43e+02
	395	12	10.1	1426	40	T12034	Streptococcus pneumon 5.43e+02
	396	12	10.1	1426	17	T00937	Almond N-glycosidase 5.43e+02
	397	12	10.1	1435	17	T00935	Almond N-glycosidase 5.43e+02
c	398	12	10.1	1464	4	Q24053	FK-506 binding protei 5.43e+02
c	399	12	10.1	1466	1	N90714	Gene II of vaccinia W 5.43e+02

c	400	12	10.1	1468	23	T33568	Cowpox virus CrmA gen 5.43e+02
	401	12	10.1	1469	4	N30039	Sequence encoding bov 5.43e+02
c	402	12	10.1	1472	3	Q20509	FKBP. 5.43e+02
c	403	12	10.1	1498	14	Q83837	B. burgdorferi OspE-O 5.43e+02
c	404	12	10.1	1500	30	T60306	Caenorhabditis elegan 5.43e+02
	405	12	10.1	1511	4	Q23524	ORF1 encoding Abi 105 5.43e+02
c	406	12	10.1	1512	37	T94625	Borna disease virus (5.43e+02
c	407	12	10.1	1515	23	T38102	ORF p57 on Borna dise 5.43e+02
	408	12	10.1	1548	1	N70650	Cytochrome-P450 chime 5.43e+02
	409	12	10.1	1560	6	Q38905	Chimeric cytochrome P 5.43e+02
	410	12	10.1	1560	2	N70919	Chimeric cytochrome P 5.43e+02
c	411	12	10.1	1568	15	Q89840	Human death associate 5.43e+02
c	412	12	10.1	1573	11	Q63998	Synthase factor B gen 5.43e+02
c	413	12	10.1	1573	3	Q22620	Brassica synthase fac 5.43e+02
	414	12	10.1	1575	2	N70924	Sequence encoding rat 5.43e+02
	415	12	10.1	1578	4	N40290	Sequence of the alpha 5.43e+02
c	416	12	10.1	1579	11	Q54895	Maize pollen-specific 5.43e+02
	417	12	10.1	1608	22	T34162	Fas promoter region. 5.43e+02
	418	12	10.1	1622	24	T37309	Aromatic acyl transfe 5.43e+02
c	419	12	10.1	1660	5	Q32869	Human PMP-22 cDNA. 5.43e+02
	420	12	10.1	1662	37	T84711	Human ShcB cDNA. 5.43e+02
c	421	12	10.1	1662	36	T88864	Internal FAD2-N DNA e 5.43e+02
c	422	12	10.1	1672	11	Q63995	Synthase factor B gen 5.43e+02
c	423	12	10.1	1683	39	T94198	Microscilla furvescen 5.43e+02
c	424	12	10.1	1683	12	Q70424	Sequence encoding nuc 5.43e+02
	425	12	10.1	1689	11	Q66824	AmEPV NPH-1 gene. 5.43e+02
c	426	12	10.1	1694	10	Q58735	Hantaan virus cDNA. 5.43e+02
	427	12	10.1	1713	17	Q99981	Almond N-glycosidase 5.43e+02
c	428	12	10.1	1736	5	Q32870	Rat PMP cDNA. 5.43e+02
c	429	12	10.1	1753	17	T01051	Mouse B7-1 exon 1. 5.43e+02
	430	12	10.1	1762	5	Q29112	EIV HA (A2/Suffolk/89 5.43e+02
	431	12	10.1	1767	3	Q14338	Rat epithelin precurs 5.43e+02
c	432	12	10.1	1771	30	T62138	Leishmania tropica Lt 5.43e+02
c	433	12	10.1	1775	12	Q73303	Sequence describing M 5.43e+02
c	434	12	10.1	1775	12	Q73302	Sequence describing M 5.43e+02
c	435	12	10.1	1775	12	Q73304	Sequence describing M 5.43e+02
c	436	12	10.1	1775	12	Q73301	Generic sequence desc 5.43e+02
	437	12	10.1	1788	1	N71067	Sequence encoding equ 5.43e+02
	438	12	10.1	1796	7	Q45658	Murine somatostatin r 5.43e+02
	439	12	10.1	1797	1	Q04597	Equine hemagglutinin 5.43e+02
c	440	12	10.1	1802	17	T05712	STLV-PH969 cDNA fragm 5.43e+02
	441	12	10.1	1823	35	T75633	Canine herpesvirus ge 5.43e+02
c	442	12	10.1	1845	3	Q22614	R. communis synthase 5.43e+02
c	443	12	10.1	1845	11	Q63996	Synthase factor B 2-8 5.43e+02
c	444	12	10.1	1876	34	T87067	Chicken platelet-acti 5.43e+02
c	445	12	10.1	1893	2	Q12008	TrpE/androgen recepto 5.43e+02
c	446	12	10.1	1935	15	Q92069	Carbonic-anhydrase cD 5.43e+02
c	447	12	10.1	1943	34	T66496	Rat glucose transport 5.43e+02
c	448	12	10.1	1959	38	T93069	Streptococcus pneumon 5.43e+02
	449	12	10.1	1977	23	T36794	Prostate-specific mem 5.43e+02
c	450	12	10.1	2000	13	Q74367	Bovine beta3 adrenerg 5.43e+02

	451	12	10.1	2032	18	T13381	Pea ADPG-PPase SH2 su 5.43e+02
	452	12	10.1	2038	15	Q89844	Human death associate 5.43e+02
	453	12	10.1	2061	13	Q81180	B.t. toxin PS71M3 gen 5.43e+02
	454	12	10.1	2061	3	Q14670	Dipteran active toxin 5.43e+02
	455	12	10.1	2072	24	T39928	Humanised tilapia ins 5.43e+02
	456	12	10.1	2075	24	T39929	Humanised tilapia ins 5.43e+02
	457	12	10.1	2078	24	T39930	Tilapia insulin gene. 5.43e+02
	458	12	10.1	2083	38	V02472	Bacillus amyloliquefa 5.43e+02
	459	12	10.1	2084	15	Q88067	Bacillus amyloliquefa 5.43e+02
	460	12	10.1	2084	16	Q95032	Bacillus amyloliquefa 5.43e+02
c	461	12	10.1	2094	24	T30126	Alpha-amylase 1 promo 5.43e+02
	462	12	10.1	2098	10	Q55320	Recombinant hyaluroni 5.43e+02
c	463	12	10.1	2151	3	Q20684	pY53.3 (NCIMB 40308). 5.43e+02
c	464	12	10.1	2170	1	N80169	TetM:Tn 1545 gene enc 5.43e+02
c	465	12	10.1	2178	15	Q92066	Maize carbonic-anhydr 5.43e+02
c	466	12	10.1	2190	15	Q92068	Carbonic-anhydrase cD 5.43e+02
c	467	12	10.1	2197	16	T05684	Lymphocyte activation 5.43e+02
	468	12	10.1	2247	30	T47040	Human protein ubiquit 5.43e+02
c	469	12	10.1	2259	26	T48032	Porcine lactoferrin g 5.43e+02
c	470	12	10.1	2260	34	T88030	Brevibacterium flavum 5.43e+02
c	471	12	10.1	2287	2	Q13802	Non-receptor linked p 5.43e+02
c	472	12	10.1	2296	2	Q11559	Partial porcine lacto 5.43e+02
c	473	12	10.1	2306	2	Q11560	Partial porcine/human 5.43e+02
	474	12	10.1	2325	36	T95400	Flavobacterium odorat 5.43e+02
	475	12	10.1	2333	33	T61973	Pseudomonas pseudoalc 5.43e+02
	476	12	10.1	2352	2	Q11490	gSPO-B1 insert encodi 5.43e+02
	477	12	10.1	2366	1	Q06634	Clone encoding rat te 5.43e+02
c	478	12	10.1	2390	30	T67024	Rat beta-amyloid prec 5.43e+02
	479	12	10.1	2466	10	Q58336	S. cerevisiae acetyl- 5.43e+02
c	480	12	10.1	2483	30	T45982	Human cocaine and amp 5.43e+02
c	481	12	10.1	2492	11	Q68479	fedE and fedF genes. 5.43e+02
	482	12	10.1	2520	22	T27766	AD 16c human neural t 5.43e+02
	483	12	10.1	2520	3	Q15180	DNA encoding capsid r 5.43e+02
	484	12	10.1	2520	13	Q77884	Neural thread protein 5.43e+02
c	485	12	10.1	2529	21	T31531	Human 3' apolipoprote 5.43e+02
	486	12	10.1	2534	2	Q10167	Cephalosporin C amida 5.43e+02
	487	12	10.1	2547	30	T67908	H. pylori cytoplasmic 5.43e+02
c	488	12	10.1	2636	37	V01871	Human brain specific 5.43e+02
	489	12	10.1	2646	11	Q61607	Mutated GAL4 gene cod 5.43e+02
c	490	12	10.1	2650	3	Q22831	Sequence encoding ovi 5.43e+02
c	491	12	10.1	2658	23	T38106	Borna disease virus H 5.43e+02
c	492	12	10.1	2687	11	Q54896	Maize pollen-specific 5.43e+02
	493	12	10.1	2696	2	N60909	Plasmid pAU157 insert 5.43e+02
	494	12	10.1	2696	2	N60889	Plasmid pAU157 sequen 5.43e+02
	495	12	10.1	2697	2	N60864	Sequence of plasmid p 5.43e+02
c	496	12	10.1	2720	3	Q21137	Sequence encoding a p 5.43e+02
	497	12	10.1	2730	22	T32583	Alfalfa isoflavone re 5.43e+02
	498	12	10.1	2751	12	Q69001	Product of alternativ 5.43e+02
	499	12	10.1	2751	20	T03625	Alternatively spliced 5.43e+02
c	500	12	10.1	2762	35	T91038	Yeast checkpoint cont 5.43e+02
	501	12	10.1	2770	33	T78599	AS-30D tumour Type II 5.43e+02

	502	12	10.1	2784	20	T09940	Anther ear1 gene from 5.43e+02
c	503	12	10.1	2784	3	Q21138	Sequence of the iron- 5.43e+02
c	504	12	10.1	2818	17	T10554	Sheep mammary gland f 5.43e+02
c	505	12	10.1	2845	6	Q32766	APase gene. 5.43e+02
	506	12	10.1	2846	36	T87040	DNA encoding human TR 5.43e+02
c	507	12	10.1	2873	11	Q54889	Maize pollen-specific 5.43e+02
	508	12	10.1	2923	37	T99233	Coding sequence for m 5.43e+02
	509	12	10.1	2923	31	T76529	DNA encoding macropha 5.43e+02
	510	12	10.1	2946	34	T84563	Human SATB1 cDNA. 5.43e+02
	511	12	10.1	2946	10	Q44407	Sequence encoding spe 5.43e+02
	512	12	10.1	2946	27	T43379	Human cytokine respon 5.43e+02
c	513	12	10.1	2947	15	Q87269	Human MSH2 gene. 5.43e+02
	514	12	10.1	2980	3	N50139	Sequence of Hepatitis 5.43e+02
	515	12	10.1	2980	3	N50274	Partial sequence of h 5.43e+02
	516	12	10.1	2988	33	T84245	Rat ICAM-4 cDNA. 5.43e+02
c	517	12	10.1	3011	2	Q12865	Human Cytotoxic Cell 5.43e+02
	518	12	10.1	3089	10	Q58337	S. cerevisiae acetyl- 5.43e+02
c	519	12	10.1	3095	15	Q93911	Human mismatch repair 5.43e+02
c	520	12	10.1	3095	24	T31290	Mouse poly-immunoglob 5.43e+02
c	521	12	10.1	3109	15	Q93901	Human MSH2 gene, homo 5.43e+02
	522	12	10.1	3185	9	Q50430	Human MTP. 5.43e+02
c	523	12	10.1	3211	40	V05541	Actinobacillus pleuro 5.43e+02
c	524	12	10.1	3217	1	N91578	Rat androgen receptor 5.43e+02
c	525	12	10.1	3217	2	Q12002	Full-length rat andro 5.43e+02
	526	12	10.1	3225	28	T47238	FIV env gene. 5.43e+02
	527	12	10.1	3225	32	T49091	Feline immunodeficien 5.43e+02
	528	12	10.1	3225	29	T62514	FIV envelope gene. 5.43e+02
	529	12	10.1	3225	32	T78402	Feline immunodeficien 5.43e+02
	530	12	10.1	3249	14	Q82748	pRb2 retinoblastoma t 5.43e+02
	531	12	10.1	3273	6	Q36934	5' end of Hepatitis A 5.43e+02
	532	12	10.1	3274	3	N50330	Sequence of hepatitis 5.43e+02
c	533	12	10.1	3348	12	Q70611	Yeast HMG-CoA reducta 5.43e+02
c	534	12	10.1	3360	2	Q05602	SSC1 gene. 5.43e+02
c	535	12	10.1	3360	19	T03677	Hepatitis C genome fr 5.43e+02
	536	12	10.1	3365	7	Q41287	Ubiquitin fusion prot 5.43e+02
	537	12	10.1	3372	1	N80755	Encodes all HAV struc 5.43e+02
	538	12	10.1	3394	16	Q99223	Japanese oyster trans 5.43e+02
	539	12	10.1	3399	1	N81747	Plasmid pALP4 5.43e+02
c	540	12	10.1	3409	33	T47544	CMV IE1. lacking amin 5.43e+02
	541	12	10.1	3420	7	Q43965	Pyrococcus thermostab 5.43e+02
	542	12	10.1	3420	7	Q43518	Pyrococcus sp. DNA po 5.43e+02
	543	12	10.1	3420	18	T17981	Pyrococcus sp. DNA po 5.43e+02
	544	12	10.1	3435	1	N81743	Plasmid pAMP19 5.43e+02
c	545	12	10.1	3437	3	Q14937	Abelson Related Gene, 5.43e+02
	546	12	10.1	3453	1	N81744	Plasmid pALP1 5.43e+02
	547	12	10.1	3453	1	N81746	Plasmid pALP25 5.43e+02
c	548	12	10.1	3480	23	T29388	Human CLAP protein cD 5.43e+02
	549	12	10.1	3489	1	N81745	Plasmid pALP17 5.43e+02
c	550	12	10.1	3499	33	T47542	H6 promoted cytomegal 5.43e+02
c	551	12	10.1	3519	29	T65000	Mouse cell cycle regu 5.43e+02
	552	12	10.1	3581	10	Q44391	Sequence of murine OS 5.43e+02

	553	12	10.1	3635	36	V00086	Rat hexokinase II enc 5.43e+02
	554	12	10.1	3635	36	V00125	Rat hexokinase II enc 5.43e+02
c	555	12	10.1	3670	4	Q22501	Human MIP gene. 5.43e+02
c	556	12	10.1	3680	22	T33559	Leukaemia inhibiting 5.43e+02
c	557	12	10.1	3684	40	V13115	Tobacco necrosis viru 5.43e+02
c	558	12	10.1	3692	36	V00090	Rat hexokinase III en 5.43e+02
c	559	12	10.1	3692	36	V00160	Rat hexokinase III en 5.43e+02
c	560	12	10.1	3746	1	N90693	Bovine peptidyl-glyci 5.43e+02
	561	12	10.1	3756	3	N50525	Bacillus thuringiensi 5.43e+02
c	562	12	10.1	3789	36	T74991	DNA polymerase 385 kD 5.43e+02
c	563	12	10.1	3850	17	T05713	PH969-PTLV-L genomic 5.43e+02
	564	12	10.1	3875	2	Q11561	Partial endothelial N 5.43e+02
	565	12	10.1	3901	12	Q71190	PLRV genomic fragment 5.43e+02
	566	12	10.1	3979	39	V14518	CF-5 pathogen resista 5.43e+02
	567	12	10.1	3979	39	V14519	CF-5 pathogen resista 5.43e+02
	568	12	10.1	3993	37	T88023	Human TRAF5 cDNA. 5.43e+02
	569	12	10.1	4123	39	V14523	CF-5 pathogen resista 5.43e+02
c	570	12	10.1	4180	1	N91773	Rat androgen receptor 5.43e+02
	571	12	10.1	4322	35	T93461	Protein phosphatase 1 5.43e+02
	572	12	10.1	4440	35	T89487	Human A20 protein cDN 5.43e+02
c	573	12	10.1	4483	1	Q03253	Sequence of clone pBS 5.43e+02
c	574	12	10.1	4483	1	N90843	Tobacco basic beta-1, 5.43e+02
c	575	12	10.1	4529	8	Q46121	Porcine sodium ion/gl 5.43e+02
c	576	12	10.1	4550	3	Q14936	Abelson Related Gene, 5.43e+02
c	577	12	10.1	4594	33	T47535	CMV gB transmembrane 5.43e+02
c	578	12	10.1	4594	33	T47533	CMV gB transmembrane 5.43e+02
	579	12	10.1	4595	37	T65123	Human GCAP-I gene. 5.43e+02
c	580	12	10.1	4665	38	T97955	Human epididymis-spec 5.43e+02
	581	12	10.1	4698	16	T01176	P815A antigen precurs 5.43e+02
	582	12	10.1	4698	5	Q32370	P1A genomic DNA. 5.43e+02
	583	12	10.1	4698	19	T14692	MAGE genomic DNA. 5.43e+02
	584	12	10.1	4698	13	Q72475	Entire P1A genomic DN 5.43e+02
	585	12	10.1	4707	31	T70813	Deep Vent DNA polymer 5.43e+02
	586	12	10.1	4707	11	Q66294	Deep Vent DNA polymer 5.43e+02
c	587	12	10.1	4768	33	T47531	H6 promoted cytomegal 5.43e+02
	588	12	10.1	4770	29	T59450	Human papillomavirus 5.43e+02
c	589	12	10.1	4809	11	Q62185	pAQ1 plasmid fragment 5.43e+02
c	590	12	10.1	4835	23	T37050	Drosophila Fru 1 cDNA 5.43e+02
c	591	12	10.1	4946	1	Q03115	Nucleotide sequence a 5.43e+02
c	592	12	10.1	4993	12	Q73385	HincII fragment of Tn 5.43e+02
c	593	12	10.1	4993	2	Q12994	HincII fragment of Tn 5.43e+02
	594	12	10.1	5008	4	Q27223	MSF precursor. 5.43e+02
c	595	12	10.1	5062	33	T47569	CMV gL gene, truncate 5.43e+02
	596	12	10.1	5091	1	Q06164	The iga gene, encodin 5.43e+02
c	597	12	10.1	5233	25	T58225	DNA encoding inositol 5.43e+02
	598	12	10.1	5294	37	T89253	cDNA for human CASPR/ 5.43e+02
c	599	12	10.1	5340	18	T10258	DNA polymerase gene f 5.43e+02
c	600	12	10.1	5340	31	T71296	KOD1 thermostable DNA 5.43e+02
c	601	12	10.1	5342	39	V14525	Full length heat-resi 5.43e+02
c	602	12	10.1	5342	23	T14694	DNA polymerase. 5.43e+02
c	603	12	10.1	5342	23	T28360	DNA polymerase gene, 5.43e+02

	604	12	10.1	5362	39	T97296	Mouse additional sex 5.43e+02
	605	12	10.1	5362	39	T97294	Mouse additional sex 5.43e+02
	606	12	10.1	5398	2	Q13284	P.denitrificans cob a 5.43e+02
	607	12	10.1	5490	38	T86433	DNA for DNA polymeras 5.43e+02
	608	12	10.1	5643	18	T09946	Aspergillus oryzae ar 5.43e+02
	609	12	10.1	5643	38	T97394	Aspergillus oryzae ar 5.43e+02
c	610	12	10.1	6063	6	Q37205	Delta-amino levulinat 5.43e+02
c	611	12	10.1	6645	29	T65002	Mouse cell cycle regu 5.43e+02
	612	12	10.1	6796	4	Q27433	HVTA antigen (gA) gen 5.43e+02
	613	12	10.1	6995	14	Q85466	Ras oncogene/LacZ rep 5.43e+02
	614	12	10.1	7101	19	T29247	C. difficile toxin B 5.43e+02
c	615	12	10.1	7174	2	N91062	Sequence of plasmid p 5.43e+02
	616	12	10.1	7216	17	T12249	Cytochrome P450 isoen 5.43e+02
	617	12	10.1	7291	12	Q67860	ALVAC subgenomic frag 5.43e+02
	618	12	10.1	7323	4	Q25033	Bovine RSV strain A 5 5.43e+02
	619	12	10.1	7350	30	T69982	Canarypox virus DNA c 5.43e+02
	620	12	10.1	7351	3	Q22770	Canarypox virus BglII 5.43e+02
	621	12	10.1	7351	33	T47529	Canarypox virus DNA c 5.43e+02
	622	12	10.1	7366	2	Q10342	Modified lac operon. 5.43e+02
	623	12	10.1	7493	1	Q03512	Attenuated hepatitis 5.43e+02
	624	12	10.1	7493	38	T93023	Hepatitis A virus HM- 5.43e+02
	625	12	10.1	7494	2	N60080	Complete nucleotide s 5.43e+02
	626	12	10.1	7852	20	T27556	Shuttle vector pAdDel 5.43e+02
c	627	12	10.1	7897	20	T27555	Shuttle vector pAdDel 5.43e+02
	628	12	10.1	8136	29	T63236	Plasmid pCR(TM)3lacZ. 5.43e+02
	629	12	10.1	8342	13	Q75209	ALL-1 (acute lymphocy 5.43e+02
	630	12	10.1	8391	19	T16333	MLL gene 8.3 kb fragm 5.43e+02
	631	12	10.1	8392	9	Q53478	MLL gene 8.3 kb BamHI 5.43e+02
c	632	12	10.1	8450	36	T74994	Plasmid pFU1001 conta 5.43e+02
	633	12	10.1	8457	11	Q66797	AmEPV spheroidin gene 5.43e+02
	634	12	10.1	8509	31	T59271	Plasmid pAV.CMVLacZ. 5.43e+02
	635	12	10.1	8668	9	Q50433	Partial human genomic 5.43e+02
c	636	12	10.1	8710	17	Q89650	pSC11 FIPV E1. 5.43e+02
c	637	12	10.1	8906	36	T85392	Human TRIO phosphopro 5.43e+02
c	638	12	10.1	8910	23	T38104	Borna disease virus a 5.43e+02
c	639	12	10.1	9020	17	T07652	pSC11 FIPV N. 5.43e+02
	640	12	10.1	9100	27	T28521	H. influenzae detecti 5.43e+02
c	641	12	10.1	9413	22	T03960	Partial HCV non-struc 5.43e+02
c	642	12	10.1	9413	15	Q81559	Hepatitis C virus hel 5.43e+02
c	643	12	10.1	9413	14	Q80498	DNA encoding HCV prot 5.43e+02
	644	12	10.1	9432	17	Q99312	LKP1 operon. 5.43e+02
	645	12	10.1	9511	6	Q38221	NANBH virus strain HC 5.43e+02
	646	12	10.1	9593	6	Q35297	ZYMV genome. 5.43e+02
c	647	12	10.1	9633	1	N80890	Sequence of cDNA clon 5.43e+02
	648	12	10.1	9661	12	Q73733	Retro virus vector p5 5.43e+02
c	649	12	10.1	9693	1	N92119	Sequence of clone HIV 5.43e+02
	650	12	10.1	9997	36	T97925	Composite sequence of 5.43e+02
	651	12	10.1	9997	40	V05839	Mouse huntingtin codi 5.43e+02
	652	12	10.1	10216	24	T39279	Transposon delivery v 5.43e+02
c	653	12	10.1	10241	18	T10690	Expression vector pBl 5.43e+02
	654	12	10.1	10306	12	Q73734	Retro virus vector p5 5.43e+02

	655	12	10.1	10398	22	T15286	Ad.AV.CMVLacZ hybrid 5.43e+02
c	656	12	10.1	10897	20	T09187	MuTu putative oncogen 5.43e+02
	657	12	10.1	10952	23	T33345	Human CAPL gene. 5.43e+02
	658	12	10.1	10970	12	Q73735	Retro virus vector p5 5.43e+02
	659	12	10.1	11236	12	Q70447	Rat nestin gene - its 5.43e+02
c	660	12	10.1	11715	36	T91087	Bovine lysosomal alph 5.43e+02
c	661	12	10.1	11785	12	Q74694	Genomic RNA sequence 5.43e+02
	662	12	10.1	11800	11	Q67676	Acetyl CoA carboxylas 5.43e+02
	663	12	10.1	11873	3	N40176	Sequence of human fac 5.43e+02
c	664	12	10.1	15894	7	Q40480	Attenuated measles vi 5.43e+02
	665	12	10.1	22481	24	T11658	PEDF full length sequ 5.43e+02
c	666	12	10.1	24026	21	T17528	Mutated BRCA1 genomic 5.43e+02
c	667	12	10.1	29392	40	V15422	Mouse poly Ig recepto 5.43e+02
	668	12	10.1	35100	40	V20441	Human c-fms oncogene. 5.43e+02
	669	12	10.1	35408	31	T59272	Recombinant adenoviru 5.43e+02
	670	12	10.1	37895	24	T34137	Gene for biosynthetic 5.43e+02
	671	12	10.1	37913	10	Q51128	Eicosapentaenoic acid 5.43e+02
c	672	12	10.1	134525	1	Q04525	Total base sequence o 5.43e+02
	673	12	10.1	580073	27	T58840	Mycoplasma genitalium 5.43e+02
c	674	12	10.1	580073	27	T58840	Mycoplasma genitalium 5.43e+02
	675	11	9.2	33	37	V02273	Human parathyroid hor 1.92e+03
c	676	11	9.2	33	24	T31848	Human geranylgeranyl 1.92e+03
c	677	11	9.2	37	40	V10643	A. thaliana PDF1.2 PC 1.92e+03
	678	11	9.2	38	32	T81752	Human c-myb hammerhea 1.92e+03
	679	11	9.2	38	38	V06323	Human Col III gene 3' 1.92e+03
c	680	11	9.2	60	10	Q57696	Diversified HIV-1 tat 1.92e+03
c	681	11	9.2	81	5	Q33791	Sequence upstream fro 1.92e+03
c	682	11	9.2	84	27	T34316	Bovine preprolactin 3 1.92e+03
c	683	11	9.2	108	22	T24229	Human gene signature 1.92e+03
	684	11	9.2	131	12	Q68978	Human-21 cadherin-re 1.92e+03
	685	11	9.2	156	5	Q30054	Sequence of cowpea mo 1.92e+03
	686	11	9.2	159	38	T99939	Chimeric epidermal gr 1.92e+03
	687	11	9.2	159	38	T99941	Chimeric epidermal gr 1.92e+03
	688	11	9.2	163	1	Q06654	Feline T-cell lymphot 1.92e+03
	689	11	9.2	186	30	T68189	H. pylori cytoplasmic 1.92e+03
c	690	11	9.2	197	30	T68552	Butyrivibrio isolate 1.92e+03
c	691	11	9.2	197	30	T68544	Butyrivibrio isolate 1.92e+03
	692	11	9.2	207	20	T24168	Human gene signature 1.92e+03
c	693	11	9.2	208	29	T63479	Type 5 17-beta-hydrox 1.92e+03
	694	11	9.2	219	19	T20858	Human gene signature 1.92e+03
c	695	11	9.2	219	18	T21017	Human gene signature 1.92e+03
	696	11	9.2	223	17	T01041	Mouse B7-2 exon mlB (1.92e+03
c	697	11	9.2	236	25	T39218	Cryptococcus laurenti 1.92e+03
c	698	11	9.2	240	30	T67781	H. pylori membrane pr 1.92e+03
c	699	11	9.2	240	20	T22226	Human gene signature 1.92e+03
c	700	11	9.2	253	10	Q76848	Human genome fragment 1.92e+03
	701	11	9.2	255	35	T69375	Murine metastatic nuc 1.92e+03
c	702	11	9.2	258	8	Q60485	Human brain Expressed 1.92e+03
	703	11	9.2	261	22	T24838	Human gene signature 1.92e+03
	704	11	9.2	288	21	T25374	Human gene signature 1.92e+03
c	705	11	9.2	298	8	Q60083	Human brain Expressed 1.92e+03

c	706	11	9.2	299	8	Q60556	Human brain Expressed 1.92e+03
c	707	11	9.2	308	40	V09118	3' nucleotide sequenc 1.92e+03
c	708	11	9.2	311	8	Q61112	Human brain Expressed 1.92e+03
c	709	11	9.2	312	30	T67709	H. pylori cytoplasmic 1.92e+03
	710	11	9.2	330	20	T22154	Human gene signature 1.92e+03
	711	11	9.2	330	22	T13213	Plasmid pBlue-TH8 fra 1.92e+03
c	712	11	9.2	330	20	T23534	Human gene signature 1.92e+03
	713	11	9.2	336	20	T22217	Human gene signature 1.92e+03
c	714	11	9.2	338	8	Q60300	Human brain Expressed 1.92e+03
	715	11	9.2	341	8	Q59729	Human brain Expressed 1.92e+03
c	716	11	9.2	342	8	Q59382	Human brain Expressed 1.92e+03
c	717	11	9.2	347	8	Q60377	Human brain Expressed 1.92e+03
c	718	11	9.2	354	10	Q55233	Grapevine ribosomal c 1.92e+03
c	719	11	9.2	355	18	T10430	CTP-1 chloroplast tra 1.92e+03
	720	11	9.2	363	5	Q28275	PCR prod. contg. part 1.92e+03
c	721	11	9.2	367	8	Q59872	Human brain Expressed 1.92e+03
	722	11	9.2	375	10	Q57918	Diversified HIV-1 rev 1.92e+03
	723	11	9.2	378	3	N50055	Sequence encoding hum 1.92e+03
c	724	11	9.2	384	30	T77446	H. pylori secreted or 1.92e+03
	725	11	9.2	390	19	T22717	Human gene signature 1.92e+03
	726	11	9.2	395	8	Q59961	Human brain Expressed 1.92e+03
c	727	11	9.2	399	8	Q61271	Human brain Expressed 1.92e+03
	728	11	9.2	400	8	Q59949	Human brain Expressed 1.92e+03
c	729	11	9.2	411	8	Q60549	Human brain Expressed 1.92e+03
	730	11	9.2	423	19	T20515	Human gene signature 1.92e+03
	731	11	9.2	442	3	Q21857	Vector phNF117 contg. 1.92e+03
	732	11	9.2	449	24	T44723	Human papillomavirus 1.92e+03
c	733	11	9.2	462	40	V19628	Human interleukin-4 (1.92e+03
c	734	11	9.2	479	38	T91307	Human H83-22 secreted 1.92e+03
	735	11	9.2	507	19	T14354	Nak2 receptor cDNA al 1.92e+03
c	736	11	9.2	530	24	T43686	Medium chain-specific 1.92e+03
c	737	11	9.2	540	4	N40291	Sequence of the alpha 1.92e+03
c	738	11	9.2	542	40	V10165	Stealth virus nucleic 1.92e+03
c	739	11	9.2	549	19	T19862	Human gene signature 1.92e+03
	740	11	9.2	580	17	T16308	Pea HY4 cDNA partial 1.92e+03
c	741	11	9.2	600	18	T03456	Turkey prolactin gene 1.92e+03
c	742	11	9.2	603	37	T79556	PRRSV field strain 46 1.92e+03
	743	11	9.2	609	25	T44968	Chlorella virus AMT-2 1.92e+03
c	744	11	9.2	616	13	Q53979	Human OTC gene exon 7 1.92e+03
c	745	11	9.2	642	39	V13305	B. burgdorferi sensu 1.92e+03
c	746	11	9.2	645	9	Q54080	Monkey GRP DNA. 1.92e+03
c	747	11	9.2	688	40	V10180	Stealth virus nucleic 1.92e+03
c	748	11	9.2	723	22	T32986	E.coli nitroreductase 1.92e+03
	749	11	9.2	730	29	T63470	Type 5 17-beta-hydrox 1.92e+03
	750	11	9.2	741	10	Q45219	Recombinant M19-M6-M5 1.92e+03
c	751	11	9.2	743	4	Q25871	D4 dopamine receptor 1.92e+03
c	752	11	9.2	747	28	T36612	Hamster islet neogene 1.92e+03
	753	11	9.2	763	1	Q04018	Granulocyte macrophag 1.92e+03
	754	11	9.2	789	8	Q49958	Impatiens Necrotic Sp 1.92e+03
c	755	11	9.2	804	30	T67632	H. pylori cytoplasmic 1.92e+03
	756	11	9.2	813	13	Q75532	Type I ribosome-inact 1.92e+03

	757	11	9.2	848	40	V09121	5' nucleotide sequenc 1.92e+03
	758	11	9.2	858	38	T92922	Promoter probe vector 1.92e+03
c	759	11	9.2	885	30	T67698	H. pylori transporter 1.92e+03
c	760	11	9.2	936	22	T13397	Hepatocyte proliferat 1.92e+03
c	761	11	9.2	957	30	T68046	H. pylori transporter 1.92e+03
c	762	11	9.2	969	11	Q65578	P35S-1 promoter derivr 1.92e+03
	763	11	9.2	978	30	T77462	H. pylori cytoplasmic 1.92e+03
c	764	11	9.2	1017	40	V15428	Human gene fragment Q 1.92e+03
c	765	11	9.2	1023	37	T76685	Pyrococcus furiosus F 1.92e+03
c	766	11	9.2	1078	38	T91308	Human H83-22 secreted 1.92e+03
	767	11	9.2	1084	12	Q66946	Cell death hid gene. 1.92e+03
c	768	11	9.2	1119	40	V15418	Human dendritic cell 1.92e+03
c	769	11	9.2	1199	40	V05163	FMV full length trans 1.92e+03
	770	11	9.2	1201	10	Q54689	Anther specific cDNA 1.92e+03
c	771	11	9.2	1208	22	T17733	Deoxyribonuclease, DN 1.92e+03
	772	11	9.2	1210	8	Q47760	Antisense RNA complem 1.92e+03
	773	11	9.2	1213	1	N91191	Nitilase gene carried 1.92e+03
c	774	11	9.2	1227	40	V24760	H. pylori ORF 01cel16 1.92e+03
	775	11	9.2	1234	29	T07265	Mutant LAG1 coding se 1.92e+03
	776	11	9.2	1248	40	V24743	H. pylori ORF 07cel02 1.92e+03
	777	11	9.2	1261	17	T01046	Mouse B7-2 exons m1B, 1.92e+03
c	778	11	9.2	1314	40	V09165	Nucleotide sequence o 1.92e+03
c	779	11	9.2	1317	31	T71697	Human deoxycytidylate 1.92e+03
c	780	11	9.2	1324	20	Q90738	B31 outer surface pro 1.92e+03
c	781	11	9.2	1347	24	T41599	Onchocerca volvulus b 1.92e+03
	782	11	9.2	1352	16	T02325	Mouse cardiac hypertr 1.92e+03
c	783	11	9.2	1365	38	V03980	P. furiosus ADPHK cod 1.92e+03
	784	11	9.2	1382	13	Q77982	cDNA corresp. to rat 1.92e+03
	785	11	9.2	1412	5	Q32365	MAGE-9 genomic DNA. 1.92e+03
	786	11	9.2	1412	16	T01169	MAGE-9 gene. 1.92e+03
c	787	11	9.2	1419	11	Q63208	Human c-myc exon 3 en 1.92e+03
	788	11	9.2	1446	34	T70001	C5a anaphylatoxin rec 1.92e+03
c	789	11	9.2	1453	18	T08710	Sak serine-threonine 1.92e+03
	790	11	9.2	1494	15	Q87730	Human auxillary cytoc 1.92e+03
	791	11	9.2	1494	15	Q87731	Human auxillary cytoc 1.92e+03
	792	11	9.2	1494	22	T28397	Human cytochrome P450 1.92e+03
c	793	11	9.2	1500	40	T86657	LEAPT gene from Arabi 1.92e+03
c	794	11	9.2	1500	39	V02767	Arabidopsis floral me 1.92e+03
c	795	11	9.2	1500	39	V06024	Arabidopsis floral me 1.92e+03
c	796	11	9.2	1503	39	V15518	Hamster oral papillom 1.92e+03
c	797	11	9.2	1520	2	N71320	Sequence encoding hum 1.92e+03
c	798	11	9.2	1521	1	N90369	DNA sequence of alpha 1.92e+03
	799	11	9.2	1522	14	Q89540	Wheat cold tolerance 1.92e+03
c	800	11	9.2	1542	22	T09695	Caenorhabditis elegan 1.92e+03
c	801	11	9.2	1559	17	T02714	MART-1 melanoma antig 1.92e+03
	802	11	9.2	1560	38	T91295	Human G52-24 secreted 1.92e+03
	803	11	9.2	1593	11	Q66903	P. falciparum protein 1.92e+03
c	804	11	9.2	1599	18	T08712	Sak-b serine-threonin 1.92e+03
	805	11	9.2	1610	31	T72724	p53-GM-CSF immunostim 1.92e+03
	806	11	9.2	1617	7	Q45887	Trans-acting DNA bind 1.92e+03
	807	11	9.2	1618	28	T33650	Human steroidogenesis 1.92e+03

	808	11	9.2	1625	26	T51115	Hippocampus calcium/c 1.92e+03
	809	11	9.2	1635	4	Q24322	Mutant thermostable D 1.92e+03
	810	11	9.2	1640	13	Q72480	Tumour rejection anti 1.92e+03
c	811	11	9.2	1654	38	T93210	Human neurotactin DNA 1.92e+03
c	812	11	9.2	1668	40	V15252	Ricinus communis desa 1.92e+03
c	813	11	9.2	1673	32	T77315	Class II EPSPS DNA fo 1.92e+03
	814	11	9.2	1674	39	V15091	Human sperm surface p 1.92e+03
	815	11	9.2	1679	5	Q29749	NKIR gene. 1.92e+03
c	816	11	9.2	1704	40	V18819	Human protease FMH-1 1.92e+03
	817	11	9.2	1719	38	V04304	Arabidopsis thaliana 1.92e+03
	818	11	9.2	1734	32	T71745	Paramyxovirus haemagg 1.92e+03
c	819	11	9.2	1745	40	V15235	Class I acyl-ACP thio 1.92e+03
c	820	11	9.2	1757	38	V12200	Human neuronal nicoti 1.92e+03
	821	11	9.2	1758	1	Q04719	Sequence encoding hum 1.92e+03
c	822	11	9.2	1760	40	V15310	Aspergillus nidulans 1.92e+03
	823	11	9.2	1761	20	T12920	Human serum albumin g 1.92e+03
	824	11	9.2	1761	20	T12919	Human serum albumin g 1.92e+03
	825	11	9.2	1765	20	T12921	Human serum albumin g 1.92e+03
c	826	11	9.2	1782	17	T03251	Pseudomonas mesoacido 1.92e+03
c	827	11	9.2	1804	40	V20462	Human c-yes oncogene. 1.92e+03
	828	11	9.2	1807	32	T74041	Soybean thiol proteas 1.92e+03
	829	11	9.2	1811	22	T34377	Plasmid pPDET insert. 1.92e+03
	830	11	9.2	1827	1	Q03685	cDNA sequence encodin 1.92e+03
	831	11	9.2	1827	21	T27599	HSA gene with fused m 1.92e+03
c	832	11	9.2	1855	38	T97910	Plasmid pAP-206 inser 1.92e+03
	833	11	9.2	1862	5	Q29604	Opossum kidney PTH/PT 1.92e+03
c	834	11	9.2	1879	38	T97901	Plasmid pAP-197 inser 1.92e+03
c	835	11	9.2	1881	40	V09214	Crk1 cDNA sequence. 1.92e+03
	836	11	9.2	1899	4	Q24321	Mutant thermostable D 1.92e+03
	837	11	9.2	1912	7	Q43443	AML 1 gene. 1.92e+03
c	838	11	9.2	1916	39	V00681	Borrelia burgdorferi 1.92e+03
c	839	11	9.2	1920	40	V15309	Aspergillus nidulans 1.92e+03
	840	11	9.2	1935	12	Q77856	Mycoplasma gallisepti 1.92e+03
c	841	11	9.2	1960	1	N90669	DNA sequence encoadin 1.92e+03
	842	11	9.2	1982	1	N90225	Malaria-specific Plas 1.92e+03
	843	11	9.2	1994	17	T09085	Achatina fulica cDNA. 1.92e+03
c	844	11	9.2	2004	39	V16167	DNA encoding an 80 kD 1.92e+03
c	845	11	9.2	2038	7	Q40490	Clone FG5 containing 1.92e+03
	846	11	9.2	2043	4	Q24320	Mutant thermostable D 1.92e+03
c	847	11	9.2	2057	37	V01916	Human GSPT1-TK gene. 1.92e+03
c	848	11	9.2	2116	25	T18213	Major neutralising an 1.92e+03
	849	11	9.2	2136	2	N60755	Sequence encoding hum 1.92e+03
c	850	11	9.2	2149	18	Q97826	CD69 DNA contg. trans 1.92e+03
c	851	11	9.2	2154	39	V17331	PreproHCPB-linker-hum 1.92e+03
c	852	11	9.2	2188	30	T62441	Mouse growth factor L 1.92e+03
c	853	11	9.2	2214	24	T34612	Probe for detecting p 1.92e+03
c	854	11	9.2	2219	11	Q79724	Human L5/3 tumour sup 1.92e+03
c	855	11	9.2	2229	4	Q23462	Human B-raf protein k 1.92e+03
	856	11	9.2	2280	17	T01403	Canine herpesvirus gC 1.92e+03
c	857	11	9.2	2281	38	T99936	Racillus subtilis rod 1.92e+03
c	858	11	9.2	2287	17	T00785	Human afamin cDNA. 1.92e+03

c	859	11	9.2	2304	5	Q32500	HCV NS2-NS4 clone N23 1.92e+03
	860	11	9.2	2359	24	T42058	Human rotavirus VP4 p 1.92e+03
c	861	11	9.2	2371	4	Q24477	Heat-resistant pullan 1.92e+03
c	862	11	9.2	2404	22	T33220	Oestrogen-regulated p 1.92e+03
	863	11	9.2	2420	13	Q72472	Tumour rejection anti 1.92e+03
	864	11	9.2	2421	39	V01538	Acylcoenzyme A:choles 1.92e+03
	865	11	9.2	2481	18	T13002	Human adrenergic G-pr 1.92e+03
c	866	11	9.2	2490	5	Q31544	Human heregulin-beta2 1.92e+03
	867	11	9.2	2497	7	Q40863	VAC-beta (genomic - p 1.92e+03
	868	11	9.2	2504	30	T62359	Schizosaccharomyces p 1.92e+03
	869	11	9.2	2504	31	T73286	S. pombe origin of re 1.92e+03
	870	11	9.2	2531	13	Q72482	Tumour rejection anti 1.92e+03
	871	11	9.2	2577	3	N50014	Sequence encoding the 1.92e+03
c	872	11	9.2	2589	30	T61590	Human c-IAP1. 1.92e+03
c	873	11	9.2	2628	39	T94195	Bankia gouldi endoglu 1.92e+03
c	874	11	9.2	2635	39	V15587	Human B-cell activati 1.92e+03
c	875	11	9.2	2655	39	V16166	DNA encoding a 100 kD 1.92e+03
	876	11	9.2	2694	39	V09364	GMCSF/HIV-1 chimeric 1.92e+03
	877	11	9.2	2703	39	T88340	Human MP52 cDNA. 1.92e+03
c	878	11	9.2	2706	37	T92835	Flea saliva protein f 1.92e+03
c	879	11	9.2	2724	38	T91280	Human cytomegalovirus 1.92e+03
c	880	11	9.2	2724	13	Q78166	Human cytomegalovirus 1.92e+03
	881	11	9.2	2817	39	V01590	Rat glycine transport 1.92e+03
c	882	11	9.2	2841	19	Q99257	Yeast coagulation gen 1.92e+03
c	883	11	9.2	2871	30	T58014	H. pylori transembra 1.92e+03
	884	11	9.2	2938	9	Q50644	Human Hum-Fli-1 gene 1.92e+03
	885	11	9.2	2960	7	Q42561	Histamine H1 receptor 1.92e+03
c	886	11	9.2	2970	38	T99433	Yeast MPS1 gene encod 1.92e+03
	887	11	9.2	3172	37	T96154	Human TB2 gene. 1.92e+03
	888	11	9.2	3208	12	Q71481	omp2 locus of Brucell 1.92e+03
c	889	11	9.2	3208	12	Q71481	omp2 locus of Brucell 1.92e+03
c	890	11	9.2	3264	40	V10525	Human ITAK DNA #2. 1.92e+03
c	891	11	9.2	3320	38	V09698	Porcine retrovirus po 1.92e+03
	892	11	9.2	3336	9	Q64652	Human Mannose-Binding 1.92e+03
	893	11	9.2	3346	12	Q71482	omp2 locus of Brucell 1.92e+03
	894	11	9.2	3347	12	Q71480	omp2 locus of Brucell 1.92e+03
c	895	11	9.2	3347	12	Q71486	omp2 locus of Brucell 1.92e+03
	896	11	9.2	3360	2	Q10154	Sequence encoding maj 1.92e+03
c	897	11	9.2	3381	29	T47666	Dengue virus serotype 1.92e+03
c	898	11	9.2	3382	40	V18518	cDNA encoding rat vit 1.92e+03
	899	11	9.2	3501	15	Q90529	Rat SIII 110 kDa subu 1.92e+03
	900	11	9.2	3505	2	Q11850	Glutamate receptor 2. 1.92e+03
	901	11	9.2	3521	7	Q41230	Gene encoding major s 1.92e+03
	902	11	9.2	3521	7	Q41226	Clone GP3 encoding ma 1.92e+03
c	903	11	9.2	3558	14	Q80648	Gene encoding CryIE/C 1.92e+03
c	904	11	9.2	3658	24	T39338	Plasmid pVE116 (EcoRI 1.92e+03
	905	11	9.2	3678	9	Q53996	Equine herpes virus 4 1.92e+03
	906	11	9.2	3695	23	T36391	Lactococcus lactis su 1.92e+03
	907	11	9.2	3786	39	T97229	Human ST receptor pro 1.92e+03
c	908	11	9.2	3794	39	V00680	Plasmid pTRH46 contai 1.92e+03
	909	11	9.2	3876	28	T61372	ATH1 gene. 1.92e+03

c	910	11	9.2	4041	39	V16172	DNA sequence encoding 1.92e+03
	911	11	9.2	4071	5	Q28272	A novel type III RTK 1.92e+03
c	912	11	9.2	4108	32	T70123	DNA encoding beta-xyl 1.92e+03
c	913	11	9.2	4164	40	V12216	Mouse retinoid metabo 1.92e+03
c	914	11	9.2	4165	24	T18893	Rat REK7 cDNA. 1.92e+03
c	915	11	9.2	4286	20	T18079	Heml gene. 1.92e+03
c	916	11	9.2	4344	10	Q56782	Insecticidal toxin ge 1.92e+03
c	917	11	9.2	4391	30	T60666	Mouse CD100 antigen c 1.92e+03
c	918	11	9.2	4437	34	T93390	Bloom's syndrome BLM 1.92e+03
c	919	11	9.2	4438	34	T93391	Bloom's syndrome BLM 1.92e+03
c	920	11	9.2	4438	1	N82262	Alpha-2 plasmin inhib 1.92e+03
c	921	11	9.2	4472	19	T05406	Human cell cycle gene 1.92e+03
	922	11	9.2	4487	3	Q14290	Protein C gene from p 1.92e+03
c	923	11	9.2	4568	25	T44039	Human G-protein recep 1.92e+03
	924	11	9.2	4594	33	T47533	CMV gB transmembrane 1.92e+03
	925	11	9.2	4594	33	T47535	CMV gB transmembrane 1.92e+03
c	926	11	9.2	4779	25	T58751	Nucleolar/endosomal a 1.92e+03
c	927	11	9.2	4839	33	T85642	Ras-binding protein p 1.92e+03
	928	11	9.2	4946	1	Q03115	Nucleotide sequence a 1.92e+03
c	929	11	9.2	5042	15	Q84694	Plasmid glucoamylase 1.92e+03
c	930	11	9.2	5055	24	T41471	Human 030 gene (fohy0 1.92e+03
c	931	11	9.2	5155	1	Q04303	mos gene; 5.1 DNA fra 1.92e+03
	932	11	9.2	5159	9	Q51559	Keratin K6 gene. 1.92e+03
c	933	11	9.2	5176	33	T76801	pLJ mutant Rev m10 re 1.92e+03
c	934	11	9.2	5218	39	V00682	Plasmid nMCMVintBL. 1.92e+03
	935	11	9.2	5221	35	T74202	Human LYST1 short iso 1.92e+03
	936	11	9.2	5285	40	T80199	Phosphatidyl inositol 1.92e+03
	937	11	9.2	5319	26	T51085	Full length beta-gala 1.92e+03
	938	11	9.2	5406	13	Q79070	Mouse flk-1 cDNA. 1.92e+03
	939	11	9.2	5510	2	Q12707	Rice actin gene RAc1. 1.92e+03
c	940	11	9.2	5642	8	Q51193	pDE4; plasmid DNA rep 1.92e+03
	941	11	9.2	5719	1	N90388	cDNA encoding human p 1.92e+03
c	942	11	9.2	5865	39	V04002	Retroviral vector pG1 1.92e+03
c	943	11	9.2	5898	39	V00677	DNA plasmid VR2211 co 1.92e+03
c	944	11	9.2	6049	39	V16165	DNA encoding vegetati 1.92e+03
c	945	11	9.2	6049	38	T73993	B. cereus VIP2A(a) co 1.92e+03
c	946	11	9.2	6049	18	T14548	Cytotactin gene. 1.92e+03
c	947	11	9.2	6136	9	Q53454	Urea amide lyase gene 1.92e+03
c	948	11	9.2	6259	38	T86366	SM22 protein gene nuc 1.92e+03
	949	11	9.2	6340	28	T48871	IamS expression vecto 1.92e+03
	950	11	9.2	6819	32	T77782	cDNA encoding GAL4/HA 1.92e+03
c	951	11	9.2	6978	2	N90096	Sequence of plasmid p 1.92e+03
	952	11	9.2	7091	33	T47530	H6 promoted cytomegal 1.92e+03
c	953	11	9.2	7152	2	N81390	Sequence correspondin 1.92e+03
	954	11	9.2	7320	4	Q24171	M13IX22. 1.92e+03
c	955	11	9.2	7326	16	Q99456	Sequence encoding CRE 1.92e+03
	956	11	9.2	7394	4	Q24175	M13ED04. 1.92e+03
c	957	11	9.2	7451	40	V05856	Complete sequence of 1.92e+03
	958	11	9.2	7456	2	Q10686	Plasmid pBHA-1. 1.92e+03
c	959	11	9.2	7617	39	V14354	Plasmid pGen-PGKgfo25 1.92e+03
	960	11	9.2	7652	7	Q45682	Sequencing vector JAN 1.92e+03

c	961	11	9.2	7653	9	Q54401	ced-3 gene. 1.92e+03
c	962	11	9.2	7653	9	Q64740	ced-3 (C63227) gene. 1.92e+03
c	963	11	9.2	7920	20	T05899	Human IL-4 gene. 1.92e+03
	964	11	9.2	8041	18	T10105	Adherence conferring 1.92e+03
c	965	11	9.2	8082	40	V20463	Human c-myc oncogene. 1.92e+03
c	966	11	9.2	8196	38	V09699	Porcine retrovirus DN 1.92e+03
c	967	11	9.2	8209	39	V09700	Porcine retrovirus DN 1.92e+03
c	968	11	9.2	8332	39	V05061	Moloney murine leukem 1.92e+03
c	969	11	9.2	8332	20	T07374	MoMLV gag/pol gene re 1.92e+03
	970	11	9.2	8363	16	Q92408	Human cyclin A gene. 1.92e+03
	971	11	9.2	8543	28	T47073	Maize dwarf mosaic vi 1.92e+03
c	972	11	9.2	9009	40	V12112	Homo sapiens factor V 1.92e+03
	973	11	9.2	9046	16	T02792	Transcription factor 1.92e+03
c	974	11	9.2	9096	38	T88015	Human interleukin 16 1.92e+03
c	975	11	9.2	9144	24	T18951	Plasmid pBE346. 1.92e+03
c	976	11	9.2	9405	6	Q40426	Full-length Hepatitis 1.92e+03
c	977	11	9.2	9636	12	Q67190	P. falciparum transmi 1.92e+03
c	978	11	9.2	9737	5	Q31936	ARV-2 proviral DNA fr 1.92e+03
c	979	11	9.2	9997	40	V05839	Mouse huntingtin codi 1.92e+03
	980	11	9.2	10366	12	Q67401	Huntingtin DNA\IT15 g 1.92e+03
	981	11	9.2	10366	40	V05828	Human huntingtin codi 1.92e+03
	982	11	9.2	10607	6	Q37894	Beta-casein coding se 1.92e+03
	983	11	9.2	10930	20	Q81226	Plasmid pM16-1. 1.92e+03
	984	11	9.2	11994	25	T43075	Wheat acetyl-CoA carb 1.92e+03
c	985	11	9.2	11994	25	T43075	Wheat acetyl-CoA carb 1.92e+03
c	986	11	9.2	12151	11	Q62137	Rice starch branching 1.92e+03
c	987	11	9.2	12492	1	N70543	Complete sequence of 1.92e+03
c	988	11	9.2	14237	39	V05384	Human cathepsin K gen 1.92e+03
	989	11	9.2	14255	7	Q43526	ALL-1 gene cDNA. 1.92e+03
c	990	11	9.2	14556	15	Q90511	CEA genomic clone. 1.92e+03
	991	11	9.2	19307	20	T27558	Shuttle vector pAdDel 1.92e+03
c	992	11	9.2	20303	31	T71699	Human deoxycytidylate 1.92e+03
c	993	11	9.2	24025	21	T17515	Mutated BRCA1 genomic 1.92e+03
c	994	11	9.2	24026	23	T32612	BRCA1, human breast a 1.92e+03
c	995	11	9.2	24029	21	T17520	Mutated BRCA1 genomic 1.92e+03
c	996	11	9.2	24031	21	T17525	Mutated BRCA1 genomic 1.92e+03
c	997	11	9.2	29544	17	T12460	Ovine adenovirus geno 1.92e+03
c	998	11	9.2	40352	40	V02032	MAGE-B cluster DNA se 1.92e+03
c	999	11	9.2	53526	40	T94101	Human PKD1 gene. 1.92e+03
c	1000	11	9.2	53577	40	T94108	Human PKD1 locus betw 1.92e+03

ALIGNMENTS

```
RESULT    1
ID   T21058 standard; cDNA to mRNA; 120 BP.
AC   T21058;
DT   03-JUL-1996  (first entry)
DE   Human gene signature HUMGS02342.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
```



```
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 791; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   120 BP;     45 A;     24 C;     22 G;     28 T;

  Query Match          100.0%;  Score 119;  DB 18;  Length 120;
  Best Local Similarity 100.0%;  Pred. No. 3.34e-94;
  Matches   120;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db       1 gatctttactcttgagaaatgaataagctttctctcagaaatgctgtccctatacactag 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy       1 GATCTTTACTCTTGAGAAATGAATAAGCTTTCTCTCAGAAATGCTGTCCCTATACACTAG 60

Db      61 acaaaactgagcctgtataaggaataaatgggagcgccgaaaagctccntaaaaagcaaa 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      61 ACAAAACTGAGCCTGTATAAGGAATAAATGGGAGCGCCGAAAAGCTCCNTAAAAAGCAAA 120


RESULT    2
ID   Q94109 standard; DNA; 7849 BP.
AC   Q94109;
DT   22-FEB-1996  (first entry)
DE   hML genomic DNA.
KW   Human; thrombopoietin; TPO; mpl ligand; hML; fragment polypeptide;
```



```
KW   megakaryocytopoietic cytokine receptor; thrombopoietic signal;
KW   EPO-domain fragment; erythropoietin; hEPO; haematopoietic cell;
KW   megakaryocyte; thrombocytopenia; myeloproliferative disease;
KW   inflammatory thrombocytosis; iron deficiency; EPO; platelet;
KW   red blood cell; progenitor; hML-2; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   prim_transcript 1166..7289
FT                   /*tag= a
FT   exon            1161..1232
FT                   /*tag= b
FT                   /number= 1
FT   intron          1233..2900
FT                   /*tag= c
FT   exon            2901..3058
FT                   /*tag= d
FT                   /number= 2
FT   signal_peptide  3046..3058
FT                   /*tag= e
FT                   /note= "Encodes partial signal peptide"
FT   intron          3059..3289
FT                   /*tag= f
FT   exon            3290..3417
FT                   /*tag= g
FT                   /number= 3
FT   signal_peptide  3290..3339
FT                   /*tag= h
FT                   /note= "Encodes partial signal peptide"
FT   intron          3418..3703
FT                   /*tag= i
FT   exon            3704..3790
FT                   /*tag= j
FT                   /number= 4
FT   intron          3791..5722
FT                   /*tag= k
FT   exon            5723..5890
FT                   /*tag= l
FT                   /number= 5
FT   intron          5891..6126
FT                   /*tag= m
FT   intron          5891..6138
FT                   /*tag= n
FT                   /note= "Utilises alternative splice site, causes
FT                   formation of cDNA encoding hML-2 comprising
FT                   a four amino acid deletion"
FT   exon            6127..7289
FT                   /*tag= o
FT                   /number= 6
FT   misc_signal     6790..6792
FT                   /*tag= p
FT                   /note= "Stop codon"
```



```
FT   misc_feature    7290
FT                   /*tag= q
FT                   /standard_name= PolyA_site
PN   GB2285446-A.
PD   12-JUL-1995.
PF   21-DEC-1994; 025831.
PR   03-JAN-1994; US-176553.
PR   21-JAN-1994; US-185607.
PR   15-FEB-1994; US-196689.
PR   04-APR-1994; US-223263.
PR   25-MAY-1994; US-249376.
PR   02-DEC-1994; US-348658.
PR   02-DEC-1994; US-348657.
PA   (GETH ) GENENTECH INC.
PI   Eaton DL,  de Sauvage FJ;
DR   WPI; 95-234018/31.
DR   P-PSDB; R76164.
PT   Thrombopoietin polypeptide, ligand for mpl cytokine receptor  -
PT   useful for treating thrombocytopenia and related diseases
PS   Example 8; Fig 14; 192pp; English.
CC   This sequence represents a genomic clone encoding human thrombopoietin
CC   (TPO), also known as mpl ligand (hML).  The human gene comprises 6 exons
CC   within 7 kb of genomic DNA.  Exons 1 and 2 contain 5' untranslated
CC   sequence and the initial four amino acids of the signal peptide.  The
CC   remainder of the secretory signal and the first 26 amino acids of the
CC   mature protein are encoded within exon 3.  The entire carboxyl domain
CC   and 3' untranslated sequence, as well as approx. 50 amino acids of the
CC   erythropoietin-like domain are encoded within exon 6.  The four amino
CC   acids involved in the deletion observed in hML-2 (hTPO-2) are encoded at
CC   the 5' end of exon 6.  Analysis of human genomic DNA by Southern blot
CC   indicated the gene for TPO is present in a single copy. The chromosomal
CC   location of the gene was determined by fluorescent in situ hybridisation
CC   which mapped to chromosome 3q27-28.  The protein encoded by this sequence
CC   binds to mpl which is a megakaryocytopoietic cytokine receptor which may
CC   transduce a thrombopoietic signal.  The N-terminal fragment of hML shows
CC   high levels of homology with human erythropoietin (hEPO).  This
CC   constitutes the EPO-domain.  The EPO-domain spans from Cys7 to Cys151
CC   of mature hML.
SQ   Sequence   7849 BP;    1985 A;    2151 C;    1881 G;    1831 T;

  Query Match          99.2%;  Score 118;  DB 16;  Length 7849;
  Best Local Similarity 99.2%;  Pred. No. 3.13e-93;
  Matches   118;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db    7252 gatctttactcttgagaaatgaataagctttctctcagaaatgctgtccctatacactag 7311
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy       1 GATCTTTACTCTTGAGAAATGAATAAGCTTTCTCTCAGAAATGCTGTCCCTATACACTAG 60

Db    7312 acaaaactgagcctgtataaggaataaatgggagcgccgaaaagctccctaaaaagcaa 7370
           |||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Qy      61 ACAAAACTGAGCCTGTATAAGGAATAAATGGGAGCGCCGAAAAGCTCCNTAAAAAGCAA 119
```

```
RESULT    3
ID   Q99554 standard; cDNA to mRNA; 1721 BP.
AC   Q99554;
DT   29-APR-1996  (first entry)
DE   Thrombopoietin coding sequence.
KW   Thrombopoietin; TPO; humoral factor; platelet; antibody; therapy; AIDS;
KW   thrombocytopenia; hypoplastic anaemia; thrombotic thombocytopenia;
KW   disseminated intravascular coagulation syndrome; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   misc_feature    78..1344
FT                   /*tag= a
FT                   /note= "see Q99553 (claimed)"
FT   cds             25..1086
FT                   /*tag= b
FT                   /product= thrombopoietin
FT   signal_peptide  25..87
FT                   /*tag= c
FT   mat_peptide     88..1083
FT                   /*tag= d
PN   WO9521919-A1.
PD   17-AUG-1995.
PF   14-FEB-1995; J00208.
PR   14-FEB-1994; JP-039090.
PR   14-MAR-1994; US-212164.
PR   25-MAR-1994; JP-079842.
PR   01-APR-1994; US-221020.
PR   01-JUN-1994; JP-155126.
PR   15-JUN-1994; JP-167328.
PR   20-JUL-1994; US-278083.
PR   17-AUG-1994; JP-193169.
PR   17-AUG-1994; JP-227159.
PR   18-AUG-1994; JP-193916.
PR   11-OCT-1994; US-320300.
PR   01-NOV-1994; JP-304167.
PR   01-DEC-1994; JP-298669.
PR   22-DEC-1994; US-361811.
PR   28-DEC-1994; JP-341200.
PR   31-JAN-1995; US-381478.
PA   (KIRI ) KIRIN BREWERY KK.
PI   Akahori H,  Iwamatsu A,  Kato T,  Kuroki R,  Miyazaki H;
PI   Muto T,  Ohgami K, Shimizu T;
DR   WPI; 95-293120/38.
DR   P-PSDB; R81341.
PT   Thrombopoietin polypeptide which specifically stimulates or
PT   increases platelet production - useful in treatment of
PT   thrombocytopenia, also related DNA and vectors
PS   Example 22; Page 282-284; 383pp; English.
CC   This sequence represents the coding sequence for a thrombopoietin (TPO)
```



```
CC   protein.  This sequence contains the sequence represented by Q99553.  TPO
CC   is a humoral factor capable of promoting platelet production.  The
CC   encoded protein can be purified by using a gel affinity column where Mpl
CC   has been coupled to a resin.  Q99551-Q99553 also encode TPO proteins.
CC   The DNA sequences are inserted into vectors which are used to transform
CC   prokaryotic and eukaryotic host cells.  Using the proteins encoded by
CC   these sequences, and derivatives of them, antibodies specifically
CC   immunoreactive with a TPO protein can be created.  The antibodies, DNA
CC   sequences and vectors are used to isolate the protein sequences.  The TPO
CC   proteins can then be used in the treatment of platelet disorders.  These
CC   include thrombocytopenia, hypoplastic anaemia, AIDS, disseminated
CC   intravascular coagulation syndrome and thrombotic thrombocytopenia.
SQ   Sequence   1721 BP;    413 A;    542 C;    371 G;    395 T;

  Query Match           58.8%;  Score 70;  DB 17;  Length 1721;
  Best Local Similarity 100.0%;  Pred. No. 3.1le-47;
  Matches   70;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1623 gatctttactcttgagaaatgaataagctttctctcagaaatgctgtccctatacactag 1682
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      1 GATCTTTACTCTTGAGAAATGAATAAGCTTTCTCTCAGAAATGCTGTCCCTATACACTAG 60

Db   1683 acaaaactga 1692
          ||||||||||
Qy     61 ACAAAACTGA 70


RESULT    4
ID   Q94107 standard; cDNA; 1795 BP.
AC   Q94107;
DT   22-FEB-1996  (first entry)
DE   hML cDNA.
KW   Human; thrombopoietin; TPO; mpl ligand; hML; fragment polypeptide;
KW   megakaryocytopoietic cytokine receptor; thrombopoietic signal;
KW   EPO-domain fragment; erythropoietin; hEPO; haematopoietic cell;
KW   megakaryocyte; thrombocytopenia; myeloproliferative disease;
KW   inflammatory thrombocytosis; iron deficiency; EPO; platelet;
KW   red blood cell; progenitor; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   5'utr           1..215
FT                   /*tag= a
FT   cds             216..1277
FT                   /*tag= b
FT                   /product= human mpl ligand
FT   signal_peptide  216..278
FT                   /*tag= c
FT   mat_peptide     279..274
FT                   /*tag= d
FT   3'utr           278..1795
FT                   /*tag= e
```



```
FT   exon            229..356
FT                   /*tag= f
FT                   /note= "Presumed exon 3"
PN   GB2285446-A.
PD   12-JUL-1995.
PF   21-DEC-1994; 025831.
PR   03-JAN-1994; US-176553.
PR   21-JAN-1994; US-185607.
PR   15-FEB-1994; US-196689.
PR   04-APR-1994; US-223263.
PR   25-MAY-1994; US-249376.
PR   02-DEC-1994; US-348658.
PR   02-DEC-1994; US-348657.
PA   (GETH ) GENENTECH INC.
PI   Eaton DL,  de Sauvage FJ;
DR   WPI; 95-234018/31.
DR   P-PSDB; R76164.
PT   Thrombopoietin polypeptide, ligand for mpl cytokine receptor  -
PT   useful for treating thrombocytopenia and related diseases
PS   Claim 10; Fig 1; 192pp; English.
CC   This sequence encodes human thrombopoietin (TPO), also known as mpl
CC   ligand (hML).  The protein encoded by this sequence binds to mpl which
CC   is a megakaryocytopoietic cytokine receptor which may transduce a
CC   thrombopoietic signal.  The N-terminal fragment of hML shows high levels
CC   of homology with human erythropoietin (hEPO).  This constitutes the EPO-
CC   domain.  The EPO-domain spans from Cys7 to Cys151 of mature hML.  The
CC   full length hML protein and fragment polypeptides derived from the
CC   EPO-domain, may be used to stimulate survival, proliferation,
CC   differentiation, and maturation of haematopoietic cells, esp.
CC   megakaryocytes.  They are useful in the treatment and prevention
CC   of thrombocytopenia, myeloproliferative diseases, inflammatory
CC   thrombocytosis, iron deficiency etc.  Chimera's of the fragment
CC   polypeptides and EPO can regulate growth and proliferation of both
CC   platelet and red blood cell progenitors.
SQ   Sequence   1795 BP;    420 A;    586 C;    382 G;    407 T;

  Query Match           34.5%;  Score 41;  DB 16;  Length 1795;
  Best Local Similarity 100.0%;  Pred. No. 1.08e-20;
  Matches   41;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1737 gatctttactcttgagaaatgaataagctttctctcagaaa 1777
          |||||||||||||||||||||||||||||||||||||||||
Qy      1 GATCTTTACTCTTGAGAAATGAATAAGCTTTCTCTCAGAAA 41


RESULT    5
ID   T64318 standard; cDNA; 1795 BP.
AC   T64318;
DT   21-MAY-1997  (first entry)
DE   Human thrombopoietin cDNA.
KW   TPO; thrombopoietin; C-terminal fragment; platelet count; increase;
```



```
KW   thrombocytopenia; acquired immune deficiency syndrome; AIDS; HIV;
KW   human immunodeficency virus; cytokine; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             216..1277
FT                   /*tag= a
FT                   /product= thrombopoietin
FT                   /note= "C-terminal domain is used for increasing
FT                   platelet count"
FT   signal_peptide  216..278
FT                   /*tag= b
FT   mat_peptide     279..1274
FT                   /*tag= c
PN   US5593666-A.
PD   14-JAN-1997.
PF   16-AUG-1994; 291376.
PR   16-AUG-1994; US-291376.
PR   27-OCT-1994; US-330517.
PA   (UYTE-) UNIV TENNESSEE RES CORP.
PI   McDonald TP;
DR   WPI; 97-099409/09.
DR   P-PSDB; W14905.
PT   Compsns. contg. C-terminal fragments of thrombopoietin - useful for
PT   increasing platelet cell count esp. to treat thrombocytopenia in
PT   patients having HIV/AIDS or having undergone radio- or chemotherapy
PS   Disclosure; Fig 9; 26pp; English.
CC   Platelet counts are increased (especially in AIDS patients) by at
CC   least 20% above a baseline count within 2-4 days of administration
CC   of a C-terminal fragment of thrombopoietin. The C-terminal fragment
CC   consists of amino acids 154-332 of the full-length TPO encoded by
CC   the present sequence.
SQ   Sequence   1795 BP;    420 A;    586 C;    382 G;    407 T;

  Query Match           34.5%;  Score 41;  DB 28;  Length 1795;
  Best Local Similarity 100.0%;  Pred. No. 1.08e-20;
  Matches   41;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1737 gatctttactcttgagaaatgaataagctttctctcagaaa 1777
          |||||||||||||||||||||||||||||||||||||||||
Qy      1 GATCTTTACTCTTGAGAAATGAATAAGCTTTCTCTCAGAAA 41


RESULT    6
ID   T26842 standard; cDNA to mRNA; 25 BP.
AC   T26842;
DT   19-NOV-1996  (first entry)
DE   Human GS HUMGS02342-based reverse transcriptase primer PA-6.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; primer; PCR; amplification;
KW   polymerase chain reaction; yeast; ribosome; ss.
```

```
OS   Synthetic.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Example 10; Page 2186; 2245pp; Japanese.
CC   This is the sequence of a primer based on the sequence of the human gene
CC   signature (GS) sequence HUMGS02342 (T21342) which was used to reverse
CC   transcribe mRNA isolated from human foetal liver polyA+ RNA for cloning
CC   the cDNA encoding the human mpl-ligand, using the 5'-amplifinder (RTM)
CC   kit.  The resultant cDNA strand was ligated to the ampliFINDER Anchor
CC   primer (T26843) and the clone was amplified using PCR primers T26844 and
CC   T26846.
CC   The GS sequences (T19001-T26837) were obtained from 3'-directed cDNA
CC   libraries prepared from various human tissues; synthesis of cDNA was
CC   initiated from the 3'-end of mRNA by using poly(T) as the sole primer.
CC   Each library is constructed so as to reflect accurately the relative
CC   abundance of different mRNAs in the particular tissue from which it was
CC   derived.  The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS sequences)
CC   as a means of diagnosing abnormal cell function or for recognising
CC   different cell types.
SQ   Sequence   25 BP;    2 A;    8 C;    3 G;    12 T;

  Query Match           21.0%;  Score 25;  DB 23;  Length 25;
  Best Local Similarity 100.0%;  Pred. No. 3.59e-07;
  Matches   25;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 ttttcggcgctcccatttattcctt 25
          |||||||||||||||||||||||||
Cp    103 TTTTCGGCGCTCCCATTTATTCCTT 79


RESULT    7
ID   T26844 standard; cDNA to mRNA; 37 BP.
AC   T26844;
DT   19-NOV-1996  (first entry)
DE   Human mpl-ligand gene PA-5 PCR primer.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; primer; PCR; amplification;
KW   polymerase chain reaction; ss.
OS   Synthetic.
```



```
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Example 10; Page 2187; 2245pp; Japanese.
CC   Primers T26844-6 were used to PCR amplify the gene sequence encoding the
CC   human mpl-ligand. T26844 and T26846 were used in a primary PCR whereas
CC   T26845, which is internal to T6844, was used in a secondary nested
CC   amplification reaction.  The template for the PCR was cDNA derived human
CC   foetal liver polyA+ RNA which was reverse transcribed using a primer
CC   based on the  human gene signature (GS) sequence HUMGS02342 (T21342).
CC   The GS sequences (T19001-T26837) were obtained from 3'-directed cDNA
CC   libraries prepared from various human tissues; synthesis of cDNA was
CC   initiated from the 3'-end of mRNA by using poly(T) as the sole primer.
CC   Each library is constructed so as to reflect accurately the relative
CC   abundance of different mRNAs in the particular tissue from which it was
CC   derived.  The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS sequences)
CC   as a means of diagnosing abnormal cell function or for recognising
CC   different cell types.
SQ   Sequence   37 BP;    5 A;    13 C;    6 G;    13 T;

  Query Match           21.0%;  Score 25;  DB 23;  Length 37;
  Best Local Similarity 100.0%;  Pred. No. 3.59e-07;
  Matches   25;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     13 tccttatacaggctcagttttgtct 37
          |||||||||||||||||||||||||
Cp     83 TCCTTATACAGGCTCAGTTTTGTCT 59


RESULT    8
ID   T26845 standard; cDNA to mRNA; 37 BP.
AC   T26845;
DT   19-NOV-1996  (first entry)
DE   Human mpl-ligand gene PA-4 PCR primer.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; primer; PCR; amplification;
KW   polymerase chain reaction; ss.
OS   Synthetic.
PN   WO9514772-A1.
PD   01-JUN-1995.
```



```
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K, Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Example 10; Page 2187; 2245pp; Japanese.
CC   Primers T26844-6 were used to PCR amplify the gene sequence encoding the
CC   human mpl-ligand. T26844 and T26846 were used in a primary PCR whereas
CC   T26845, which is internal to T6844, was used in a secondary nested
CC   amplification reaction.  The template for the PCR was cDNA derived human
CC   foetal liver polyA+ RNA which was reverse transcribed using a primer
CC   based on the  human gene signature (GS) sequence HUMGS02342 (T21342).
CC   The GS sequences (T19001-T26837) were obtained from 3'-directed cDNA
CC   libraries prepared from various human tissues; synthesis of cDNA was
CC   initiated from the 3'-end of mRNA by using poly(T) as the sole primer.
CC   Each library is constructed so as to reflect accurately the relative
CC   abundance of different mRNAs in the particular tissue from which it was
CC   derived.  The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS sequences)
CC   as a means of diagnosing abnormal cell function or for recognising
CC   different cell types.
SQ   Sequence   37 BP;    8 A;    10 C;    10 G;    9 T;

  Query Match           21.0%;  Score 25;  DB 23;  Length 37;
  Best Local Similarity 100.0%;  Pred. No. 3.59e-07;
  Matches   25;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     13 tgtatagggacagcatttctgagag 37
          |||||||||||||||||||||||||
Cp     56 TGTATAGGGACAGCATTTCTGAGAG 32


RESULT    9
ID   Q43202 standard; cDNA; 1350 BP.
AC   Q43202;
DT   01-OCT-1993  (first entry)
DE   Sequence encoding delta-15 desaturase.
KW   Lipid composition; modification; fatty acid desaturase; enzyme; ss.
OS   Arabidopsis thaliana, clone pCF3.
FH   Key             Location/Qualifiers
FT   cds             46..1206
FT                   /*tag=  a
PN   WO9311245-A.
PD   10-JUN-1993.
PF   03-DEC-1992; U10284.
PR   04-DEC-1991; US-804259.
```



```
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
PI   Browse J,  Grau LP,  Kinney AJ,  Pierce JW,  Wierzbicki AM,
PI   Yadav NS;
DR   WPI; 93-197063/24.
DR   P-PSDB; R37592.
PT   Isolated nucleic acid fragment, for plant lipid compsn.
PT   modification - comprises nucleic acid sequence encoding fatty
PT   acid desaturase or related enzyme with high aminoacid identity to
PT   specific polypeptide
PS   Claim 1; Page 127-129; 167pp; English.
CC   pCF3 was isolated from a cDNA library using a 5.2 kb Hind III
CC   fragment containing wild-type genomic DNA as a radiolabeled
CC   hybridization probe. One of the sequenceing primers made to the pCF3
CC   insert was also used. The identity of the expression product of pCF3
CC   as the Arabidopsis microsomal delta-15 desaturase was confirmed by
CC   its biological overexpression in plant tissues. Conserved regions of
CC   R37592 are useful in designing long oligomers for hybridization as
CC   well as shorter ones for use as primers in the PCR. The sequences
CC   of useful regions are given in Q43213-Q43224.
SQ   Sequence   1350 BP;    344 A;    315 C;    282 G;    409 T;

  Query Match           12.6%;  Score 15;  DB 7;  Length 1350;
  Best Local Similarity 100.0%;  Pred. No. 7.94e+00;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    670 gagagaaagcttatt 684
          |||||||||||||||
Cp     36 GAGAGAAAGCTTATT 22


RESULT   10
ID   Q73481 standard; cDNA; 1728 BP.
AC   Q73481;
DT   02-JUN-1995  (first entry)
DE   Low molecular weight urokinase bioactive protein cDNA.
KW   Urokinase bioactive glycoprotein; cardiovascular diseases;
KW   pulmonary embolism; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             14..796
FT                   /*tag=  a
PN   EP-620279-A.
PD   19-OCT-1994.
PF   14-APR-1983; 103629.
PR   15-APR-1982; US-368773.
PR   14-MAR-1983; US-474930.
PA   (GETH ) GENENTECH INC.
PI   Heyneker HL,  Holmes WE,  Vehar GA;
DR   WPI; 94-318362/40.
DR   P-PSDB; R63140.
PT   Prodn. of human urokinase glycoproteins - using a recombinant
```

```
PT   expression system used for the treatment of vascular diseases or
PT   conditions.
PS   Example; Fig 2; 41pp; English.
CC   Q73481 is the cDNA sequence which encodes R63140 a low molecular
CC   weight 33000 dalton urokinase (UK) bioactive protein. This cDNA
CC   was used in the construction of a plasmid capable of transforming
CC   either yeast or vertebrate cells, enabling them to produce the
CC   33000 dalton UK bioactive protein. The UK glycoprotein produced
CC   could then be used in the treatment of cardiovascular diseases,
CC   including pulmonary embolism. The UK produced using this method
CC   had the advantage of a specific activity towards fibrin and extant
CC   thrombi, not demonstrated previously with UK isolated from natural
CC   sources.
SQ   Sequence   1728 BP;    419 A;    419 C;    459 G;    431 T;

  Query Match           12.6%;  Score 15;  DB 12;  Length 1728;
  Best Local Similarity 100.0%;  Pred. No. 7.94e+00;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1159 ttgagaaatgaataa 1173
          |||||||||||||||
Qy     12 TTGAGAAATGAATAA 26


RESULT   11
ID   Q21400 standard; DNA; 2000 BP.
AC   Q21400;
DT   03-JUN-1992  (first entry)
DE   Vibrolysin gene sequence.
KW   Protease; wound treatment; necrotic tissue; collagen; elastin;
KW   fibrin; wound debridement; vascularisation; ss.
OS   Vibrio proteolyticus.
FH   Key             Location/Qualifiers
FT   cds             61..1890
FT                   /*tag=  a
FT                   /product= vibriolysin
PN   EP-472011-A.
PD   26-FEB-1992.
PF   29-JUL-1991; 112732.
PR   15-AUG-1990; US-567884.
PR   13-MAR-1991; US-670612.
PA   (GRAC ) GRACE W R & CO-CONN.
PI   Fortney DZ, Durham DR;
DR   WPI; 92-066263/09.
DR   P-PSDB; R21414.
PT   Protease compsn. for wound treatment - contains protease from
PT   Vibrio, esp. V proteolyticus, which hydrolyses components of
PT   necrotic tissue but not native tissue
PS   Claim 3; Fig 1; 23pp; English.
CC   The vibrolysin protease can be produced by aerobic fermentation of
CC   Vibrio species in nutrient medium.  The DNA sequence was obtd. by
```



```
CC   cloning the Vibrio protease.  A gene library was prepd. using Vibrio
CC   chromosomal DNA.  The DNA underwent partial digestion with Sau3A
CC   and ligation into E. coli cosmid vector pHC79.  The recombinant
CC   vectors were packaged into bacteriophage lambda and used to
CC   transform E. coli strain HB101.  Clones contg. vibriolysin were
CC   found to produce a zone of clearing on milk agar plates due to
CC   proteolytic hydrolysis of the caesin component of milk. The
CC   protease is capable of digesting necrotic tissue while viable
CC   living tissue is not injured.  It causes wound debridement and
CC   stimulates vascularisation and healing of traumatised tissue. The
CC   protease can be used in management of full and partial thickness
CC   wounds, burn wounds, debridement of ulcerative lesions, principally
CC   pressure (decubitus) ulcers and varicose stasis and trophic ulcers,
CC   prepn. of skin graft sites and general surgical wounds such as
CC   amputation, incisional, traumatic and pyogenic wounds and for the
CC   treatment of vaginitis, cervicitis, circumcisions, episitomy, cyst
CC   wounds, carbuncles, sunburn, frostbite and cataract scar tissue.
SQ   Sequence   2000 BP;    537 A;    467 C;    502 G;    494 T;

  Query Match           12.6%;  Score 15;  DB 3;  Length 2000;
  Best Local Similarity 100.0%;  Pred. No. 7.94e+00;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     54 ttgagaaatgaataa 68
          |||||||||||||||
Qy     12 TTGAGAAATGAATAA 26


RESULT   12
ID   N93079 standard; DNA; 2266 BP.
AC   N93079;
DT   27-JUN-1990  (first entry)
DE   Sequence encoding prourokinase.
KW   Prourokinase; CGE 195; plasminogen activator; blood clot lysis.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   signal_peptide  39..98
FT                   /*tag=  a
FT   mat_peptide     99..1331
FT                   /*tag=  b
FT                   /product=prourokinase
FT   cds             80..2266
FT                   */tag=  c
FT                   /label=CGF31 clone
PN   EP-299706-A.
PD   18-JAN-1989.
PF   12-JUL-1988; 306334.
PR   13-JUL-1987; US-072426.
PR   29-JUN-1988; US-211279.
PA   (COLB) Collaborative Research Inc.
PI   Baltimore D, Moir DT, Broeze RJ;
```



```
DR   WPI; 89-017204/03.
DR   P-PSDB; P94764.
PT   New non-glycosylated, secreted plasminogen activator - pref. with
PT   asparagine replaced or deleted, useful for treating blood clots,
PT   expressed in non-mammalian cells.
PS   Disclosure; fig. 1; 26pp; English.
CC   myocardial infarction; ss.
CC   The DNA was sequenced from plasmid pCGE195, a subclone of two inserts
CC   isolated by screening a cDNA library prepd. from kidney cell RNA. One
CC   of the original inserts, clone CGP31 (tag c) started in the middle of
CC   the signal sequence. Mutants of the sequence, pref. in which gcc (Ala)
CC   replaces aat (Asn) at nucleotides 1002-1004 are used to transform hosts
CC   for the prodn. of non-glycosylated prourokinase.
SQ   Sequence   2266 BP;    547 A;    581 C;    595 G;    543 T;

  Query Match           12.6%;  Score 15;  DB 1;  Length 2266;
  Best Local Similarity 100.0%;  Pred. No. 7.94e+00;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1697 ttgagaaatgaataa 1711
          |||||||||||||||
Qy     12 TTGAGAAATGAATAA 26


RESULT   13
ID   N80981 standard; DNA; 2298 BP.
AC   N80981;
DT   14-SEP-1990  (first entry)
DE   Sequence of the single chain urokinase plasminogen activator (SCU-PA)
DE   cDNA insert prepared from human Hep3 cells
KW   Single chain urokinase plasminogen activator (SCU-PA); human Hep3 cells;
KW   thrombosis treatment; thrombosis prevention; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             70..1365
FT                   /*tag=  a
FT                   /product=peptide
FT   mat_peptide     130..1365
FT                   /*tag=  b
FT   misc_feature    2298
FT                   /*tag=  c
FT                   /note="polyA site"
PN   EP-288435-A.
PD   26-OCT-1988.
PF   11-APR-1988; 810234.
PR   16-JUN-1987; GB-014059,
PA   (CIBA) Ciba Geigy Ag.
PI   Meyhack B, Heim J, Burgi R;
DR   WPI; 88-301440/43.
DR   P-PSDB; P80430.
PT   Prodn. of human single chain urokinase type plasminogen activator -
```



```
PT   by culturing yeast strain transformed with hybrid vector contg.
PT   yeast expression control sequences
PS   Example 1; Fig 2; 48pp; English.
CC   The patent is for the prodn. of human single chain urokinase-type
CC   plasminogen activator (UTPA).  Mutants of scu-PA are especially those
CC   which render the protein protease resistant.  The UTPA proteins exhibit
CC   the biological activity of natural human UTPA without any refolding
CC   procedure being necessary.  They can be used as for known PAs in humans
CC   for the prevention or treatment of thrombosis or other conditions where
CC   it is desired to produce local fibrinolytic or proteolytic activity.
SQ   Sequence   2298 BP;    548 A;    594 C;    607 G;    548 T;    1 Others;

  Query Match           12.6%;  Score 15;  DB 1;  Length 2298;
  Best Local Similarity 100.0%;  Pred. No. 7.94e+00;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1728 ttgagaaatgaataa 1742
          |||||||||||||||
Qy     12 TTGAGAAATGAATAA 26


RESULT   14
ID   N60703 standard; DNA; 2299 BP.
AC   N60703;
DT   23-OCT-1991  (first entry)
DE   Sequence encoding human urokinase.
KW   E.coli; high molecular urokinase; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             67..1362
FT                   /*tag=  a
FT   mat_peptide     127..1362
FT                   /*tag=  b
PN   J61181377-A.
PD   14-AUG-1986.
PF   25-JAN-1985; 011032.
PR   25-JAN-1985; JP-011032.
PA   (NISC ) NISSAN CHEM IND KK.
PA   (HODO ) HODOGAYA CHEM IND KK.
PA   (SAGA ) SAGAMI CHEM RES CENTRE.
PA   (CENG ) CENTRAL GLASS KK.
PA   (NIPS ) NIPPON SODA KK.
PA   (TOYJ ) TOYO SODA MFG KK.
DR   WPI; 86-254744/39.
DR   P-PSDB; P60703.
PT   Human urokinase gene - has N-end of aminoacid sequence coded by
PT   codon used in Escherichia coli.
PS   Disclosure; Fig 2; 19pp; Japanese.
CC   The claimed gene product may be expressed in a transformed E.coli
CC   host, for the efficient production of high molecular human urokinase.
CC   The N-terminal of the protein expressed by the transforming plasmid
```

```
CC   is replaced with a codon frequently used in E.coli.
SQ   Sequence   2299 BP;    555 A;    596 C;    606 G;    542 T;

  Query Match          12.6%;  Score 15;  DB 2;  Length 2299;
  Best Local Similarity 100.0%;  Pred. No. 7.94e+00;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1724 ttgagaaatgaataa 1738
         |||||||||||||||
Qy     12 TTGAGAAATGAATAA 26


RESULT   15
ID   N60591 standard; DNA; 2301 BP.
AC   N60591;
DT   16-OCT-1991  (first entry)
DE   Sequence encoding modified human pro-urokinase clone.
KW   Fibrinolytic; protease; trypsin; thrombin; plasmin.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             67..1362
FT                   /*tag=  a
PN   WO8604351-A.
PD   31-JUL-1986.
PF   24-JAN-1986; J00031.
PR   25-JAN-1985; JP-011033.
PR   20-APR-1985; JP-083611.
PR   20-JUL-1985; JP-159294.
PA   (SAGA ) SAGAMI CHEM RES CENTRE
PA   (MIYA/) MIYAKE T
PA   (HODO ) HODOGAYA-CHEM CO LTD
PA   (NIPS ) NIPPON SODA CO LTD
PA   (TOYJ ) TOYO SODA MFG CO LTD
PA   (NISC ) NISSAN CHEM IND LTD
PA   (TOYJ ) TOYO SODA MFG CO LTD
PA   (NIPC ) NIPPON CHEM IND LTD
PA   (TOYJ ) TOYO SODA MFG CO LTD
PA   (NISC ) NISSAN CHEM IND LTD
PA   (NISC ) NISSAN CHEM IND KK
PA   (TOYJ ) TOYO SODA MFG KK
PI   MIYAKE T,HIBINO Y,KOBAYASHI Y,WATABE K,OMORI M,MIKI T,YOKOYAMA M;
PI   MATSUMOTO R;
DR   WPI; 86-212124/32.
DR   P-PSDB; P60674.
PT   Human pro-urokinase stable to protease and 135,157-modified - prepd.
PT   using transformant e.coli and having long-acting fibrinolytic activity
PS   Disclosure; Fig 4; 98pp; Japanese.
CC   Sequence is modified from the human pro-urokinase with a non-basic
CC   residue replacing Lys135. The product may be expressed from a
CC   transformed host such as E.coli, it has long-acting fibrinolytic
CC   activity and is stable to proteases such as trypsin, thrombin and
```



```
CC   plasmin.
SQ   Sequence   2301 BP;    555 A;    597 C;    606 G;    543 T;

  Query Match          12.6%;  Score 15;  DB 2;  Length 2301;
  Best Local Similarity 100.0%;  Pred. No. 7.94e+00;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1726 ttgagaaatgaataa 1740
         |||||||||||||||
Qy     12 TTGAGAAATGAATAA 26


RESULT   16
ID   N70390 standard; DNA; 2301 BP.
AC   N70390;
DT   19-MAR-1991  (first entry)
DE   cDNA encoding human prourokinase in plasmid pKU22.
KW   Cardiovascular disease treatment; fibrin affinity; thrombolytic;
KW   enzyme; protease; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             67..126
FT                   /*tag=  a
FT                   /product= leader
FT   mat_peptide     127..1363
FT                   /*tag=  b
FT                   /product= prourokinase
PN   EP-231883-A.
PD   12-AUG-1987.
PF   29-JAN-1987; 101209.
PR   31-JAN-1986; JP-017734.
PR   30-JAN-1987; JP-018626.
PA   (SAGA ) SAGAMI CHEM RES CENTRE.
PA   (NIPS ) NIPPON SODA KK.
PA   (CENG ) CENTRAL GLASS CO LTD.
PA   (TOYJ ) TOYO SODA MFG KK.
PA   (NISC ) NISSAN CHEMICAL INDS KK.
PI   Tagawa M, Wada M, Yamada M, Yokoyama M, Numao N;
DR   WPI; 87-222882/32.
DR   P-PSDB; P70258.
PT   Hybrid plasminogen activator-like polypeptide - having a region
PT   for affinity to fibrin from tissue plasminogen activator and a
PT   region from prourokinase
PS   Disclosure; Fig 2(1-5); 64pp; English.
CC   The TPA portion of the claimed hybrid polypeptide (see FT) may
CC   consist of 2 kringles from N-terminal first serine to 219th glycine
CC   of human TPA, 1 kringle from 128th serine to 219th glycine of human
CC   TPA or half a kringle from 161st methionine to 219th glycine (see
CC   P70257). The C-terminal half of the hybrid polypeptide may contain
CC   an AA sequence from 150th glutamine to C-terminal 411th leucine of
CC   prourokinase (see P70258).
```



```
SQ   Sequence   2301 BP;    555 A;    596 C;    606 G;    544 T;

  Query Match          12.6%;  Score 15;  DB 2;  Length 2301;
  Best Local Similarity 100.0%;  Pred. No. 7.94e+00;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1726 ttgagaaatgaataa 1740
         |||||||||||||||
Qy     12 TTGAGAAATGAATAA 26


RESULT   17
ID   N91075 standard; DNA; 2301 BP.
AC   N91075;
DT   29-JUN-1990  (first entry)
DE   DNA encoding natural human prourokinase and 5' UTR and 3' UTR
KW   Human prourokinase; antithrombotic.
OS   Homo sapiens
FH   Key             Location/Qualifiers
FT   cds             67..1362
FT                   /*tag=  a
FT                   /product=Human prourokinase
PN   WO8901513-A.
PD   23-FEB-1989.
PF   18-AUG-1988; J00815.
PR   19-AUG-1987; JP-204149.
PA   (CENG) Central Glass Kk (Saga Hodo Nips).
PI   Kobayashi Y, Omori M, Yamanda C;
DR   WPI; 89-068869/09.
DR   P-PSDB; P92119.
PT   Antithrombotic fast-acting pro-urokinase deriv. -
PT   produced by culture of E. coli transformant contg. new plasmid
PT   of PMUT9Q family.
PS   Fig 1; ; 75pp; Japanese.
CC   A human prourokinase (PU) deriv. is new which is based upon residues
CC   2-155 of natural human prourokinase. The new deriv. is produced by
CC   E. coli J103/pMUT9Q-RPK in culture.
SQ   Sequence   2301 BP;    555 A;    596 C;    606 G;    544 T;

  Query Match          12.6%;  Score 15;  DB 1;  Length 2301;
  Best Local Similarity 100.0%;  Pred. No. 7.94e+00;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1726 ttgagaaatgaataa 1740
         |||||||||||||||
Qy     12 TTGAGAAATGAATAA 26


RESULT   18
ID   N91740 standard; DNA; 2303 BP.
AC   N91740;
```



```
DT   16-APR-1990  (first entry)
DE   Sequence of prourokinase cDNA
KW   Low mol. wt. plasminogen activator; prourokinase; dissolve blood clot.
FH   Key             Location/Qualifiers
FT   cds             76..1371
FT                   /*tag=  a
FT   mat_protein     136..1371
FT                   /*tag=  b
FT   misc_feature    117
FT                   /*tag=  c
FT                   /note= "Start of initial isolated clone."
FT   misc_feature    583..609
FT                   /*tag=  d
FT                   /note= "Preferred initiation region for low mol. wt.
FT                   plasminogen activators."
PN   EP-316068-A.
PD   17-MAY-1989.
PF   07-OCT-1988; 309417.
PR   27-SEP-1988; US-248727, US-107370.
PA   (COLB) Collaborative Res Inc.
PI   Mao J;
DR   WPI; 89-146601/20.
DR   P-PSDB; P91886.
PT   Modified low mol. wt. plasmingen activator- formed of
PT   amino acids comprising the amino acid portion of
PT   prourokinase from 150 to 411
PS   Disclosure; Figure 1; 27pp; English.
CC   A low mol. wt. polypeptide plasminogen activator (PA) is claimed which is
CC   formed from amino acids 150-411 of prourokinase. The preferred intiation
CC   region for the low mol. wt. PA is indicated (tag d). A low mol wt. PA
CC   can be injected into blood in the body in vivo to dissolve clots
CC   without harm.
SQ   Sequence   2303 BP;    550 A;    600 C;    606 G;    547 T;

  Query Match          12.6%;  Score 15;  DB 1;  Length 2303;
  Best Local Similarity 100.0%;  Pred. No. 7.94e+00;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1734 ttgagaaatgaataa 1748
         |||||||||||||||
Qy     12 TTGAGAAATGAATAA 26


RESULT   19
ID   Q73483 standard; cDNA; 2303 BP.
AC   Q73483;
DT   02-JUN-1995  (first entry)
DE   Full length human urokinase protein cDNA.
KW   Human urokinase glycoprotein; cardiovascular diseases;
KW   pulmonary embolism; ds.
OS   Homo sapiens.
```

```
FH   Key             Location/Qualifiers
FT   cds             73..1371
FT                   /*tag= a
FT   signal_peptide  73..135
FT                   /*tag= b
FT   mat_peptide     136..1368
FT                   /*tag= c
PN   EP-620279-A.
PD   19-OCT-1994.
PF   14-APR-1983; 103629.
PR   15-APR-1982; US-368773.
PR   14-MAR-1983; US-474930.
PA   (GETH ) GENENTECH INC.
PI   Heyneker HL,  Holmes WE,  Vehar GA;
DR   WPI; 94-318362/40.
DR   P-PSDB; R63141.
PT   Prodn. of human urokinase glycoproteins - using a recombinant
PT   expression system used for the treatment of vascular diseases or
PT   conditions.
PS   Example; Fig 4; 41pp; English.
CC   Q73483 is the cDNA sequence which encodes R63141 the full length
CC   54000 dalton human urokinase (UK) protein. This cDNA was used in
CC   the construction of a plasmid capable of transforming either yeast
CC   or vertebrate cells, enabling them to produce the 54000 dalton
CC   human UK protein. The UK glycoprotein produced could then be used
CC   in the treatment of cardiovascular diseases, including pulmonary
CC   embolism. The UK produced using this method had the advantage of a
CC   specific activity towards fibrin and extant thrombi, not
CC   demonstrated previously with UK isolated from natural sources.
SQ   Sequence   2303 BP;    550 A;    600 C;    606 G;    547 T;

  Query Match           12.6%;  Score 15;  DB 12;  Length 2303;
  Best Local Similarity 100.0%;  Pred. No. 7.94e+00;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1734 ttgagaaatgaataa 1748
          |||||||||||||||
Qy     12 TTGAGAAATGAATAA 26


RESULT   20
ID   N71368 standard; DNA; 2304 BP.
AC   N71368;
DT   30-APR-1991  (first entry)
DE   Modified prourokinase (MET/SER).
KW   Prourokinase; fibrin; PUK; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             77..1369
FT                   /*tag= a
FT                   /product= prourokinase
```



```
FT   signal_peptide  77..136
FT                   /*tag= b
FT   mat_peptide     137..1369
FT                   /*tag= c
PN   EP-236040-A.
PD   09-SEP-1987.
PF   24-FEB-1987; 301567.
PR   26-FEB-1986; US-833179.
PR   19-FEB-1987; US-012023.
PA   (COLB) COLLABORATIVE RES INC.
PI   Vovis GF, Mao JI;
DR   WPI; 87-251560/36.
DR   P-PSDB; P71698.
PT   Amino acid modified form of prourokinase - having high fibrin
PT   selectively and resistant to cleavage to the two-chain form
PS   Disclosure; Fig.1; 18pp; English.
CC   The amino acid LYS at position 158 of the mature protein (see tag b)
CC   is replaced by MET and the amino acid ILE at position 160 (bases
CC   614-616) is replaced by SER, achieved by site-directed mutagenesis
CC   or by chemial modification.
CC   The modified PUK has a single polypeptide chain, high fibrin
CC   activity and is resistant to cleavage to the two-chain form to
CC   an extent greater than its original unmodified form.
CC   The therapeutic specificity of single-chain PUK can be increased
CC   without decreasing fibrinolytic activity.  The single chain form
CC   can have high fibrin activity and thus does not cause general
CC   bleeding in the body but rather can be localised to the site of
CC   the blood clot directly.
CC   See also N71368-69 and N71338.
SQ   Sequence   2304 BP;    547 A;    621 C;    588 G;    547 T;

  Query Match           12.6%;  Score 15;  DB 2;  Length 2304;
  Best Local Similarity 100.0%;  Pred. No. 7.94e+00;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1735 ttgagaaatgaataa 1749
          |||||||||||||||
Qy     12 TTGAGAAATGAATAA 26


RESULT   21
ID   N71330 standard; DNA; 2304 BP.
AC   N71330;
DT   15-MAY-1991  (first entry)
DE   Sequence encoding modified prourokinase.
KW   Fibrin selectivity; ds.
FH   Key             Location/Qualifiers
FT   misc_feature    608..610
FT                   /*tag= a
FT                   /label= Absent or any codon except Leu
PN   EP-236040-A.
```



```
PD   09-SEP-1987.
PF   24-FEB-1987; 301567.
PR   26-FEB-1986; US-833179.
PR   19-FEB-1987; US-012023.
PA   (COLB ) COLLABORATIVE RES INC.
PI   Vovis GF, Mao JI;
DR   WPI; 87-251560/36.
PT   Amino acid modified form of prourokinase - having high fibrin
PT   selectively and resistant to cleavage to the two-chain form
PS   Claim 10; Fig 1; 18pp; Japanese.
CC   The modified prourokinase product has a single polypeptide chain,
CC   high fibrin selectivity and resistance to cleavage into the two-
CC   chain form. The modification is at the codon for Lys158.
SQ   Sequence   2304 BP;    546 A;    599 C;    609 G;    547 T;

  Query Match           12.6%;  Score 15;  DB 2;  Length 2304;
  Best Local Similarity 100.0%;  Pred. No. 7.94e+00;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1735 ttgagaaatgaataa 1749
          |||||||||||||||
Qy     12 TTGAGAAATGAATAA 26


RESULT   22
ID   N30030 standard; cDNA; 2304 BP.
AC   N30030;
DT   25-APR-1992  (first entry)
DE   Sequence encoding full length urokinase.
KW   Enzyme; protease; thrombolytic; fibrinolytic system; coagulation;
KW   plasminogen; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   signal_peptide  77..136
FT                   /*tag= a
FT   mat_peptide     137..1372
FT                   /*tag= b
PN   EP--92182-A.
PD   26-OCT-1983.
PF   14-APR-1983; 066284.
PR   15-APR-1982; US-368773.
PR   14-MAR-1983; US-474930.
PA   (GETH ) GENENTECH INC.
PI   Heyneker HL, Holmes WE, Vehar GA;
DR   WPI; 83-802735/44.
DR   P-PSDB; P30001.
PT   Protein comprising enzymatic portion of human urokinase - prepd.
PT   by recombinant DNA methods for cardiovascular use
PS   Disclosure; Fig 4; 64pp; English.
CC   N30030 was prepd. from a cDNA library from Detroit 562 (human
CC   pharangeal carcinoma) cells (38)(ATCC No. CCL138).  P30041 is a
```



```
CC   bioinactive protein.  Proteolytic cleavage, e.g. with trypsin, at
CC   the N-terminus or at the Lys-Ile junction of the sequence Lys-Ile-
CC   Ile-Gly-Gly (see FT) gives bioactive human urokinase (2-chain
CC   urokinase).
SQ   Sequence   2304 BP;    550 A;    600 C;    608 G;    546 T;

  Query Match           12.6%;  Score 15;  DB 4;  Length 2304;
  Best Local Similarity 100.0%;  Pred. No. 7.94e+00;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1735 ttgagaaatgaataa 1749
          |||||||||||||||
Qy     12 TTGAGAAATGAATAA 26


RESULT   23
ID   N71338 standard; DNA; 2304 BP.
AC   N71338;
DT   30-APR-1991  (first entry)
DE   Modified prourokinase.
KW   Prourokinase; fibrin; PUK; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             77..1369
FT                   /*tag= a
FT                   /product= prourokinase
FT   misc_difference 608..610
FT                   /*tag= b
FT                   /note= "codon encoding amino acid other than LYS"
FT   signal_peptide  77..136
FT                   /*tag= c
FT   mat_peptide     137..1369
FT                   /*tag= d
PN   EP-236040-A.
PD   09-SEP-1987.
PF   24-FEB-1987; 301567.
PR   26-FEB-1986; US-833179.
PR   19-FEB-1987; US-012023.
PA   (COLB) COLLABORATIVE RES INC.
PI   Vovis GF, Mao JI;
DR   WPI; 87-251560/36.
DR   P-PSDB; P71663.
PT   Amino acid modified form of prourokinase - having high fibrin
PT   selectively and resistant to cleavage to the two-chain form
PS   Disclosure; Fig. 1; 18pp; English.
CC   The amino acid LYS at position 158 of the mature protein (see tag b)
CC   is replaced by another amino acid, achieved by site-directed
CC   mutagenesis or by chemial modification.
CC   The modified PUK has a single polypeptide chain, high fibrin
CC   activity and is resistant to cleavage to the two-chain form to
CC   an extent greater than its original unmodified form.
```

```
CC   The therapeutic specificity of single-chain PUK can be increased
CC   without decreasing fibrinolytic activity.  The single chain form
CC   can have high fibrin activity and thus does not cause general
CC   bleeding in the body but rather can be localised to the site of
CC   the blood clot directly.
CC   See also N71368-69.
SQ   Sequence   2304 BP;    547 A;    620 C;    587 G;    547 T;

  Query Match          12.6%;  Score 15;  DB 2;  Length 2304;
  Best Local Similarity 100.0%;  Pred. No. 7.94e+00;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1735 ttgagaaatgaataa 1749
          |||||||||||||||
Qy     12 TTGAGAAATGAATAA 26


RESULT   24
ID   N71369 standard; DNA; 2304 BP.
AC   N71369;
DT   30-APR-1991  (first entry)
DE   Modified prourokinase (ALA).
KW   Prourokinase; fibrin; PUK; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds              77..1369
FT                    /*tag=  a
FT                    /product= prourokinase
FT   signal_peptide   77..136
FT                    /*tag=  b
FT   mat_peptide      137..1369
FT                    /*tag=  c
PN   EP-236040-A.
PD   09-SEP-1987.
PF   24-FEB-1987; 301567.
PR   26-FEB-1986; US-833179.
PR   19-FEB-1987; US-012023.
PA   (COLB) COLLABORATIVE RES INC.
PI   Vovis GF, Mao JI;
DR   WPI; 87-251560/36.
DR   P-PSDB; P71699.
PT   Amino acid modified form of prourokinase - having high fibrin
PT   selectively and resistant to cleavage to the two-chain form
PS   Disclosure; Fig.1; 18pp; English.
CC   The amino acid LYS at position 158 of the mature protein (bases
CC   610-612) is replaced by another amino acid (ALA), achieved by
CC   site-directed mutagenesis.
CC   The modified PUK has a single polypeptide chain, high fibrin
CC   activity and is resistant to cleavage to the two-chain form to
CC   an extent greater than its original unmodified form.
CC   The therapeutic specificity of single-chain PUK can be increased
```



```
CC   without decreasing fibrinolytic activity.  The single chain form
CC   can have high fibrin activity and thus does not cause general
CC   bleeding in the body but rather can be localised to the site of
CC   the blood clot directly.
CC   See also N71368-69.
SQ   Sequence   2304 BP;    547 A;    620 C;    591 G;    546 T;

  Query Match          12.6%;  Score 15;  DB 2;  Length 2304;
  Best Local Similarity 100.0%;  Pred. No. 7.94e+00;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1735 ttgagaaatgaataa 1749
          |||||||||||||||
Qy     12 TTGAGAAATGAATAA 26


RESULT   25
ID   Q68538 standard; DNA; 2377 BP.
AC   Q68538;
DT   20-FEB-1995  (first entry)
DE   Extracellular neutral protease (vibriolysin) gene.
KW   Extracellular; neutral protease; vibriolysin; V. proteolyticus;
KW   recombinant; gram-negative; microorganisms; E. coli; bacteria;
KW   secretion; lysis; ds.
OS   Vibrio proteolyticus ATCC 53559.
FH   Key             Location/Qualifiers
FT   cds              249..2081
FT                    /*tag= a
FT                    /note= "Large open reading frame - encodes neutral
FT                    protease"
PN   EP-605073-A.
PD   06-JUL-1994.
PF   21-SEP-1988; 115439.
PR   01-OCT-1987; US-103983.
PR   19-NOV-1987; US-123038.
PA   (GRAC ) GRACE & CO-CONN W R.
PI   David VA,  Deutch AH;
DR   WPI; 94-210320/26.
DR   P-PSDB; R59765.
PT   Recombinant DNA contg neutral protease genes - obtd from Vibrio
PT   species or Bacillus stearothermophilus, used for large scale
PT   prodn of the enzyme.
PS   Claim 5; Fig 1; 15pp; English.
CC   This sequence encodes the extracellular neutral protease (vibriolysin)
CC   from V. proteolyticus ATCC 53559.  This sequence may be used in the
CC   production of recombinant gram-negative microorganisms, such as E. coli,
CC   which secrete the heterologous protease.  The enzyme may also be
CC   liberated by lysis of the microbial cell.  Recombinant E. coli producing
CC   the protease may be used for the large scale production of active,
CC   functional, neutral protease enzymes for industrial use.
SQ   Sequence   2377 BP;    646 A;    528 C;    586 G;    617 T;
```



```
  Query Match          12.6%;  Score 15;  DB 11;  Length 2377;
  Best Local Similarity 100.0%;  Pred. No. 7.94e+00;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    245 ttgagaaatgaataa 259
          |||||||||||||||
Qy     12 TTGAGAAATGAATAA 26


RESULT   26
ID   Q20360 standard; cDNA; 2377 BP.
AC   Q20360;
DT   20-MAY-1992  (first entry)
DE   Human pro-urokinase cDNA.
KW   PFC44; proUK; plasminogen; ss.
FH   Key             Location/Qualifiers
FT   cds              69..1364
FT                    /*tag=  a
FT                    /label= PUK
FT   polya_signal     2260..2265
FT                    /*tag=  b
FT                    /number= 1
FT   polya_signal     2273..2278
FT                    /*tag=  c
FT                    /number= 2
PN   DE4122688-A.
PD   16-JAN-1992.
PF   09-JUL-1991; 122688.
PR   12-JUL-1990; GB-015369.
PR   10-JUL-1991; GB-014846.
PA   (FARM ) FARMITALIA C ERBA SRL.
PI   Gozzini L, Visco C, Perego R, Roncucci R, Sarmientos P;
DR   WPI; 92-025815/04.
PT   New amidated derivs. of human pro-urokinase - are fibrinolytic
PT   and can be used to treat acute myocardial infarction, pulmonary
PT   embolism or deep venous thrombosis
PS   Disclosure; Fig 4; 18pp; German.
CC   The sequence is used in the construction of pFC44 (WO9004023),
CC   contg. the wild-type proUK gene and a tetracycline resistance gene.
CC   This gene can be modified by usual methods of oligonucleotide-
CC   directed mutagenesis to encode for the variants represented in
CC   R20536-38.  The prod. is converted to amidated form by treatment
CC   with an amidating enzyme, specifically peptidylglycine alpha-
CC   amidating monooxygenase.
CC   See also Q20360, Q20754-58 and R20536-38.
SQ   Sequence   2377 BP;    631 A;    593 C;    604 G;    549 T;

  Query Match          12.6%;  Score 15;  DB 3;  Length 2377;
  Best Local Similarity 100.0%;  Pred. No. 7.94e+00;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```



```
Db   1726 ttgagaaatgaataa 1740
          |||||||||||||||
Qy     12 TTGAGAAATGAATAA 26


RESULT   27
ID   N90726 standard; DNA; 2377 BP.
AC   N90726;
DT   19-JUN-1990  (first entry)
DE   DNA sequence of extracellular neutral protease (vibriolysin) gene from
DE   Vibrio proteolyticus ATCC 53559
KW   Extracellular neutral protease gene; vibriolysin gene;
KW   Vibrio proteolyticus; ATCC 53559.
OS   Vibrio proteolyticus.
FH   Key             Location/Qualifiers
FT   cds              3..35
FT                    /*tag=  a
FT                    /note="p93568"
FT   cds              39..68
FT                    /*tag=  b
FT                    /note="p93569"
FT   cds              72..86
FT                    /*tag=  c
FT                    /note="p93570"
FT   cds              90..104
FT                    /*tag=  d
FT                    /note="p93571"
FT   cds              108..116
FT                    /*tag=  e"
FT   cds              120..155
FT                    /*tag=  f
FT                    /note="p93573"
FT   cds              159..200
FT                    /*tag=  g
FT                    /note="p93575"
FT   cds              204..215
FT                    /*tag=  h
FT                    /note="p93576"
FT   cds              219..244
FT                    /*tag=  i
FT                    /note="p93577"
FT   cds              248..2078
FT                    /*tag=  j
FT                    /note="p93578"
FT   cds              2082..2111
FT                    /*tag=  k
FT                    /note="p93579"
FT   cds              2115..2144
FT                    /*tag=  l
FT                    /note="p93580"
```

```
FT   cds             2148..2150
FT                   /*tag= m
FT   cds             2154..2240
FT                   /*tag= n
FT                   /note="p93582"
FT   cds             2244..2309
FT                   /*tag= o
FT                   /note="p93583"
FT   cds             2313..2317
FT                   /*tag= p
FT   cds             2321..2324
FT                   /*tag= q
FT   cds             2328..2377
FT                   /*tag= r
FT                   /note="p93586"
PN   EP-309879-A.
PD   05-APR-1989.
PF   21-AUG-1988; 115439.
PR   19-NOV-1987; US-123038, US-103983.
PA   (GRAC) Grace WR Co.
PI   Deutch AH, David VA;
DR   WPI; 89-101218/14.
DR   P-PSDB; P93568, p93569, p93570, p93571, p93573, p93575, p93576, p93577,
DR   p93578, p93579, p93580, p93582, p93583, p93586.
PT   Cloning and expression of neutral protease genes -
PT   using DNA encoding enzymes of Vibrio proteolyticus or Bacillus
PT   species in gram-negative microorganisms
PS   Figure 1; 16pp; English.
CC   Figure 1 gives the translation of the entire sequence, hence the large
CC   number of CDS, which are separated by stop codons. Recombinant DNA
CC   contg. a Vibrio nuclear protease enzyme (NPE) gene is claimed as are
CC   gram-negative microorganisms contg. DNA coding from more than one NPE of
CC   Vibrio proteolyticus. The microorganisms are pref. E.coli or Serratia
CC   species, esp. E.coli ATCC 67499 or 67501. The gram-negative
CC   microorganisms synthesise active protease enzyme. The genes can be
CC   manipulated for the overprodn. of NPE.
SQ   Sequence   2377 BP;   646 A;   527 C;   585 G;   618 T;   1 Others;

  Query Match           12.6%;  Score 15;  DB 1;  Length 2377;
  Best Local Similarity 100.0%;  Pred. No. 7.94e+00;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    245 ttgagaaatgaataa 259
          ||||||||||||||| 
Qy     12 TTGAGAAATGAATAA 26


RESULT   28
ID   Q04107 standard; cDNA; 2427 BP.
AC   Q04107;
DT   12-SEP-1990  (first entry)
```



```
DE   Human pro-urokinase cDNA of clone pcUK176..
KW   Non-glycosylated; pro-urokinase; E. coli; Ptrp promoter; MS-2 RBS; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   polya_site      2294..3082
FT                   /*tag= a
FT   cds             7..1363
FT                   /*tag= b
FT                   /product=proUK
FT   polya_signal    2261..2266
FT                   /*tag= c
FT                   /number=1
FT   polya_signal    2274..2279
FT                   /*tag= d
FT                   /number=2
PN   EP-365894-A.
PD   02-MAY-1990.
PF   06-OCT-1989; 110586.
PR   11-OCT-1988; GB-023833.
PA   (FARM) Farmitalia Carlo  ERBA SPA.
PI   Brandazza A, Sarmientos P, Orsini G;
DR   WPI; 90-133447/18.
DR   P-PSDB; R04253.
PT   Non-glycosylated pro-urokinase prodn. - using E.coli B strains and E.coli
PT   promoter Ptrp and Shine-Dalgarno sequence MS-2.
PS   Disclosure; p; English.
CC   SER residue encoded by the bases AGC at position 133-135 is the start of
CC   the mature proUK. cDNA is pref. obtained from mRNA of HEp-3 epidermoid
CC   carcinoma cells. Non-glycosylated proUK (MW 45kD) is produced by E.coli B
CC   strain containing the sequence..
CC   See also Q04101-07.
SQ   Sequence   2427 BP;   680 A;   593 C;   605 G;   549 T;

  Query Match           12.6%;  Score 15;  DB 1;  Length 2427;
  Best Local Similarity 100.0%;  Pred. No. 7.94e+00;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1727 ttgagaaatgaataa 1741
          ||||||||||||||| 
Qy     12 TTGAGAAATGAATAA 26


RESULT   29
ID   T51124 standard; DNA; 2919 BP.
AC   T51124;
DT   18-AUG-1997  (first entry)
DE   Human homeoprotein regulator of insulin HoxB13 gene.
KW   Antibody; diabetes; breast cancer; insulin dependent diabetes;
KW   hypoglycaemia; hyperglycaemia; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
```



```
FT   tata_signal     484..489
FT                   /*tag= a
FT                   /note= "Candidate for the TATA box"
FT   intron          1..582
FT                   /*tag= b
FT                   /number= 1
FT   exon            583..1183
FT                   /*tag= c
FT                   /number= 1
FT                   /transl_except= (pos: 790..792, aa: Ala)
FT   misc_feature    790..792
FT                   /*tag= d
FT                   /note= "Ala in specification"
FT   intron          1184..2132
FT                   /*tag= e
FT                   /number= 2
FT   exon            2133..2919
FT                   /*tag= f
FT                   /number= 2
FT   misc_feature    2174..2356
FT                   /*tag= g
FT                   /label= Encodes_homeodomain
PN   WO9636711-A2.
PD   21-NOV-1996.
PF   09-MAY-1996; U06608.
PR   09-MAY-1995; US-437607.
PA   (SALK ) SALK INST.
PA   (STRA-) STRANG CANCER PREVENTION CENT.
PI   Leonard JN,  Montminy MR;
DR   WPI; 97-012086/01.
DR   P-PSDB; W11259.
PT   New homeo:protein regulator of insulin gene expression - and related
PT   DNA and antibodies, useful for detecting e.g diabetes or breast
PT   cancer
PS   Claim 14; Fig 2A; 96pp; English.
CC   The present sequence encodes the human HoxB13, homeoprotein
CC   regulator of insulin gene expression. The novel homeoprotein regulator
CC   of insulin gene expression is a protein (or its active fragments,
CC   agonists and/or mimics) which binds the FLAT element of an insulin gene
CC   promoter; which is modulated by a Ca2+-dependent CaM kinase IV; and
CC   which is homologous to a sequence encoded by a Hox gene complex.
CC   Detection of altered levels of this protein indicates a disease related
CC   to glucose homeostasis, particularly (non-)insulin dependent diabetes,
CC   hypo- or hyper- glycaemia, or also breast cancer. Antibodies are useful
CC   as reagents for immunoassays. The protein and compounds that promote and
CC   inhibit its production or activity or its specific binding partners, are
CC   useful for preventing or treating the specified diseases, especially
CC   where insulin is also being administered.
SQ   Sequence   2919 BP;   645 A;   739 C;   912 G;   621 T;

  Query Match           12.6%;  Score 15;  DB 30;  Length 2919;
```



```
  Best Local Similarity 100.0%;  Pred. No. 7.94e+00;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2792 aagctttctctcaga 2806
          ||||||||||||||| 
Qy     25 AAGCTTTCTCTCAGA 39


RESULT   30
ID   T33537 standard; DNA; 12412 BP.
AC   T33537;
DT   15-FEB-1998  (first entry)
DE   BCG deletion region 3 and flanking sequences.
KW   BCG delta 3; virulence; avirulence; attenuation; gene deletion;
KW   mycobacteria; vaccine; infection; marker; ss.
OS   Mycobacterium bovis strain BCG.
FH   Key             Location/Qualifiers
FT   misc_feature    1406..10673
FT                   /*tag= a
FT                   /note= "BCG delta 1 deletion region"
PN   WO9625519-A1.
PD   22-AUG-1996.
PF   15-FEB-1996; U01938.
PR   17-FEB-1995; US-390878.
PA   (PATH-) PATHOGENESIS CORP.
PI   Mahairas GG,  Stover CK;
DR   WPI; 96-393419/39.
PT   Detecting markers for avirulence in Mycobacterium - used in
PT   production of vaccines against bacterial infection, and to detect
PT   bacterial infection
PS   Example 1; Fig 3; 66pp; English.
CC   This DNA sequence comprises Mycobacterium bovis BCG deletion
CC   sequence BCGdelta3.  A specific genetic deletion of this region
CC   results in an avirulence phenotype of the mycobacterium.  2 Other
CC   deletion regions (see T33535 and T33536) have also been detected.
CC   Identification involved screening a BCG cosmid library with a
CC   radiolabeled probe obtained following DNA subtraction between
CC   virulent Mycobacterium tuberculosis H37Rv and avirulent BCG.
CC   The deletions provide useful markers for the identification of an
CC   avirulent, or a virulent, mycobacterial phenotype.  Determination
CC   of avirulence requires the detection of the presence or absence of
CC   the deletion; the deletions are detected either by detecting the
CC   presence or absense of deletion junctions (see T33538-46), or by
CC   detecting the presence or absence of the sequences contained within
CC   the deletion.  Deletion polypeptides are used as components of
CC   immunological assays and in vaccines.
SQ   Sequence   12412 BP;   2146 A;   4273 C;   3876 G;   1917 T;

  Query Match           12.6%;  Score 15;  DB 35;  Length 12412;
  Best Local Similarity 100.0%;  Pred. No. 7.94e+00;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```

```
Db    7370 gcttttcggcgctcc 7384
           ||||||||||||||||
Cp     105 GCTTTTCGGCGCTCC 91


RESULT   31
ID   T44575 standard; mRNA; 18 BP.
AC   T44575;
DT   04-FEB-1997  (first entry)
DE   Antisense oligo HUPA-1728 targetting human UPA mRNA.
KW   antisense; protease; tumour; metastasis; inhibition; breast cancer;
KW   phosphorothioate; cell invasion; urokinase; human; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   modified_base    1..18
FT                    /*tag= a
FT                    /note= "phosphorothioate backbone"
PN   US5552390-A.
PD   03-SEP-1996.
PF   09-DEC-1993; 164200.
PR   09-DEC-1993; US-164200.
PA   (UYNE-) UNIV NEBRASKA.
PI   Iversen PL,  Scholar EM;
DR   WPI; 96-412096/41.
PT   Anti:sense oligo:nucleotide(s) to murine urokinase mRNA - for
PT   inhibiting tumour cell invasion and metastasis
PS   Claim 1; Column 9-10; 14pp; English.
CC   T44560-77 are antisense phosphorothioate oligonucleotides (PONs)
CC   complementary to the mRNA of protease enzymes involved in tumour
CC   invasion. Urokinase (UK) is important in metastasis of tumours and PONs
CC   to urokinase plasminogen activator (UPA) have demonstrated in vivo and in
CC   vitro reduction of metastasis of breast cell cancer. The effect of
CC   several PONs on 410.4 cell invasion through matrigel coated membranes was
CC   tested. The present sequence, which is complementary to the mRNA of
CC   human UPA (position 1728) was tested for effects on the invasion of
CC   the cell line. The inhibitory effect on invasion was specific to the
CC   sequence of the urokinase type plasminogen activator mRNA.
SQ   Sequence   18 BP;    5 A;    5 C;    0 G;    8 T;

  Query Match           11.8%;  Score 14;  DB 24;  Length 18;
  Best Local Similarity 100.0%;  Pred. No. 3.46e+01;
  Matches   14;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 tattcatttctcaa 14
          ||||||||||||||
Cp     25 TATTCATTTCTCAA 12


RESULT   32
ID   T02529 standard; DNA; 32 BP.
```



```
AC   T02529;
DT   10-JUN-1996  (first entry)
DE   PenI gene upstream primer CLC107.
KW   Antibody; penI; fusion protein; T cell; C9; PCR; primer; amplify;
KW   polymerase chain reaction; GAP-N; SH2; human; immunisation;
KW   immune response; proteolytic cleavage; ss.
OS   Synthetic.
PN   WO9530016-A1.
PD   09-NOV-1995.
PF   01-MAY-1995; U05361.
PR   29-APR-1994; US-235178.
PA   (DADE-) DADE INT INC.
PI   Casipit C,  Chavaillas P,  Tal R,  Wittman V,  Wong HC;
DR   WPI; 95-393087/50.
PT   Target peptide-penI protein fusion polypeptide - purified by
PT   reducing ionic strength of soln. to precipitate the fusion
PT   polypeptide
PS   Example 5; Page 31; 67pp; English.
CC   T02529-T02532 represent amplification primers used in the construction of
CC   a GAP-N SH2 domain:penI fusion protein.  This sequence and T02530 amplify
CC   the penI sequence, while T02531 and T02532 amplify the GAP-N SH2 domain.
CC   The fusion protein is provided in solution and is contacted with an
CC   aqueous solution with sufficient ionic strength to precipitate the
CC   protein (preferably a buffer solution with a potassium chloride salt
CC   concentration of 200 mM or less).  The fusion protein is cleaved to
CC   provide the target polypeptide.  This method is useful for the synthesis
CC   and purification of any target peptide.  The fusion protein can be used
CC   to immunise a mammal against a specific antigen, by eliciting an immune
CC   response.  Also, for fine epitope mapping of a peptide by determining the
CC   binding activity of the fusion protein to the binding domain of a peptide
CC   (preferrably an antibody peptide).  The method can be used for selection
CC   of a target peptide, and for the identification of a peptide that
CC   modulates the activity of T cells.  The advantage of this method, is that
CC   it allows the purification of the fusion protein in a single step.  Also,
CC   for certain target peptides, the ability to induce or enhance T cell
CC   activity may be enhanced when the peptide is fused to penI.
SQ   Sequence   32 BP;    17 A;    4 C;    5 G;    6 T;

  Query Match           11.8%;  Score 14;  DB 18;  Length 32;
  Best Local Similarity 100.0%;  Pred. No. 3.46e+01;
  Matches   14;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 gagagaaagcttat 14
          ||||||||||||||
Cp     36 GAGAGAAAGCTTAT 23


RESULT   33
ID   T34332 standard; DNA; 35 BP.
AC   T34332;
DT   04-OCT-1996  (first entry)
```



```
DE   3' glaA non-coding sequence amplification primer, AB2155.
KW   Polymerase chain reaction; primer; amplify; PCR; acetamidase gene;
KW   amdS; Aspergillus nidulans; selection; transformation; glaA;
KW   filamentous fungi; marker gene; antibiotic; selection marker;
KW   glyceraldehyde-3-phosphate dehydrogenase; gpdA; amyloglucosidase; ss.
OS   Synthetic.
PN   EP-635574-A1.
PD   25-JAN-1995.
PF   30-JUN-1994; 201896.
PR   23-JUL-1993; EP-202195.
PA   (KONN ) GIST-BROCADES NV.
PI   Selten GCM,  Swinkels BW,  Van Gorcom RFM;
DR   WPI; 95-053686/08.
PT   Selection marker gene free recombinant strains, esp. filamentous
PT   fungi, and methods for obtaining them  - for improved selection
PT   without use of antibiotics and with no undesired residual marker DNA
PT   following transformation
PS   Example 1; Page 14; 109pp; English.
CC   The sequences given in T34331-36 are primers which were used in the
CC   construction of the glaA gene replacement vector pGBDEL4L.  pGBDEL4L
CC   contains the 5' part of the A. niger amyloglucosidase (glaA) promoter
CC   region, a synthetic DNA sequence of 16 bp providing stop codons in all
CC   three reading frames, the A. nidulans acetamidase (amdS) gene under
CC   control of the A. nidulans glyceraldehyde-3-phosphate dehydrogenase
CC   (gpdA) promoter, flanked on both sides by 3' glaA non-coding sequences.
CC   The resultant vector was used in the method of the invention for the
CC   selection of transformed filamentous fungi from which the marker gene
CC   has been deleted.  This selection system reduces the need for antibiotic
CC   selection markers which give rise to an undesired spread of strains that
CC   have become resistant to antibiotics.
SQ   Sequence   35 BP;    9 A;    8 C;    7 G;    11 T;

  Query Match           11.8%;  Score 14;  DB 21;  Length 35;
  Best Local Similarity 100.0%;  Pred. No. 3.46e+01;
  Matches   14;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      8 gaaagcttattcat 21
          ||||||||||||||
Cp     32 GAAAGCTTATTCAT 19


RESULT   34
ID   T26260 standard; cDNA to mRNA; 62 BP.
AC   T26260;
DT   10-OCT-1996  (first entry)
DE   Human gene signature HUMGS08499.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
```



```
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 2041; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   62 BP;    29 A;    6 C;    12 G;    15 T;

  Query Match           11.8%;  Score 14;  DB 21;  Length 62;
  Best Local Similarity 93.3%;  Pred. No. 3.46e+01;
  Matches   14;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db     41 aaaagctccctaaaa 55
          ||||||||| |||||
Qy    100 AAAAGCTCCNTAAAA 114


RESULT   35
ID   Q34036 standard; DNA; 80 BP.
AC   Q34036;
DT   02-FEB-1993  (first entry)
DE   Downstream sequence of microsatellite from clone TGLA431.
KW   PCR; selection; primers; OPTIPRIM; breeding; cattle; parentage;
KW   genetic mapping; traits; amplification; ss.
OS   Bos taurus.
PN   WO9213102-A.
PD   06-AUG-1992.
PF   15-JAN-1992; U00340.
PR   15-JAN-1991; US-642342.
PA   (GENM-) GENMARK.
```

```
PI   Georges M, Massey JM;
DR   WPI; 92-284684/34.
PT   Polymorphic bovine DNA markers - used in genetic identification,
PT   gene mapping, and selective breeding
PS   Table 7; Page 345; 517pp; English.
CC   The sequence is that downstream of a bovine microsatellite sequence
CC   obtd. by screening a library of bovine MboI DNA fragments of between
CC   250 and 500 bp with an (AC)15 and a (TC)15 oligonucleotide probe.
CC   One out of 50 clones cross-hybridised.  Assuming independent
CC   distribution of microsatellites and MboI sites, the frequency of
CC   (T6)n >9 microsatellites in the bovine genome is estimated at >100,
CC   000.  The sequence information for ca. 230 such bovine microsatellites
CC   is summarised in the specification and indexed herein (see below).
CC   The sequences upstream and downstream of the microsatellite sequence
CC   were used to generate the required PCR primers for in vitro
CC   amplification of the corr.esp. microsatellite (using the program
CC   OPTIPRIM).  The microsatellites may be used to identify individuals,
CC   for parentage testing, and in the genetic mapping of economic trait
CC   loci, or genes involved the determinism of economically important
CC   traits esp. in cattle, to allow selective breeding.
CC   See also Q33501-34437.
SQ   Sequence   80 BP;    26 A;    16 C;    10 G;    28 T;

  Query Match           11.8%;  Score 14;  DB 5;  Length 80;
  Best Local Similarity 100.0%;  Pred. No. 3.46e+01;
  Matches   14;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     13 tcccatttattcct 26
          ||||||||||||||
Cp     93 TCCCATTTATTCCT 80


RESULT   36
ID   T21050 standard; cDNA to mRNA; 129 BP.
AC   T21050;
DT   03-JUL-1996  (first entry)
DE   Human gene signature HUMGS02333.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
```



```
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 789; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   129 BP;    35 A;    15 C;    12 G;    60 T;

  Query Match           11.8%;  Score 14;  DB 18;  Length 129;
  Best Local Similarity 93.3%;  Pred. No. 3.46e+01;
  Matches   14;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db      6 aaagcttattcnttt 20
          |||||||||||| |||
Cp     31 AAAGCTTATTCATTT 17


RESULT   37
ID   T07068 standard; cDNA; 216 BP.
AC   T07068;
DT   05-JUL-1996  (first entry)
DE   Immunogen DNA from n-(ABCDE) hepatitis virus.
KW   Non-A, non-B, non-C, non-D, non-E hepatitis virus; n-(ABCDE);
KW   My 131 clone 430-4-8; immunogen; antibody; vaccine; phage library;
KW   ds.
OS   Non-A, non-B, non-C, non-D, non-E hepatitis virus.
PN   WO9532290-A2.
PD   30-NOV-1995.
PF   17-MAY-1995; U05980.
PR   20-MAY-1994; US-246986.
PA   (GENE-) GENELABS TECHNOLOGIES INC.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI   Kim JP,  Purcell RH;
DR   WPI; 96-020585/02.
PT   New non-A, -B, -C, -D and -E (n-(ABCDE)) hepatitis DNA libraries -
PT   used to develop prods. for the detection, diagnosis, prevention and
PT   treatment of n-(ABCDE) hepatitis.
PS   Disclosure; Page 91; 165pp; English.
CC   The sequence represents clone 430-4-8 which encodes an immunogenic
```



```
CC   polypeptide associated with non-A, non-B, non-C, non-D, non-E
CC   hepatitis virus (n-(ABCDE) virus) infection; and is obtained
CC   by preparing a phage from a phage lambda gt11 library of My 190 DNA
CC   (ATCC 75284) n-(ABCDE) hepatitis serum, plating to form plaques,
CC   and screening the phage plaques for the production of polypeptides
CC   immunoreactive with n-(ABCDE) serum.  n-(ABCDE) hepatitis virus
CC   polypeptides can be used for the production or detection of
CC   antibodies, and in vaccines.  The antibodies can be used for
CC   detection, diagnosis and in passive immunotherapy.  The DNA can
CC   also be used in detection and diagnosis, and as hybridisation
CC   probes for identification of further n-(ABCDE) hepatitis coding
CC   sequences.  Culture systems producing the n-(ABCDE) polypeptides
CC   can be used in screening studies.
SQ   Sequence   216 BP;    58 A;    60 C;    54 G;    44 T;

  Query Match           11.8%;  Score 14;  DB 19;  Length 216;
  Best Local Similarity 100.0%;  Pred. No. 3.46e+01;
  Matches   14;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    168 cgccgaaaagctcc 181
          ||||||||||||||
Qy     95 CGCCGAAAAGCTCC 108


RESULT   38
ID   Q59885 standard; cDNA; 281 BP.
AC   Q59885;
DT   16-MAR-1994  (first entry)
DE   Human brain Expressed Sequence Tag EST00783.
KW   Gene transcription product; genetic markers; tagging; in vivo;
KW   transcription; mapping; locations; chromosomes; chromosomal; ss.
OS   Homo sapiens.
PN   WO9316178-A.
PD   19-AUG-1993.
PF   12-FEB-1993; U01294.
PR   12-FEB-1992; US-837195.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
PI   Adams MD, Moreno RF, Venter CJ;
DR   WPI; 93-272882/34.
PT   Enriched oligonucleotides and corresp. sequences - used as
PT   markers for human genes transcribed in-vivo, facilitate tagging
PT   of most human genes
PS   Example 4; Page 247; 500pp; English.
CC   The Expressed Sequence Tag was isolated from a human brain cDNA
CC   library as part of a large set of ESTs which can be used as markers
CC   for human genes transcribed in vivo. They can be used to facilitate
CC   tagging of most human genes, for mapping locations of expressed genes
CC   on chromosomes, for individual or forensic identification, for mapping
CC   locations of disease-associated genes, for identification of tissue
CC   type, and for prepn. of antisense sequences, probes and constructs.
CC   EST00783 has a "poor" coding probability as evaluated using the
```



```
CC   coding-region prediction program CRM. See also Q59041-Q61440.
SQ   Sequence   281 BP;    57 A;    50 C;    48 G;    119 T;

  Query Match           11.8%;  Score 14;  DB 8;  Length 281;
  Best Local Similarity 100.0%;  Pred. No. 3.46e+01;
  Matches   14;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    184 tgaataagctttct 197
          ||||||||||||||
Qy     20 TGAATAAGCTTTCT 33


RESULT   39
ID   T19568 standard; cDNA to mRNA; 291 BP.
AC   T19568;
DT   27-JUN-1996  (first entry)
DE   Human gene signature HUMGS00633.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 420; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   291 BP;    95 A;    60 C;    61 G;    71 T;
```

```
  Query Match          11.8%;  Score 14;  DB 18;  Length 291;
  Best Local Similarity 100.0%;  Pred. No. 3.46e+01;
  Matches   14;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    238 agctttctctcaga 251
          ||||||||||||||
Qy     26 AGCTTTCTCTCAGA 39


RESULT   40
ID   Q14619 standard; cDNA; 522 BP.
AC   Q14619;
DT   29-JAN-1992  (first entry)
DE   Fragment 83 from human pancreatitis associated protein gene.
KW   hPAP; pancreas; imaging; ss.
OS   Homo sapiens.
PN   WO9116428-A.
PD   31-OCT-1991.
PF   18-APR-1991; F00323.
PR   20-APR-1990; FR-005062.
PA   (INRM ) INSERM INST NAT SANTE.
PI   Iovanna JL, Keim V, Dagron JC;
DR   WPI; 91-339816/46.
DR   P-PSDB; R14795.
PT   cDNA fragment of human pancreatitis associated protein -
PT   produced by genetic engineering, and the corresponding antibody,
PT   for use in the diagnosis of pancreatitis
PS   Claim 3; Page 26; 49pp; French.
CC   This cDNA fragment was obtained by screening a human pancreatic cDNA
CC   library by hybridisation with a rat PAP cDNA probe, followed by a
CC   second screening of positive clones with a PSP cDNA probe (PSP is a
CC   protein with some structural similarities with rat PAP) and then
CC   recovering clones not reactive with the PSP probe and isolating cDNA
CC   fragments from them. Fragment S3 was cloned into pEX and the
CC   resultant recombinant transformed into E.coli. The expression
CC   product was used to generate antibodies for in vitro diagnosis of
CC   acute pancreatitis and for imaging the pancreas.
SQ   Sequence   522 BP;    113 A;    143 C;    141 G;    125 T;

  Query Match          11.8%;  Score 14;  DB 3;  Length 522;
  Best Local Similarity 100.0%;  Pred. No. 3.46e+01;
  Matches   14;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    457 acagcatttctgag 470
          ||||||||||||||
Cp     47 ACAGCATTTCTGAG 34


RESULT   41
ID   Q64172 standard; cDNA; 528 BP.
AC   Q64172;
```



```
DT   01-FEB-1995  (first entry)
DE   Mouse PAP.
KW   Pancreatitis-Associated Protein; PAP; mouse; human; bacteria;
KW   aggregation; antibacterial; primer; PCR; ss.
OS   Mus musculus.
FH   Key             Location/Qualifiers
FT   cds             1..525
FT                   /*tag=  a
PN   J06135998-A.
PD   17-MAY-1994.
PF   22-OCT-1992; 284765.
PR   22-OCT-1992; JP-284765.
PA   (SHIO ) SHIONOGI & CO LTD.
DR   WPI; 94-197103/24.
DR   P-PSDB; R54098.
PT   Human and mouse pancreatitis-associated protein (PAP) - useful
PT   for prepn. of proteinaceous antibacterial agent
PS   Claim 7-12; Page 6-7; 7pp; Japanese.
CC   PAP DNA has a high bacteria-aggregating activity and is useful for
CC   the prepn. of PAP which can act as a proteinaceous antibacterial
CC   agent.  Total RNA was isolated from mouse.  It was hybridised with a
CC   modified salmon spermatozoa DNA and known rat PAP cDNA as a probe.
CC   Two primers are given in Q64173-74.
SQ   Sequence   528 BP;    114 A;    144 C;    143 G;    127 T;

  Query Match          11.8%;  Score 14;  DB 11;  Length 528;
  Best Local Similarity 100.0%;  Pred. No. 3.46e+01;
  Matches   14;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    457 acagcatttctgag 470
          ||||||||||||||
Cp     47 ACAGCATTTCTGAG 34


RESULT   42
ID   T43460 standard; DNA; 593 BP.
AC   T43460;
DT   08-AUG-1997  (first entry)
DE   ATM gene exon 20.
KW   ATM gene; ataxia-telangiectasia; progressive genetic disorder; antibody;
KW   central nervous system; immune system; chromosomal instability; therapy;
KW   cancer predisposition; radiation sensitivity; cell cycle abnormality;
KW   multi-system disease; autosomal recessive; cerebellar ataxia; cerebellum;
KW   general motor dysfunction; Purkinje cell; oculocutaneous telangiectasia;
KW   blood vessel; bulbar conjunctiva; facial skin; A-T; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   intron          1..197
FT                   /*tag=  a
FT                   /number= 19
FT   exon            198..397
```



```
FT                   /*tag=  b
FT                   /number= 20
FT   intron          398..593
FT                   /*tag=  c
FT                   /number= 20
PN   WO9636691-A1.
PD   21-NOV-1996.
PF   16-MAY-1996; U07025.
PR   16-MAY-1995; US-441822.
PR   08-APR-1996; US-629001.
PA   (KOHN/) KOHN K I.
PA   (UYRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.
PI   Shiloh Y;
DR   WPI; 97-012070/01.
PT   New isolated ataxia-telangiectasia gene - used to develop prods. for
PT   the study, diagnosis and treatment of ataxia-telangiectasia.
PS   Claim 1; Page 61-62; 153pp; English.
CC   T43444-T43496 represent exons of the ATM gene of the invention.
CC   Ataxia-telangiectasia (A-T) is caused by mutations, insertions, or
CC   deletions in the coding region of the ATM gene. A-T is a progressive
CC   genetic disorder affecting the central nervous and immune systems. A-T
CC   involves chromosomal instability, cancer predisposition, radiation
CC   sensitivity, and cell cycle abnormalities. A-T is a multi-system disease
CC   inherited in an autosomal recessive manner. Cerebellar ataxia that
CC   gradually develops into general motor dysfunction is the first clinical
CC   hallmark of A-T, and results from progressive loss of Purkinje cells in
CC   the cerebellum. Oculocutaneous telangiectasia (dilation of blood vessels)
CC   develops in the bulbar conjunctiva and facial skin, and is later
CC   accompanied by graying of the hair and atrophic changes in the skin. The
CC   co-occurrence of cerebellar ataxia and telangiectases in the conjunctivae
CC   and occasionally on the facial skin (the second early hallmark of A-T)
CC   usually establishes the differential diagnosis of A-T from other
CC   cerebellar ataxias. The ATM gene, can be used in methods for detecting
CC   carriers of a defective gene that causes A-T. The gene can also be used
CC   to generate antibodies. The antibodies and methods can be used in the
CC   study, diagnosis and therapy of A-T.
SQ   Sequence   593 BP;    176 A;    85 C;    111 G;    221 T;

  Query Match          11.8%;  Score 14;  DB 31;  Length 593;
  Best Local Similarity 100.0%;  Pred. No. 3.46e+01;
  Matches   14;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    509 tcttgagaaatgaa 522
          ||||||||||||||
Qy     10 TCTTGAGAAATGAA 23


RESULT   43
ID   Q69201 standard; DNA; 797 BP.
AC   Q69201;
DT   21-FEB-1995  (first entry)
```



```
DE   Human Pancreatitis-Associated Protein coding sequence.
KW   Human PAP; pancreatitis-associated protein; diagnosis; detection;
KW   cystic fibrosis transmembrane conductance regulator; CFTR gene;
KW   mucoviscidosis; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             43..570
FT                   /*tag=  a
FT                   /product= PAP
FT   polya_signal    772..777
FT                   /*tag=  b
PN   WO9415218-A.
PD   07-JUL-1994.
PF   23-DEC-1993; F01299.
PR   24-DEC-1992; FR-015730.
PA   (INRM ) INST NAT SANTE & RECH MEDICALE.
PI   Dagorn J,  Iovanna J,  Keim V,  Sarles J;
DR   WPI; 94-234870/28.
DR   P-PSDB; R57117.
PT   In vitro detection of pancreatic disease associated with CTFR
PT   gene mutation - by measuring concn. of pancreatitis associated
PT   protein in blood etc., suitable for neonatal diagnosis of, e.g.
PT   cystic fibrosis
PS   Claim 1; Fig 3; 33pp; French.
CC   Antibodies which recognise the previously isolated and sequenced
CC   human Pancreatitis-Associated protein (PAP) are claimed. They are
CC   useful for measuring the amount of PAP in a sample to detect
CC   pancreatic disease associated with heterozygotic mutations in the
CC   CTFR gene. Such conditions result in expression of PAP at levels
CC   at least 2-3 times higher than normal.
SQ   Sequence   797 BP;    180 A;    213 C;    198 G;    206 T;

  Query Match          11.8%;  Score 14;  DB 11;  Length 797;
  Best Local Similarity 100.0%;  Pred. No. 3.46e+01;
  Matches   14;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    499 acagcatttctgag 512
          ||||||||||||||
Cp     47 ACAGCATTTCTGAG 34


RESULT   44
ID   Q14623 standard; DNA; 798 BP.
AC   Q14623;
DT   29-JAN-1992  (first entry)
DE   Fragment S4 from human pancreatitis associated protein gene.
KW   hPAP; pancreas; imaging; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             43..567
FT                   /*tag=  a
```

```
FT                      /product= human PAP
PN   WO9116428-A.
PD   31-OCT-1991.
PF   18-APR-1991; F00323.
PR   20-APR-1990; FR-005062.
PA   (INRM ) INSERM INST NAT SANTE.
PI   Iovanna JL, Keim V, Dagron JC;
DR   WPI; 91-339816/46.
DR   P-PSDB; R14795.
PT   cDNA fragment of human pancreatitis associated protein -
PT   produced by genetic engineering, and the corresponding antibody,
PT   for use in the diagnosis of pancreatitis
PS   Claim 4; Fig 4; 49pp; French.
CC   This cDNA fragment was obtained by screening a human pancreatic cDNA
CC   library by hybridisation with a rat PAP cDNA probe, followed by a
CC   second screening of positive clones with a PSP cDNA probe (PSP is a
CC   protein with some structural similarities with rat PAP) and then
CC   recovering clones not reactive with the PSP probe and isolating cDNA
CC   fragments from them. This sequence includes fragment S3 (see
CC   Q14619) which corresponds to the CDS.
SQ   Sequence   798 BP;     181 A;     213 C;     198 G;     206 T;

  Query Match           11.8%;  Score 14;  DB 3;  Length 798;
  Best Local Similarity 100.0%;  Pred. No. 3.46e+01;
  Matches    14;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    499 acagcatttctgag 512
          ||||||||||||||
Cp     47 ACAGCATTTCTGAG 34


RESULT   45
ID   T29435 standard; cDNA; 1044 BP.
AC   T29435;
DT   27-OCT-1996  (first entry)
DE   Mouse pancreas beta-cell line MIN6 galanin receptor cDNA sequence.
KW   Galanin receptor; stomach ulcer; antiulcer; diabetes; antidiabetic;
KW   dementia; neuropeptide; sedative; Alzheimer's disease; obesity; ds.
OS   Mus musculus.
FH   Key             Location/Qualifiers
FT   mat_peptide     1..1044
FT                   /*tag= a
PN   EP-711830-A2.
PD   15-MAY-1996.
PF   11-OCT-1995; 115996.
PR   13-OCT-1994; JP-247599.
PR   28-DEC-1994; JP-326610.
PR   31-MAY-1995; JP-134412.
PA   (TAKE ) TAKEDA CHEM IND LTD.
PI   Fujii R,  Fukusumi S,  Hinuma S,  Hosoya M,  Ohgi K;
PI   Ohtaki T,  Onda H;
```



```
DR   WPI; 96-232095/24.
DR   P-PSDB; R95069.
PT   Isolated galanin receptor proteins - used partic. to identify
PT   agonists or antagonists, which can be used to treat, e.g. ulcers,
PT   diabetes or dementia
PS   Claim 1; Page 48; 71pp; English.
CC   This sequence encoding a mouse galanin receptor protein was obtained
CC   from the mouse pancreatic beta-cell line, MIN6 (FERM BP-4954), and
CC   may be expressed recombinantly, using a vector, in e.g. CHO cells.
CC   The encoded protein is particularly used to identify agonists or
CC   antagonists which can be used to treat e.g. ulcers, diabetes or
CC   dementia, etc.  The protein can also be used for the detection of
CC   galanin, in the production of antibodies, and transgenic animals.
SQ   Sequence   1044 BP;     216 A;     325 C;     256 G;     247 T;

  Query Match           11.8%;  Score 14;  DB 22;  Length 1044;
  Best Local Similarity 100.0%;  Pred. No. 3.46e+01;
  Matches    14;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    909 ctttctctcagaaa 922
          ||||||||||||||
Qy     28 CTTTCTCTCAGAAA 41


RESULT   46
ID   T13905 standard; cDNA; 1044 BP.
AC   T13905;
DT   27-AUG-1996  (first entry)
DE   Mouse pancreas G-protein coupled receptor protein cDNA pMGR20.
KW   G-protein coupled receptor protein; G-PCRP; agonist; antagonist;
KW   cystic fibrosis; incontinence; diabetes; diagnosis; gene therapy;
KW   ds.
OS   Mus sp.
PN   WO9605302-A1.
PD   22-FEB-1996.
PF   10-AUG-1995; J01599.
PR   11-AUG-1994; JP-189272.
PR   11-AUG-1994; JP-189273.
PR   11-AUG-1994; JP-189274.
PR   30-SEP-1994; JP-236356.
PR   30-SEP-1994; JP-236357.
PR   02-NOV-1994; JP-270017.
PR   28-DEC-1994; JP-326611.
PR   20-JAN-1995; JP-007177.
PR   16-MAR-1995; JP-057186.
PR   19-APR-1995; JP-093989.
PA   (TAKE ) TAKEDA CHEM IND LTD.
PI   Fujii R,  Fukusumi S,  Hinuma S,  Hosoya M,  Ohgi K;
PI   Ohtaki T;
DR   WPI; 96-139698/14.
DR   P-PSDB; R91229.
```



```
PT   G-protein coupled receptor protein DNA and protein - also methods
PT   for isolating (ant)agonists for treatment of cystic fibrosis,
PT   incontinence and diabetes
PS   Example 10; Page 260-61; 360pp; English.
CC   A cDNA clone (T13905), designated pMFR20, codes for a novel mouse
CC   pancreatic beta-cell strain MIN6-derived G-protein coupled receptor
CC   protein (G-PCRP) (R91229).  It was obtd. by PCR amplification of
CC   MIN6 cDNA using primers (T18347-48) based on the transmembrane
CC   domains of known G-PCRPs.  Such cDNA sequences (see also T18366-72,
CC   and T13901-09) may be used in the prodn. of recombinant G-PCRPs,
CC   in the prepn. of transgenic animals, and also for gene therapy.
SQ   Sequence   1044 BP;     216 A;     325 C;     256 G;     247 T;

  Query Match           11.8%;  Score 14;  DB 21;  Length 1044;
  Best Local Similarity 100.0%;  Pred. No. 3.46e+01;
  Matches    14;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    909 ctttctctcagaaa 922
          ||||||||||||||
Qy     28 CTTTCTCTCAGAAA 41


RESULT   47
ID   T48100 standard; DNA; 1293 BP.
AC   T48100;
DT   31-MAR-1997  (first entry)
DE   Immunogenic type F botulinum toxin polypeptide (aa848-1278) DNA.
KW   Botulinum toxin; neurotoxin; BoBT/F; immunogen; vaccine; botulism;
KW   ds.
OS   Clostridium botulinum type F strain Langeland.
PN   WO9641881-A1.
PD   27-DEC-1996.
PF   12-JUN-1996; G01409.
PR   12-JUN-1995; GB-011909.
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
PI   Elmore MJ,  Mauchline ML,  Minton NP,  Pasechnik VA;
DR   WPI; 97-065467/06.
DR   P-PSDB; W09014.
PT   Immunogenic type F botulinum toxin polypeptide(s) - allows
PT   recombinant vaccine prodn.
PS   Disclosure; Page 20-21; 37pp; English.
CC   A segment (T48100) of Clostridium botulinum type F strain Langeland
CC   genomic DNA codes for the heavy chain (amino acids 848-1278) of the
CC   type F neurotoxin (see also W09014).  A synthetic gene (T48101) for
CC   the heavy chain utilises preferred codons for high level expression
CC   in E. coli.  The heavy chain polypeptides (see also W09015-17) are
CC   free of botulinum toxin activity but can induce protective immunity
CC   to a type F botulinum toxin, making them useful for vaccine prodn.
SQ   Sequence   1293 BP;     539 A;     114 C;     191 G;     449 T;

  Query Match           11.8%;  Score 14;  DB 27;  Length 1293;
```



```
  Best Local Similarity 100.0%;  Pred. No. 3.46e+01;
  Matches    14;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    337 tgtataaggaataa 350
          ||||||||||||||
Qy     74 TGTATAAGGAATAA 87


RESULT   48
ID   T14038 standard; cDNA; 1399 BP.
AC   T14038;
DT   07-JUL-1996  (first entry)
DE   Transcription factor-IIIA gene.
KW   Human; transcription factor-IIIA; hTFIIIA; DNA binding protein;
KW   ribosome; zinc finger; rapid amplification of cDNA ends; 5'-RACE;
KW   primer; PCR; polymerase chain reaction; foetal brain;
KW   anchor primer; diagnostic; probe; transcription control;
KW   antitumour; cancer; therapy; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             20..1291
FT                   /*tag= a
FT                   /product= Transcription factor-IIIA
FT                   /note= 'Claim 2'
FT   misc_feature    317..1096
FT                   /*tag= b
FT                   /product= zinc finger domains
FT   polya_signal    1363..1368
FT                   /*tag= c
PN   EP-704526-A1.
PD   03-APR-1996.
PF   05-SEP-1995; 113908.
PR   05-SEP-1994; JP-211022.
PA   (SAKA ) OTSUKA PHARM CO LTD.
PI   Fujiwara T,  Ozaki K,  Shimada Y,  Shin S,  Takeda S;
DR   WPI; 96-173033/18.
DR   P-PSDB; R91305.
PT   Human Transcription Factor III A gene - useful in regulation of
PT   transcription and for diagnosis and treatment of e.g. cancer related
PT   diseases
PS   Example 1; Page 11-13; 17pp; English.
CC   The sequence encodes human transcription factor-IIIA (hTFIIIA), a
CC   DNA binding protein with 9 zinc finger domains, which is necessary
CC   for the initiation of 5S RNA gene transcription, binding to an
CC   internal control region of the 5S gene.  The coding region
CC   (claimed) is given in T14037.  A fragment lacking a 5'-portion of
CC   the gene has been isolated from a human foetal brain cDNA library
CC   (OTK7-1), and the 5'-portion of the gene has been isolated by
CC   5'-rapid amplification of cDNA ends using primers H11-R (T14039),
CC   H11-E (T14040), H11-H (T14041) and AP-2 (T14042), complementary to
CC   anchor primer T14043.  Reverse transcription using H-11R is
```

```
CC   followed by anchor primer ligation and PCR using AP-2 and H11-E, to
CC   give a full-length cDNA, OTK7.  The gene and its encoded protein
CC   may be used in diagnosis, identification or therapy of hereditary
CC   diseases such as cancer, or other diseases resulting from abnormal
CC   transcriptional control, and to analyse the mechanisms involved in
CC   their activity.
SQ   Sequence   1399 BP;    405 A;    349 C;    354 G;    291 T;

  Query Match           11.8%;  Score 14;  DB 19;  Length 1399;
  Best Local Similarity 100.0%;  Pred. No. 3.46e+01;
  Matches    14;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1335 agcttttctctcaga 1348
          ||||||||||||||
Qy     26 AGCTTTCTCTCAGA 39


RESULT   49
ID   T72167 standard; cDNA to mRNA; 1420 BP.
AC   T72167;
DT   25-FEB-1998  (first entry)
DE   Alzheimer's disease DNA sequence from plasmid pGCS198.
KW   Human; brain; Alzheimer's disease; diagnosis; antibody; expression; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             378..578
FT                   /*tag=  a
FT                   /note= "Protein encoded by DNA from the human brain which
FT                          has different expression from the normal protein
FT                          when in an Alzheimer's disease patient"
FT   CDS             590..702
FT                   /*tag=  b
FT                   /note= "Protein encoded by DNA from the human brain which
FT                          has different expression from the normal protein
FT                          when in an Alzheimer's disease patient"
FT   CDS             1105..1209
FT                   /*tag=  c
FT                   /note= "Protein encoded by DNA from the human brain which
FT                          has different expression from the normal protein
FT                          when in an Alzheimer's disease patient"
PN   WO9721807-A1.
PD   19-JUN-1997.
PF   12-DEC-1996; JO3630.
PR   12-DEC-1995; JP-322745.
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
PI   Hashida H,  Kuga T,  Nakagawa S,  Sakaki Y,  Zhao N;
DR   WPI; 97-332779/30.
DR   P-PSDB; W19939. W19940, W19941.
PT   DNA sequences with varied expression in normal and Alzheimer
PT   patients - used for diagnosis and treatment of Alzheimer's disease
PS   Claim 3; Page 34-35; 90pp; Japanese.
```



```
CC   The present sequence represents a DNA fragment which is complementary to
CC   mRNA sequences whose level of expression in the brains of Alzheimer's
CC   disease patients varies from normal individuals. The present sequence
CC   represents a specifically claimed DNA fragment in which the expression
CC   is augmented to at least twice that in normal cases. The sequence, and
CC   antibodies to the proteins encoded by the sequence, may be used in the
CC   diagnosis, treatment and investigation of Alzheimer's disease.
SQ   Sequence   1420 BP;    470 A;    234 C;    283 G;    433 T;

  Query Match           11.8%;  Score 14;  DB 35;  Length 1420;
  Best Local Similarity 100.0%;  Pred. No. 3.46e+01;
  Matches    14;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     18 gagagaaagcttat 31
          ||||||||||||||
Cp     36 GAGAGAAAGCTTAT 23


RESULT   50
ID   T39671 standard; cDNA; 1424 BP.
AC   T39671;
DT   03-JUN-1997  (first entry)
DE   D-aminotransferase coding sequence.
KW   D-aminotransferase; Bacillus sphaericus; D-amino acid; alpha-keto acid;
KW   phenylpyruvate; D-phenylalanine; ss.
OS   Bacillus sphaericus.
FH   Key             Location/Qualifiers
FT   cds             427..1278
FT                   /*tag=  a
FT                   /product= D-aminotransferase
PN   EP-736604-A2.
PD   09-OCT-1996.
PF   30-MAR-1996; 105167.
PR   03-APR-1995; US-415716.
PR   19-APR-1995; US-424797.
PA   (NUTR-) NUTRASWEET CO.
PI   Fotheringham I,  Taylor PB,  Yoshida RK;
DR   WPI; 96-444891/45.
DR   P-PSDB; W05302.
PT   Prodn. of D-phenylalanine in E. coli - using recombinant E. coli
PT   contg. a new isolated gene encoding a Bacillus sphaericus
PT   D-amino:transferase
PS   Claim 15; Page 11-12; 24pp; English.
CC   This sequence represents the coding sequence for the D-aminotransferase
CC   of Bacillus sphaericus. D-aminotransferases reversibly catalyse the
CC   transamination of various D-amino acids and corresponding alpha-keto
CC   acids. This sequence was isolated by MboI digestion of B. sphaericus
CC   chromosomal DNA, transforming E. coli with the digested DNA, and
CC   analysing the DNA in transformants plated on appropriate medium
CC   (preferably a medium containing L-aspartic acid, L-alanine, and
CC   phenylpyruvate) that produced D-phenylalanine. The protein encoded by
```



```
CC   this sequence is used in the method of the invention for producing
CC   D-phenylalanine in E. coli.  The method comprises incorporating this gene
CC   into an E. coli, culturing the microorganism in a culture medium and
CC   isolating the D-phenylalanine from the culture.  The method provides for
CC   the high level production of D-phenylalanine, particularly
CC   enantiomerically-pure D-phenylalanine.
SQ   Sequence   1424 BP;    473 A;    242 C;    288 G;    421 T;

  Query Match           11.8%;  Score 14;  DB 29;  Length 1424;
  Best Local Similarity 100.0%;  Pred. No. 3.46e+01;
  Matches    14;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    611 ttattccttataca 624
          ||||||||||||||
Cp     87 TTATTCCTTATACA 74


Search completed: Tue Dec 22 21:38:05 1998
Job time : 149 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Tue Dec 22 21:30:02 1998; MasPar time 203.34 Seconds
1056.813 Million cell updates/sec

Tabular output not generated.

Title: >US-08-530-112A-2058
Description: (1-120) from US08530112A.seq
Perfect Score: 119
N.A. Sequence: 1 GATCTTTACTCTTGAGAAAT.....AAAGCTCCNTAAAAAGCAAA 120
Comp: CTAGAAATGAGAACTCTTTA.....TTTCGAGGNATTTTTCGTTT

Scoring table: TABLE jmetric
Gap 60

Nmatch STD : Dbase 0; Query 0

Searched: 2275026 seqs, 895388244 bases x 2

Post-processing: Minimum Match 0%
Listing first 1000 summaries

Database: embl-est55
1:em_est1 2:em_gss1 3:em_gss2 4:em_gss3
Database: genbank-est107
5:gb_est1 6:gb_est10 7:gb_est11 8:gb_est12 9:gb_est13
10:gb_est14 11:gb_est15 12:gb_est16 13:gb_est17
14:gb_est18 15:gb_est19 16:gb_est2 17:gb_est20
18:gb_est21 19:gb_est3 20:gb_est4 21:gb_est5 22:gb_est6
23:gb_est7 24:gb_est8 25:gb_est9 26:gb_gss1 27:gb_gss2
28:gb_gss3 29:gb_gss4

Statistics: Mean 7.592; Variance 1.157; scale 6.561

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
c	1	119	100.0	404	8	AA479058	zv18g02.s1 Soares NhHM	2.13e-254
	2	116	97.5	117	6	C21276	HUMGS0002342, Human Ge	2.17e-246
	3	18	15.1	247	13	AA832233	oc98b06.s1 NCI_CGAP_GC	1.90e-06
c	4	18	15.1	430	9	AA605560	30596 Lambda-PRL2 Arab	1.90e-06
c	5	18	15.1	455	6	W89200	zh69f06.s1 Soares feta	1.90e-06
	6	18	15.1	473	13	AA805195	oa89h05.s1 NCI_CGAP_GC	1.90e-06
c	7	17	14.3	299	20	ATTS2628	A. thaliana transcribe	8.61e-05
	8	17	14.3	385	12	AA390332	LD03274.5prime LD Dros	8.61e-05
	9	17	14.3	406	28	AQ116088	RPCI11-57B18.TK RPCI11	8.61e-05
	10	17	14.3	422	12	AA793171	vp21g10.r1 Soares mous	8.61e-05
	11	17	14.3	426	24	AA288616	vb23d10.r1 Soares mous	8.61e-05
	12	17	14.3	523	29	AQ252689	HS_3202_B1_E08_T7 CIT	8.61e-05
	13	17	14.3	541	14	AA978637	LD32174.5prime LD Dros	8.61e-05
	14	17	14.3	546	24	AA270045	va57f08.r1 Soares mous	8.61e-05
	15	16	13.4	258	8	AA505495	nh96c11.s1 NCI_CGAP_Br	3.31e-03
c	16	16	13.4	312	20	W36278	HFBEST-55 Human fetal	3.31e-03
	17	16	13.4	328	28	AQ197563	CIT-HSP-2383H11.TR CIT	3.31e-03
c	18	16	13.4	382	10	AA649647	ns46f02.s1 NCI_CGAP_A1	3.31e-03
	19	16	13.4	388	19	N49999	yv24c06.r1 Homo sapien	3.31e-03
	20	16	13.4	390	17	AI015026	ot30e06.s1 Soares_test	3.31e-03
	21	16	13.4	390	17	AU015448	Mus musculus 2-cell em	3.31e-03
c	22	16	13.4	396	7	AA457655	aa91h02.s1 Stratagene	3.31e-03
	23	16	13.4	402	19	H65048	yr69c09.s1 Homo sapien	3.31e-03
c	24	16	13.4	403	16	H09691	y199d07.r1 Homo sapien	3.31e-03
c	25	16	13.4	408	29	AQ152046	HS_3104_A2_H12_MR CIT	3.31e-03
	26	16	13.4	412	13	AA598625	ae39h02.s1 Gessler Wil	3.31e-03
	27	16	13.4	414	19	H91292	yu99h08.s1 Homo sapien	3.31e-03
c	28	16	13.4	418	11	T44484	7747 Lambda-PRL2 Arabi	3.31e-03
c	29	16	13.4	421	28	AQ204013	HS_3106_B2_D03_MR CIT	3.31e-03
c	30	16	13.4	421	26	B44253	HS-1059-A1-D02-MF.abi	3.31e-03
c	31	16	13.4	430	5	R20601	yg49f06.r1 Homo sapien	3.31e-03
c	32	16	13.4	431	27	AQ079945	CIT-HSP-2369K11.TR CIT	3.31e-03
	33	16	13.4	436	26	B45942	HS-1062-B1-G08-MR.abi	3.31e-03
	34	16	13.4	442	12	AA246496	LD05301.5prime LD Dros	3.31e-03
	35	16	13.4	445	12	AA747849	nx31e07.s1 NCI_CGAP_GC	3.31e-03
	36	16	13.4	448	21	AA055645	zl75c08.s1 Stratagene	3.31e-03
	37	16	13.4	455	5	T50662	yb31b02.s1 Homo sapien	3.31e-03
	38	16	13.4	456	19	N33238	yy07e06.s1 Homo sapien	3.31e-03
c	39	16	13.4	456	26	B55083	CIT-HSP-386A19.TF CIT-	3.31e-03
c	40	16	13.4	459	19	N26980	yx18h06.r1 Homo sapien	3.31e-03
c	41	16	13.4	472	5	R18159	yf97e11.r1 Homo sapien	3.31e-03
	42	16	13.4	475	19	N20136	yx29f06.s1 Homo sapien	3.31e-03

	43	16	13.4	480	16	H08237	yl92h07.s1 Homo sapien	3.31e-03
	44	16	13.4	488	22	AA097379	mk16g10.r1 Soares mous	3.31e-03
	45	16	13.4	489	18	AI156664	ue53e05.r1 Soares mous	3.31e-03
	46	16	13.4	489	16	H48298	yq77b02.r1 Homo sapien	3.31e-03
c	47	16	13.4	490	21	W76347	zd63a09.r1 Soares feta	3.31e-03
	48	16	13.4	493	13	AA630644	zc11b02.s1 Stratagene	3.31e-03
	49	16	13.4	507	13	AA868504	ak43b03.s1 Soares test	3.31e-03
c	50	16	13.4	543	29	AQ143721	HS_3075_B1_A03_MF CIT	3.31e-03
	51	16	13.4	545	15	AA978353	oq40b07.s1 NCI_CGAP_Ki	3.31e-03
	52	16	13.4	545	26	B56240	CIT-HSP-2006G12.TR CIT	3.31e-03
c	53	16	13.4	606	26	B15454	345C17.TV CIT978SKA1 H	3.31e-03
c	54	16	13.4	621	29	AQ147005	HS_2250_A2_G01_MR CIT	3.31e-03
	55	16	13.4	637	16	T41554	10135 Arabidopsis thal	3.31e-03
	56	16	13.4	773	26	B21954	F7J2-T7 IGF Arabidopsi	3.31e-03
	57	16	13.4	1040	26	B08378	T30J15-Sp6 TAMU Arabid	3.31e-03
c	58	15	12.6	110	13	AA827849	od01g12.s1 NCI_CGAP_GC	1.06e-01
	59	15	12.6	119	6	AA386412	vc75a01.r1 Knowles Sol	1.06e-01
	60	15	12.6	157	19	H90894	yu89c04.s1 Homo sapien	1.06e-01
	61	15	12.6	164	16	T29320	EST76459 Homo sapiens	1.06e-01
c	62	15	12.6	179	14	AT000251	Malus domestica young	1.06e-01
c	63	15	12.6	180	6	CELK064C5R	C.elegans cDNA clone y	1.06e-01
	64	15	12.6	189	26	B03892	cSRL-19e11-u cSRL flow	1.06e-01
	65	15	12.6	199	26	FR0019672	F.rubripes GSS sequenc	1.06e-01
c	66	15	12.6	207	11	T43552	6815 Lambda-PRL2 Arabi	1.06e-01
	67	15	12.6	225	7	AA499659	vf98g02.r1 Soares mous	1.06e-01
c	68	15	12.6	238	14	AA895782	vy08e05.r1 Stratagene	1.06e-01
c	69	15	12.6	238	8	AA505044	aa63g01.r1 NCI_CGAP_GC	1.06e-01
c	70	15	12.6	246	10	AA665699	ag53g06.s1 Gessler Wil	1.06e-01
	71	15	12.6	265	27	AQ007252	CIT-HSP-2288L14.TF CIT	1.06e-01
c	72	15	12.6	268	11	AA726174	vu87g07.r1 Stratagene	1.06e-01
c	73	15	12.6	269	5	T12472	B040R Homo sapiens cDN	1.06e-01
c	74	15	12.6	279	18	AI169297	EST215132 Normalized r	1.06e-01
c	75	15	12.6	283	17	T46434	9697 Lambda-PRL2 Arabi	1.06e-01
c	76	15	12.6	284	12	AA762550	vv80f12.r1 Stratagene	1.06e-01
	77	15	12.6	286	26	FR0019679	F.rubripes GSS sequenc	1.06e-01
	78	15	12.6	293	5	R27515	yb47e07.r1 Homo sapien	1.06e-01
	79	15	12.6	295	14	AA959655	vw57f09.s1 Soares mous	1.06e-01
c	80	15	12.6	296	14	AA895784	vy08e07.r1 Stratagene	1.06e-01
c	81	15	12.6	299	20	ATTS0120	A. thaliana transcribe	1.06e-01
c	82	15	12.6	299	22	H95441	yw60c11.s1 Soares plac	1.06e-01
c	83	15	12.6	301	11	N38707	19934 Lambda-PRL2 Arab	1.06e-01
c	84	15	12.6	305	15	AA987676	os04d09.s1 NCI_CGAP_Lu	1.06e-01
c	85	15	12.6	317	23	AA215203	mu83c09.r1 Stratagene	1.06e-01
c	86	15	12.6	320	25	AA365137	EST75835 Pineal gland	1.06e-01
	87	15	12.6	323	25	AA372341	EST84490 Colon adenoca	1.06e-01
c	88	15	12.6	328	18	AI177999	EST221655 Normalized r	1.06e-01
	89	15	12.6	331	22	AA022326	SWMFCA431SK Brugia mal	1.06e-01
c	90	15	12.6	332	22	AA154947	mn39b11.r1 Beddington	1.06e-01
c	91	15	12.6	332	27	AQ088734	HS_2235_B1_H05_MF CIT	1.06e-01
c	92	15	12.6	335	18	AI101403	EST210692 Normalized r	1.06e-01
c	93	15	12.6	337	5	T85613	yd56c01.r1 Homo sapien	1.06e-01

c	94	15	12.6	337	20	HSBA4G112	H. sapiens partial cDN	1.06e-01
	95	15	12.6	339	7	AA474439	vd55g05.r1 Knowles Sol	1.06e-01
	96	15	12.6	339	7	AA472875	vb30a10.r1 Soares mous	1.06e-01
	97	15	12.6	342	27	AQ034780	CIT-HSP-2319O22.TR CIT	1.06e-01
c	98	15	12.6	344	14	AA806674	EST190171 Normalised r	1.06e-01
	99	15	12.6	348	7	AA478097	yh17a08.r1 Soares mous	1.06e-01
c	100	15	12.6	350	6	C21422	HUMGS0009794, Human Ge	1.06e-01
	101	15	12.6	351	16	T30152	EST12126 Homo sapiens	1.06e-01
c	102	15	12.6	351	20	ATTS0257	A. thaliana transcribe	1.06e-01
c	103	15	12.6	352	17	AA991606	os60f11.s1 NCI_CGAP_Br	1.06e-01
c	104	15	12.6	354	19	H87193	yw15h05.s1 Homo sapien	1.06e-01
	105	15	12.6	356	5	R37933	yc99a02.s1 Homo sapien	1.06e-01
c	106	15	12.6	358	25	AA372499	EST84673 Colon adenoca	1.06e-01
c	107	15	12.6	360	9	C70715	C.elegans cDNA clone y	1.06e-01
c	108	15	12.6	360	9	C66888	C.elegans cDNA clone y	1.06e-01
c	109	15	12.6	361	9	AA590974	vn61a07.r1 Barstead mo	1.06e-01
c	110	15	12.6	361	8	AA496804	ae33h04.s1 Gessler Wil	1.06e-01
	111	15	12.6	362	14	AA959200	ua08h04.r1 Soares 2NbM	1.06e-01
c	112	15	12.6	367	28	AQ099962	HS_3054_B1_B05_MF CIT	1.06e-01
	113	15	12.6	368	10	AA625662	zu91c09.s1 Soares test	1.06e-01
	114	15	12.6	370	5	R39135	yf61g10.s1 Homo sapien	1.06e-01
c	115	15	12.6	373	21	SSC7D05	S.scrofa mRNA; express	1.06e-01
	116	15	12.6	374	17	AA409938	EST01625 Mouse 7.5 dpc	1.06e-01
c	117	15	12.6	374	5	R22694	yg08d04.r1 Homo sapien	1.06e-01
	118	15	12.6	375	17	AU017219	Mus musculus 2-cell em	1.06e-01
	119	15	12.6	375	9	C69597	C.elegans cDNA clone y	1.06e-01
	120	15	12.6	375	9	C70248	C.elegans cDNA clone y	1.06e-01
	121	15	12.6	376	9	C69447	C.elegans cDNA clone y	1.06e-01
	122	15	12.6	380	5	R24159	yh29d10.r1 Homo sapien	1.06e-01
c	123	15	12.6	381	24	AA325209	EST28159 Cerebellum II	1.06e-01
c	124	15	12.6	383	11	T43891	7154 Lambda-PRL2 Arabi	1.06e-01
	125	15	12.6	384	18	AI145550	UI-R-BT0-qi-a-03-0-UI.	1.06e-01
	126	15	12.6	385	5	T62206	yb97c06.s1 Homo sapien	1.06e-01
	127	15	12.6	387	12	AA765292	oa01c11.s1 NCI_CGAP_GC	1.06e-01
	128	15	12.6	390	26	B47750	RPCI11-2D1.TV RPCI11 H	1.06e-01
	129	15	12.6	391	8	AA084323	zn18c06.s1 Stratagene	1.06e-01
	130	15	12.6	392	5	R40992	yf83g12.s1 Homo sapien	1.06e-01
	131	15	12.6	393	9	AA557520	ng50g03.s1 NCI_CGAP_Co	1.06e-01
c	132	15	12.6	393	26	AQ022027	CIT-HSP-2304B2.TF CIT-	1.06e-01
	133	15	12.6	394	8	AA470458	ne14f04.s1 NCI_CGAP_Co	1.06e-01
	134	15	12.6	397	13	AA496749	ae33h05.r1 Gessler Wil	1.06e-01
	135	15	12.6	397	21	AA048107	mj23g09.r1 Soares mous	1.06e-01
c	136	15	12.6	399	18	AI156103	ue11e07.r1 Soares mous	1.06e-01
c	137	15	12.6	400	28	AQ121089	HS_3075_A2_B05_MR CIT	1.06e-01
c	138	15	12.6	402	27	AQ027608	CIT-HSP-2309G12.TR CIT	1.06e-01
c	139	15	12.6	402	28	AQ094947	HS_3028_A2_C02_MF CIT	1.06e-01
	140	15	12.6	403	26	B76480	RPCI11-15I22.TP RPCI11	1.06e-01
	141	15	12.6	406	19	N45161	yv25c06.r1 Homo sapien	1.06e-01
c	142	15	12.6	407	28	AQ123617	HS_3103_B2_H12_MR CIT	1.06e-01
c	143	15	12.6	407	6	AA057533	zl76b09.s1 Stratagene	1.06e-01
	144	15	12.6	408	28	AQ133879	HS_3041_A2_F10_MR CIT	1.06e-01

c	145	15	12.6	413	8	AA157764	zo46f07.s1 Stratagene	1.06e-01
c	146	15	12.6	414	10	T43180	6443 Lambda-PRL2 Arabi	1.06e-01
c	147	15	12.6	415	16	H12232	yml2g12.s1 Homo sapien	1.06e-01
c	148	15	12.6	416	19	H85677	ys68a07.r1 Homo sapien	1.06e-01
	149	15	12.6	417	13	AA825163	ce62e02.s1 NCI_CGAP_Lu	1.06e-01
c	150	15	12.6	422	28	AQ206767	HS_3241_A2_C09_MR CIT	1.06e-01
c	151	15	12.6	423	28	AQ117933	HS_2207_B1_C06_MF CIT	1.06e-01
c	152	15	12.6	423	5	R18427	yg02g03.r1 Homo sapien	1.06e-01
c	153	15	12.6	423	19	N68125	yz54f09.s1 Homo sapien	1.06e-01
	154	15	12.6	424	16	H03359	yj45c05.s1 Homo sapien	1.06e-01
c	155	15	12.6	424	28	AQ199505	RPCI11-46K15.TK RPCI11	1.06e-01
	156	15	12.6	429	22	N41248	SW3ICA652SK Brugia mal	1.06e-01
	157	15	12.6	429	5	R02477	ye86b04.r1 Homo sapien	1.06e-01
	158	15	12.6	430	19	N63819	za26c10.s1 Homo sapien	1.06e-01
	159	15	12.6	431	5	R25994	yh44a12.r1 Homo sapien	1.06e-01
c	160	15	12.6	433	23	AA237338	mx19h02.r1 Soares mous	1.06e-01
	161	15	12.6	439	28	AQ217732	HS_2007_A1_E01_MR CIT	1.06e-01
c	162	15	12.6	439	12	T76711	11489 Lambda-PRL2 Arab	1.06e-01
c	163	15	12.6	440	27	B79722	CIT-HSP-2045P1.TR CIT-	1.06e-01
c	164	15	12.6	440	12	AA814696	of43g02.s1 NCI_CGAP_CN	1.06e-01
c	165	15	12.6	442	9	AA548938	nk27f10.s1 NCI_CGAP_Co	1.06e-01
c	166	15	12.6	452	16	R80184	yi92c04.s1 Homo sapien	1.06e-01
c	167	15	12.6	452	9	AA551443	nj55d09.s1 NCI_CGAP_Pr	1.06e-01
c	168	15	12.6	453	10	AA527122	ni26g08.s1 NCI_CGAP_Ew	1.06e-01
c	169	15	12.6	453	19	H99206	yz15g02.s1 Homo sapien	1.06e-01
c	170	15	12.6	454	22	AA024505	ze76d06.s1 Soares feta	1.06e-01
c	171	15	12.6	456	12	AA769300	nz36h11.s1 NCI_CGAP_GC	1.06e-01
	172	15	12.6	457	26	B46748	HS-1065-B2-P04-MF.abi	1.06e-01
	173	15	12.6	457	21	AA011243	ze19d03.r1 Soares feta	1.06e-01
	174	15	12.6	458	11	AA701360	zi63f05.s1 Soares feta	1.06e-01
c	175	15	12.6	458	22	AA045604	z166e11.s1 Stratagene	1.06e-01
	176	15	12.6	461	21	AA044238	zk54f02.s1 Soares preg	1.06e-01
c	177	15	12.6	463	14	AA938174	oc10c07.s1 NCI_CGAP_GC	1.06e-01
	178	15	12.6	464	24	AA288175	vb13d10.s1 Soares mous	1.06e-01
	179	15	12.6	464	29	AQ144593	HS_3090_A2_E08_MF CIT	1.06e-01
c	180	15	12.6	464	16	R55385	yj77b07.r1 Homo sapien	1.06e-01
	181	15	12.6	465	16	R74284	yi57e03.r1 Homo sapien	1.06e-01
c	182	15	12.6	467	16	R76285	yi72g06.r1 Homo sapien	1.06e-01
c	183	15	12.6	470	8	AA399369	zt57b10.s1 Soares test	1.06e-01
c	184	15	12.6	470	29	AQ145986	HS_2226_A1_A01_MR CIT	1.06e-01
c	185	15	12.6	471	21	AA010415	ze19d03.s1 Soares feta	1.06e-01
	186	15	12.6	473	13	AA885313	al59g11.s1 Soares NFL	1.06e-01
c	187	15	12.6	474	5	R12002	yf51e05.r1 Homo sapien	1.06e-01
	188	15	12.6	478	14	AA981776	ua36c06.r1 Soares mous	1.06e-01
	189	15	12.6	479	19	N67125	yz46d08.s1 Homo sapien	1.06e-01
c	190	15	12.6	487	26	B49082	RPCI11-4K16.TV RPCI11	1.06e-01
c	191	15	12.6	488	18	AI147289	qb45e09.x1 NCI_CGAP_Br	1.06e-01
c	192	15	12.6	489	28	AQ204086	HS_3107_A2_D06_T7 CIT	1.06e-01
c	193	15	12.6	490	17	AI091389	oo26f03.x1 Soares_NSF_	1.06e-01
	194	15	12.6	491	22	AA022845	ze72e02.s1 Soares feta	1.06e-01
c	195	15	12.6	492	19	N27256	yx17g06.r1 Homo sapien	1.06e-01

c	196	15	12.6	492	13	AA844411	aj19f01.s1 Soares para	1.06e-01
c	197	15	12.6	494	13	AA181954	zp66d09.s1 Stratagene	1.06e-01
c	198	15	12.6	494	15	AA991197	os24f11.s1 NCI_CGAP_Ki	1.06e-01
c	199	15	12.6	494	14	AA959640	vv57e05.s1 Soares mous	1.06e-01
	200	15	12.6	498	5	M89093	CEL16B6 Caenorhabditis	1.06e-01
c	201	15	12.6	499	18	AI157146	ud17a05.r1 Soares 2NbM	1.06e-01
	202	15	12.6	500	18	AI141888	qa66g11.x1 Soares_feta	1.06e-01
	203	15	12.6	504	28	AQ209784	HS_3244_A1_C07_T7 CIT	1.06e-01
c	204	15	12.6	506	16	H23430	ym52e09.r1 Homo sapien	1.06e-01
	205	15	12.6	508	19	N51385	yz16f03.s1 Homo sapien	1.06e-01
c	206	15	12.6	508	14	C89466	Mus musculus early bla	1.06e-01
c	207	15	12.6	508	29	AQ149156	HS_3165_A1_H11_T7 CIT	1.06e-01
c	208	15	12.6	513	7	AA460895	zx69g06.r1 Soares tota	1.06e-01
c	209	15	12.6	515	29	AQ141322	HS_3150_A1_D11_T7 CIT	1.06e-01
c	210	15	12.6	515	27	AQ007493	CIT-HSP-2292P13.TR CIT	1.06e-01
c	211	15	12.6	518	8	AA393095	zt69c04.r1 Soares test	1.06e-01
c	212	15	12.6	519	15	AI044347	UI-R-C0-is-c-07-0-UI.s	1.06e-01
	213	15	12.6	521	8	AA284668	zt24g06.r1 Soares ovar	1.06e-01
c	214	15	12.6	523	22	AA088339	z184f09.r1 Stratagene	1.06e-01
c	215	15	12.6	526	8	AA557463	nk36h09.s1 NCI_CGAP_GC	1.06e-01
c	216	15	12.6	529	17	AI017348	ov09b01.x1 NCI_CGAP_Ki	1.06e-01
	217	15	12.6	532	28	AQ189501	HS_3209_A2_G05_T7 CIT	1.06e-01
c	218	15	12.6	536	28	AQ211354	HS_3238_A2_B09_T7 CIT	1.06e-01
	219	15	12.6	537	14	AA906847	ol20a12.s1 Soares_NFL_	1.06e-01
c	220	15	12.6	537	10	AA648417	ns22b09.s1 NCI_CGAP_GC	1.06e-01
c	221	15	12.6	538	15	AA917666	on46d01.s1 NCI_CGAP_Co	1.06e-01
c	222	15	12.6	544	9	AA551925	nk13d04.s1 NCI_CGAP_Co	1.06e-01
c	223	15	12.6	545	11	AA707837	zh23g12.s1 Soares pine	1.06e-01
	224	15	12.6	549	29	AQ136240	HS_3063_A1_G08_MF CIT	1.06e-01
c	225	15	12.6	549	11	AA726424	vu92a09.r1 Stratagene	1.06e-01
c	226	15	12.6	549	27	AQ080778	CIT-HSP-23B2H19.TF CIT	1.06e-01
	227	15	12.6	550	13	AA884557	am38a05.s1 Soares NFL	1.06e-01
c	228	15	12.6	552	9	AA548743	nk40d06.s1 NCI_CGAP_GC	1.06e-01
c	229	15	12.6	553	10	AA650729	30843 Lambda-PRL2 Arab	1.06e-01
c	230	15	12.6	553	11	H37466	15595 Lambda-PRL2 Arab	1.06e-01
c	231	15	12.6	555	8	AA281008	zt01c04.s1 NCI_CGAP_GC	1.06e-01
	232	15	12.6	556	12	AA760198	vv74d05.r1 Stratagene	1.06e-01
	233	15	12.6	556	23	AA241868	mx21f12.r1 Soares mous	1.06e-01
	234	15	12.6	561	24	AA276754	vc45f08.r1 Soares mous	1.06e-01
	235	15	12.6	561	28	AQ158000	nbxb0010B05f CUGI Rice	1.06e-01
c	236	15	12.6	565	12	AA775203	ac79c11.s1 Stratagene	1.06e-01
	237	15	12.6	566	27	B83526	RPCI11-15I22.TPB RPCI1	1.06e-01
c	238	15	12.6	568	6	AA056544	zk80d08.s1 Soares preg	1.06e-01
	239	15	12.6	569	8	AA522880	ni41b01.s1 NCI_CGAP_Lu	1.06e-01
c	240	15	12.6	569	27	AQ044794	CIT-HSP-2319M11.TF CIT	1.06e-01
	241	15	12.6	570	7	AA451568	AOB213F Onion seedling	1.06e-01
c	242	15	12.6	570	28	AQ122448	HS_3080_A2_F02_MF CIT	1.06e-01
	243	15	12.6	579	21	W52071	zc47e10.r1 Soares sene	1.06e-01
c	244	15	12.6	582	10	AA618054	nq02g03.s1 NCI_CGAP_Lu	1.06e-01
c	245	15	12.6	589	19	N39314	yv25c06.s1 Homo sapien	1.06e-01
	246	15	12.6	591	26	FR0012384	F.rubripes GSS sequenc	1.06e-01

c	247	15	12.6	591	17	AU018608	Mus musculus 8-cell em	1.06e-01
c	248	15	12.6	598	20	W21822	57d4 Human retina cDNA	1.06e-01
c	249	15	12.6	604	6	C06049	similar to cleavage si	1.06e-01
c	250	15	12.6	610	20	AA001419	ze45e05.s1 Soares reti	1.06e-01
	251	15	12.6	616	20	W22610	70P4 Human retina cDNA	1.06e-01
c	252	15	12.6	622	27	AQ082036	RPCI11-54I9.TK RPCI11	1.06e-01
c	253	15	12.6	624	14	AA889762	al50e02.s1 Soares_NFL_	1.06e-01
	254	15	12.6	627	18	AI158980	ui42h12.x1 Sugano mous	1.06e-01
	255	15	12.6	628	12	AA764623	vp10h04.r1 Soares mous	1.06e-01
	256	15	12.6	633	27	AQ061141	CIT-HSP-2348E21.TF CIT	1.06e-01
	257	15	12.6	642	15	AI055635	coau0004J22 Cotton Bol	1.06e-01
	258	15	12.6	648	15	AI055668	coau0004L08 Cotton Bol	1.06e-01
	259	15	12.6	680	29	AQ249338	T22K9-Sp6.1 TAMU Arabi	1.06e-01
c	260	15	12.6	686	27	AQ047752	cLM-3a11-t cLM Giardia	1.06e-01
c	261	15	12.6	694	11	AA669865	ag42b03.s1 Jia bone ma	1.06e-01
c	262	15	12.6	701	17	AI088834	qb17d01.x1 Soares_preg	1.06e-01
c	263	15	12.6	705	26	B28385	F18E12TF IGF Arabidops	1.06e-01
	264	15	12.6	707	27	AQ049045	cLM-11g2-t cLM Giardia	1.06e-01
c	265	15	12.6	725	15	AI055048	coau0002007 Cotton Bol	1.06e-01
	266	15	12.6	761	28	AQ114213	CIT-HSP-2382015.TR CIT	1.06e-01
c	267	15	12.6	761	28	AF010693	Homo sapiens chromosom	1.06e-01
c	268	15	12.6	782	13	AA837102	od19e07.s1 NCI_CGAP_GC	1.06e-01
c	269	15	12.6	796	12	AA742447	nx27f07.s1 NCI_CGAP_GC	1.06e-01
c	270	15	12.6	1038	26	BI37715	ji452 Trypanosome Shoc	1.06e-01
c	271	15	12.6	1141	26	B10270	F16N3-Sp6 IGF Arabidop	1.06e-01
	272	14	11.8	59	6	C01504	HUMGS0008499, Human Ge	2.76e+00
	273	14	11.8	88	16	R79591	yi88c12.s1 Homo sapien	2.76e+00
c	274	14	11.8	88	8	AA547246	vk70d01.s1 Knowles Sol	2.76e+00
	275	14	11.8	89	5	T71779	yc64c01.s1 Homo sapien	2.76e+00
	276	14	11.8	103	18	AI142352	qg63h03.r1 Soares_test	2.76e+00
	277	14	11.8	110	6	AA011581	zi02a06.r1 Soares feta	2.76e+00
	278	14	11.8	125	23	AA139472	mr82f09.r1 Stratagene	2.76e+00
c	279	14	11.8	125	14	AA906481	oj99f12.s1 Soares_NFL_	2.76e+00
c	280	14	11.8	128	6	C21272	HUMGS0002333, Human Ge	2.76e+00
c	281	14	11.8	136	26	B97629	F16A10TPC IGF Arabidop	2.76e+00
c	282	14	11.8	156	24	AA270416	va65e03.r1 Soares mous	2.76e+00
	283	14	11.8	158	20	ATTS1137	A. thaliana transcribe	2.76e+00
	284	14	11.8	162	22	AA193802	rs03g11.r1 Sommer Pris	2.76e+00
	285	14	11.8	162	27	AQ008784	RPCI11-22B23.TPC RPCI1	2.76e+00
c	286	14	11.8	188	27	AQ045942	RPCI11-39C24.TP RPCI11	2.76e+00
c	287	14	11.8	190	14	AA851792	EST194560 Normalized r	2.76e+00
c	288	14	11.8	192	29	AQ236550	HS_2031_B2_E09_T7 CIT	2.76e+00
c	289	14	11.8	193	25	AA347715	EST54057 Fetal heart I	2.76e+00
c	290	14	11.8	194	12	AA751009	ISNJ0146 Rice Immature	2.76e+00
c	291	14	11.8	196	21	W17868	mb83h05.r1 Soares mous	2.76e+00
	292	14	11.8	204	26	B67576	T23F13TF TAMU Arabidop	2.76e+00
	293	14	11.8	207	19	N63110	yz36a12.s1 Homo sapien	2.76e+00
c	294	14	11.8	207	8	AA494408	ne28a06.s1 NCI_CGAP_Co	2.76e+00
c	295	14	11.8	216	10	T04107	57 Lambda-PRL1 Arabido	2.76e+00
	296	14	11.8	217	25	AA372726	EST84651 Colon adenoca	2.76e+00
c	297	14	11.8	218	20	ATTS0517	A. thaliana transcribe	2.76e+00

c	298	14	11.8	218	28	AQ106932	HS_3077_B2_H06_MR CIT	2.76e+00
c	299	14	11.8	222	11	AA719927	zg67h10.s1 Soares feta	2.76e+00
	300	14	11.8	223	25	AA337326	EST42121 Endometrial t	2.76e+00
c	301	14	11.8	225	21	W25225	zb68a09.r1 Soares feta	2.76e+00
c	302	14	11.8	226	14	AA915458	vz31b02.s1 Soares 2NbM	2.76e+00
c	303	14	11.8	226	20	N84445	KK8816F Homo sapiens c	2.76e+00
	304	14	11.8	228	20	ATTS0413	A. thaliana transcribe	2.76e+00
c	305	14	11.8	229	25	AA334625	EST38869 Embryo, 9 wee	2.76e+00
c	306	14	11.8	229	20	DY1B26H10	H. sapiens partial cDN	2.76e+00
	307	14	11.8	231	10	T22613	4621 Lambda-PRL2 Arabi	2.76e+00
c	308	14	11.8	233	5	T94419	ye36a05.s1 Homo sapien	2.76e+00
	309	14	11.8	234	14	AA964978	UI-R-C0-hb-d-07-0-UI.s	2.76e+00
	310	14	11.8	235	7	HUM340G10B	Human aorta cDNA 5'-en	2.76e+00
c	311	14	11.8	237	9	AA524118	ng35b12.s1 NCI_CGAP_Co	2.76e+00
	312	14	11.8	239	23	AA207738	mv78g07.r1 GuayWoodfor	2.76e+00
	313	14	11.8	241	7	HUM285G12B	Human aorta cDNA 5'-en	2.76e+00
	314	14	11.8	244	27	AQ061221	CIT-HSP-2350K17.TF CIT	2.76e+00
c	315	14	11.8	244	22	AA164054	mt67f07.r1 Soares mous	2.76e+00
	316	14	11.8	246	6	C04278	Human Heart cDNA, clon	2.76e+00
	317	14	11.8	248	7	HUM122H05B	Human fetal brain cDNA	2.76e+00
	318	14	11.8	253	28	AQ197605	CIT-HSP-2384A18.TR CIT	2.76e+00
	319	14	11.8	257	24	AA296033	EST10298 Adipose tissu	2.76e+00
	320	14	11.8	258	5	R43366	yg15g07.s1 Homo sapien	2.76e+00
	321	14	11.8	264	9	AA991259	vm22a07.r1 Knowles Sol	2.76e+00
c	322	14	11.8	269	8	AA411773	zu02e01.s1 Soares test	2.76e+00
c	323	14	11.8	269	8	AA411765	zu02c01.s1 Soares test	2.76e+00
c	324	14	11.8	270	19	H58499	yr21f07.s1 Homo sapien	2.76e+00
	325	14	11.8	271	13	AA809607	nj43h11.s1 NCI_CGAP_Pr	2.76e+00
	326	14	11.8	276	9	AA573682	nf42h10.s1 NCI_CGAP_Pr	2.76e+00
	327	14	11.8	276	28	AQ099604	HS_3056_A2_A08_MR CIT	2.76e+00
c	328	14	11.8	277	13	AA748883	ny73g11.s1 NCI_CGAP_GC	2.76e+00
c	329	14	11.8	278	15	AA933690	om65a04.s1 NCI_CGAP_GC	2.76e+00
	330	14	11.8	280	27	AQ001589	CIT-HSP-2284M16.TF CIT	2.76e+00
	331	14	11.8	281	16	N78635	EST00783 Homo sapiens	2.76e+00
	332	14	11.8	281	7	HUML11771	Human fetal-lung cDNA	2.76e+00
c	333	14	11.8	284	25	AA380406	EST93401 Supt cells Ho	2.76e+00
c	334	14	11.8	285	6	AA383312	EST96671 Testis I Homo	2.76e+00
	335	14	11.8	287	13	AA206264	zq54f12.r1 Stratagene	2.76e+00
c	336	14	11.8	287	20	HSC39C112	H. sapiens partial cDN	2.76e+00
c	337	14	11.8	288	5	R00592	ye74b02.r1 Homo sapien	2.76e+00
	338	14	11.8	289	24	AA309102	EST179878 Colon carcin	2.76e+00
c	339	14	11.8	289	8	AA195520	zr33e07.s1 Soares NhHM	2.76e+00
c	340	14	11.8	294	18	AI147345	qg63h03.s1 Soares_test	2.76e+00
	341	14	11.8	295	27	AQ072533	HS_2232_B1_G05_MF CIT	2.76e+00
c	342	14	11.8	295	16	H04683	yj10b11.r1 Homo sapien	2.76e+00
	343	14	11.8	296	6	C04545	Human Heart cDNA, clon	2.76e+00
c	344	14	11.8	296	10	AA417694	zv01a05.s1 NCI_CGAP_GC	2.76e+00
c	345	14	11.8	300	20	ATTS5338	A. thaliana transcribe	2.76e+00
c	346	14	11.8	301	5	T07559	EST05449 Homo sapiens	2.76e+00
c	347	14	11.8	301	16	T29337	EST77003 Homo sapiens	2.76e+00
c	348	14	11.8	302	20	ATTS1034	A. thaliana transcribe	2.76e+00

	349	14	11.8	303	17	AA784668	f5e02a1.f1 Aspergillus	2.76e+00
c	350	14	11.8	303	27	AQ005284	CIT-HSP-2291A15.TR CIT	2.76e+00
c	351	14	11.8	304	15	AA973054	op31e08.s1 Soares_NFL_	2.76e+00
	352	14	11.8	305	15	AI026558	TENU0770 T. cruzi epim	2.76e+00
	353	14	11.8	305	15	RICC10369A	Rice cDNA, partial seq	2.76e+00
c	354	14	11.8	306	20	ATTS0285	A. thaliana transcribe	2.76e+00
c	355	14	11.8	307	9	AA228307	nc38f01.r1 NCI_CGAP_Pr	2.76e+00
	356	14	11.8	308	16	H19514	yn59f07.s1 Homo sapien	2.76e+00
c	357	14	11.8	308	5	T77521	yd73e05.s1 Homo sapien	2.76e+00
c	358	14	11.8	309	9	AA553691	nl04g07.s1 NCI_CGAP_Co	2.76e+00
	359	14	11.8	310	5	R10383	yf37c07.s1 Homo sapien	2.76e+00
	360	14	11.8	311	6	HSPD06155	H.sapiens EST sequence	2.76e+00
	361	14	11.8	312	23	AA068443	mm46a03.r1 Stratagene	2.76e+00
	362	14	11.8	312	8	AA515577	nf66a04.s1 NCI_CGAP_Co	2.76e+00
c	363	14	11.8	313	28	AQ092257	HS_3016_B2_H02_MR CIT	2.76e+00
c	364	14	11.8	314	28	AQ221448	HS_2004_B2_E08_T7 CIT	2.76e+00
	365	14	11.8	315	15	AA953134	oo37h03.s1 NCI_CGAP_Lu	2.76e+00
c	366	14	11.8	315	27	AQ066587	HS_2251_B1_H05_MF CIT	2.76e+00
	367	14	11.8	316	26	B48749	RPCI11-3E18.TV RPCI11	2.76e+00
	368	14	11.8	316	14	AA857372	od83g06.s1 NCI_CGAP_Ov	2.76e+00
c	369	14	11.8	316	15	AA999436	MCCE3293.M1F Schistoso	2.76e+00
c	370	14	11.8	316	5	T36395	EST101325 Saccharomyce	2.76e+00
	371	14	11.8	317	28	AQ106964	HS_3086_A2_C06_MR CIT	2.76e+00
c	372	14	11.8	317	8	AA491684	ne80g10.s1 NCI_CGAP_Ew	2.76e+00
	373	14	11.8	319	8	AA477826	xu39g07.s1 Soares ovar	2.76e+00
	374	14	11.8	319	13	AA845641	ai90d04.s1 Soares para	2.76e+00
	375	14	11.8	320	15	AI034523	LmLw39p10/160C Leishma	2.76e+00
	376	14	11.8	320	29	AQ229609	HS_2028_A2_A08_MR CIT	2.76e+00
c	377	14	11.8	320	28	AQ097802	HS_3038_A2_A02_MF CIT	2.76e+00
c	378	14	11.8	320	26	AQ021294	CIT-HSP-2301M2.TR CIT-	2.76e+00
c	379	14	11.8	321	20	R58312	G1867 Fetal heart Homo	2.76e+00
c	380	14	11.8	321	10	AA651631	ns89a03.s1 NCI_CGAP_Pr	2.76e+00
	381	14	11.8	324	29	AQ153323	HS_2233_A1_C01_MR CIT	2.76e+00
	382	14	11.8	324	7	AA445924	zw57f08.r1 Soares tota	2.76e+00
c	383	14	11.8	324	26	FR0018607	F.rubripes GSS sequenc	2.76e+00
	384	14	11.8	325	14	H34799	EST112044 Rat PC-12 ce	2.76e+00
	385	14	11.8	325	25	AA347006	EST53431 Fetal heart I	2.76e+00
	386	14	11.8	325	15	AI026475	TENU0687 T. cruzi epim	2.76e+00
	387	14	11.8	326	16	H06001	yl76c03.r1 Homo sapien	2.76e+00
c	388	14	11.8	326	8	AA261953	zs23e05.s1 NCI_CGAP_GC	2.76e+00
c	389	14	11.8	326	11	AA081449	zn06g06.r1 Stratagene	2.76e+00
c	390	14	11.8	327	16	R44656	yg29f09.s1 Homo sapien	2.76e+00
	391	14	11.8	328	16	H48928	yq80h11.r1 Homo sapien	2.76e+00
c	392	14	11.8	328	5	R36747	yg15a04.r9 Homo sapien	2.76e+00
c	393	14	11.8	328	21	AA058016	mj56f10.r1 Soares mous	2.76e+00
	394	14	11.8	329	13	AA824742	CT050.SK Tomato Leaf c	2.76e+00
	395	14	11.8	330	15	AI026413	TENU0624 T. cruzi epim	2.76e+00
c	396	14	11.8	330	17	AA407660	EST01741 Mouse 7.5 dpc	2.76e+00
c	397	14	11.8	331	5	R06502	yf09e08.r1 Homo sapien	2.76e+00
	398	14	11.8	332	28	AQ195780	RPCI11-58A11.TK RPCI11	2.76e+00
	399	14	11.8	333	11	T14180	2345 Lambda-PRL2 Arabi	2.76e+00

c	400	14	11.8	334	19	N33628	SMNHADA002012SK Schist	2.76e+00
	401	14	11.8	335	8	AA085789	zl90a11.r1 Stratagene	2.76e+00
	402	14	11.8	335	24	AA309793	EST180650 Jurkat T-cel	2.76e+00
c	403	14	11.8	335	12	AA781208	aj24h02.s1 Soares test	2.76e+00
c	404	14	11.8	336	20	DY1BA7D07	H. sapiens partial cDN	2.76e+00
c	405	14	11.8	336	14	C90610	Dictyostelium discoide	2.76e+00
c	406	14	11.8	337	25	AA376752	EST89272 Small intesti	2.76e+00
c	407	14	11.8	337	16	H54548	yq91h02.r1 Homo sapien	2.76e+00
c	408	14	11.8	339	5	T11825	A1038R Homo sapiens cD	2.76e+00
	409	14	11.8	341	27	B83781	RPCI11-20A8.TV RPCI11	2.76e+00
	410	14	11.8	343	11	AA731394	nz86h12.s1 NCI_CGAP_GC	2.76e+00
	411	14	11.8	343	10	AA663956	ac07a10.s1 Stratagene	2.76e+00
c	412	14	11.8	343	21	AA007863	mg76g04.r1 Soares mous	2.76e+00
c	413	14	11.8	343	9	AA568820	nm22a09.s1 NCI_CGAP_Co	2.76e+00
	414	14	11.8	345	9	AA230171	nc38f01.s1 NCI_CGAP_Pr	2.76e+00
c	415	14	11.8	345	5	T82360	yc14c08.r1 Homo sapien	2.76e+00
c	416	14	11.8	345	27	AQ078055	CIT-HSP-2354L11.TF CIT	2.76e+00
	417	14	11.8	347	18	AI170496	EST216424 Normalized r	2.76e+00
c	418	14	11.8	347	21	T14757	5c01d09-t7 membrane-fr	2.76e+00
c	419	14	11.8	349	9	C44375	C.elegans cDNA clone y	2.76e+00
	420	14	11.8	350	18	AI170976	EST216920 Normalized r	2.76e+00
	421	14	11.8	350	23	T03000	FB1D12 Fetal brain, St	2.76e+00
	422	14	11.8	351	25	AA340556	EST45777 Fetal kidney	2.76e+00
c	423	14	11.8	352	17	AI015828	ov51g10.s1 Soares_test	2.76e+00
	424	14	11.8	353	6	AA010240	zi08e06.r1 Soares feta	2.76e+00
c	425	14	11.8	353	6	AA001307	zh82e09.r1 Soares feta	2.76e+00
c	426	14	11.8	353	9	N60505	TgESTsy13e11.r1 TgRH T	2.76e+00
	427	14	11.8	354	25	AA342616	EST48170 Fetal spleen	2.76e+00
c	428	14	11.8	354	12	AA767498	oa90f10.s1 NCI_CGAP_GC	2.76e+00
	429	14	11.8	355	8	AA397867	zt78g08.r1 Soares test	2.76e+00
c	430	14	11.8	357	27	AQ041956	CIT-HSP-2336D8.TR CIT-	2.76e+00
	431	14	11.8	359	18	AI076278	oy87d03.x1 Soares_feta	2.76e+00
	432	14	11.8	360	9	C68428	C.elegans cDNA clone y	2.76e+00
c	433	14	11.8	360	9	C60500	C.elegans cDNA clone y	2.76e+00
c	434	14	11.8	360	10	T22174	4102 Lambda-PRL2 Arabi	2.76e+00
c	435	14	11.8	361	14	AA982384	ua53h12.r1 Soares 2NbM	2.76e+00
c	436	14	11.8	362	11	AA083325	zn09b02.s1 Stratagene	2.76e+00
	437	14	11.8	363	11	T14221	2386 Lambda-PRL2 Arabi	2.76e+00
c	438	14	11.8	364	14	AA958680	vz94g07.r1 Soares 2NbM	2.76e+00
c	439	14	11.8	364	21	W79325	zd80e03.r1 Soares feta	2.76e+00
	440	14	11.8	365	17	AA937140	oo98g02.s1 NCI_CGAP_Ki	2.76e+00
c	441	14	11.8	365	5	R19468	yg21a04.r1 Homo sapien	2.76e+00
c	442	14	11.8	366	19	N47386	yy88a09.s1 Homo sapien	2.76e+00
c	443	14	11.8	366	5	T09415	EST07308 Homo sapiens	2.76e+00
c	444	14	11.8	367	19	H87691	yw15f07.r1 Homo sapien	2.76e+00
	445	14	11.8	369	9	C69183	C.elegans cDNA clone y	2.76e+00
	446	14	11.8	369	22	AA120916	zn83e11.s1 Stratagene	2.76e+00
	447	14	11.8	369	18	AI150301	qf40b08.x1 Soares_test	2.76e+00
c	448	14	11.8	369	20	HSC3HB051	H. sapiens partial cDN	2.76e+00
	449	14	11.8	371	26	B43946	HS-1058-B1-C11-MF.abi	2.76e+00
c	450	14	11.8	372	5	T58682	ya94d08.r1 Homo sapien	2.76e+00

	451	14	11.8	373	29	AQ151409	HS_2211_A2_H08_MF CIT	2.76e+00
	452	14	11.8	373	28	AQ195421	RPCI11-46P1.TK RPCI11	2.76e+00
	453	14	11.8	373	29	AQ139657	HS_3093_B1_H06_MR CIT	2.76e+00
	454	14	11.8	374	10	AA421091	zu05a02.r1 Soares test	2.76e+00
c	455	14	11.8	374	28	AQ093984	HS_2199_A2_D08_MR CIT	2.76e+00
c	456	14	11.8	374	11	T03971	174 AT-NHC Arabidopsis	2.76e+00
	457	14	11.8	375	8	AA491202	aa52h08.r1 NCI_CGAP_GC	2.76e+00
c	458	14	11.8	375	5	R00242	ye69f01.s1 Homo sapien	2.76e+00
c	459	14	11.8	375	5	T59459	yb63c01.r1 Homo sapien	2.76e+00
c	460	14	11.8	376	28	AQ113998	CIT-HSP-2360L22.TR CIT	2.76e+00
	461	14	11.8	377	29	AQ149038	HS_3145_B2_E07_MR CIT	2.76e+00
c	462	14	11.8	377	15	AA909676	om47c02.s1 Soares_NFL_	2.76e+00
c	463	14	11.8	377	15	AI060056	UI-R-C1-kx-f-08-0-UI.s	2.76e+00
	464	14	11.8	378	15	AA910247	ok86b10.s1 NCI_CGAP_Ki	2.76e+00
c	465	14	11.8	378	5	R18555	yf95a09.r1 Homo sapien	2.76e+00
c	466	14	11.8	378	27	AQ036494	CIT-HSP-2321K1.TF CIT-	2.76e+00
	467	14	11.8	380	15	AI057783	TENU1874 T. cruzi epim	2.76e+00
	468	14	11.8	381	16	H20404	yn56b12.r1 Homo sapien	2.76e+00
	469	14	11.8	381	27	AQ054598	CIT-HSP-2344P7.TR CIT-	2.76e+00
c	470	14	11.8	381	26	B63435	RPCI11-5B4.TVB RPCI11	2.76e+00
c	471	14	11.8	382	28	AQ098539	HS_3050_A1_B10_MF CIT	2.76e+00
c	472	14	11.8	383	15	AA954990	op01a08.s1 NCI_CGAP_Ki	2.76e+00
c	473	14	11.8	383	27	AQ045996	RPCI11-31M19.TK RPCI11	2.76e+00
	474	14	11.8	384	11	N65203	20243 Lambda-PRL2 Arab	2.76e+00
	475	14	11.8	384	9	AA533512	nj96b11.s1 NCI_CGAP_Pr	2.76e+00
	476	14	11.8	384	21	W35282	zc07d02.s1 Soares para	2.76e+00
	477	14	11.8	385	28	AQ166668	HS_3148_A1_D01_MR CIT	2.76e+00
c	478	14	11.8	385	23	AA106609	mm17g04.r1 Stratagene	2.76e+00
	479	14	11.8	386	22	AA101211	zn37d01.s1 Stratagene	2.76e+00
c	480	14	11.8	386	13	C87840	Mus musculus fertilize	2.76e+00
c	481	14	11.8	387	8	AA502555	ng62e09.s1 NCI_CGAP_Li	2.76e+00
	482	14	11.8	388	19	H77877	ys09d02.r1 Homo sapien	2.76e+00
c	483	14	11.8	388	8	AA491523	ne78c11.s1 NCI_CGAP_Ew	2.76e+00
c	484	14	11.8	389	13	AA837463	od20e01.s1 NCI_CGAP_GC	2.76e+00
c	485	14	11.8	389	27	AQ086069	HS_2165_B2_E04_MR CIT	2.76e+00
c	486	14	11.8	389	26	B46845	HS-1066-A1-F04-MR.abi	2.76e+00
	487	14	11.8	390	27	AQ055847	CIT-HSP-2341H11.TR CIT	2.76e+00
c	488	14	11.8	390	5	R09340	yf26d06.r1 Homo sapien	2.76e+00
	489	14	11.8	391	26	B88396	kflak011 mouse BAC lib	2.76e+00
c	490	14	11.8	391	9	AA565566	nk42f09.s1 NCI_CGAP_GC	2.76e+00
c	491	14	11.8	391	19	N44678	yy22b03.r1 Homo sapien	2.76e+00
c	492	14	11.8	392	11	H37518	15647 Lambda-PRL2 Arab	2.76e+00
	493	14	11.8	393	15	AA931637	co15c12.s1 NCI_CGAP_Lu	2.76e+00
	494	14	11.8	394	28	AQ164750	HS_3011_B1_A06_MR CIT	2.76e+00
	495	14	11.8	394	5	T05945	EST03834 Homo sapiens	2.76e+00
c	496	14	11.8	394	28	AQ219461	HS_3245_B2_C05_MR CIT	2.76e+00
c	497	14	11.8	394	15	AA954411	on85a08.s1 Soares_NFL_	2.76e+00
c	498	14	11.8	394	8	AA502447	ng62c09.s1 NCI_CGAP_Li	2.76e+00
c	499	14	11.8	394	28	AQ130619	HS_3008_A2_B06_MR CIT	2.76e+00
	500	14	11.8	395	29	AQ239620	CIT-HSP-2387H8.TR.1 CI	2.76e+00
c	501	14	11.8	395	28	AQ121077	HS_3075_A2_A04_MR CIT	2.76e+00

c	502	14	11.8	395	6	AA026184	zj99e11.s1 Soares preg	2.76e+00
	503	14	11.8	396	14	AA887508	oj37e08.s1 NCI_CGAP_Lu	2.76e+00
	504	14	11.8	396	28	AQ205760	HS_3245_A1_B05_MR CIT	2.76e+00
	505	14	11.8	396	14	AA969889	op59b06.s1 Soares_NFL_	2.76e+00
	506	14	11.8	397	27	AQ061814	CIT-HSP-2357G10.TR CIT	2.76e+00
	507	14	11.8	397	27	B86317	RPCI11-22L16.TV RPCI11	2.76e+00
	508	14	11.8	397	21	AA016495	mg88a03.r1 Soares mous	2.76e+00
	509	14	11.8	397	28	AQ183089	HS_3126_B2_H07_MR CIT	2.76e+00
c	510	14	11.8	397	13	AA861733	ak38b09.s1 Soares test	2.76e+00
c	511	14	11.8	397	21	AA032277	zf01e04.r1 Soares feta	2.76e+00
	512	14	11.8	398	26	AQ009754	HS_2263_A1_D07_MF CIT	2.76e+00
c	513	14	11.8	399	27	AQ071997	HS_3020_A2_H01_MR CIT	2.76e+00
c	514	14	11.8	399	5	T58703	ya94g11.r1 Homo sapien	2.76e+00
	515	14	11.8	401	10	AA421090	zu05a01.r1 Soares test	2.76e+00
c	516	14	11.8	402	29	AQ225705	HS_2009_B2_C09_T7 CIT	2.76e+00
	517	14	11.8	403	22	AA188521	zp74c06.r1 Stratagene	2.76e+00
c	518	14	11.8	403	9	AA012058	TgESTzz18g11.r1 TgME49	2.76e+00
	519	14	11.8	404	26	B75039	RPCI11-12L7.TP RPCI11	2.76e+00
	520	14	11.8	405	28	AQ119502	HS_2269_A2_D10_MF CIT	2.76e+00
	521	14	11.8	406	11	T45362	8625 Lambda-PRL2 Arabi	2.76e+00
	522	14	11.8	406	28	AQ222655	HS_2012_A2_A10_T7 CIT	2.76e+00
	523	14	11.8	406	18	AI075426	oz82e04.x1 Soares_sene	2.76e+00
c	524	14	11.8	406	23	W93408	zh46h08.s1 Soares feta	2.76e+00
	525	14	11.8	408	27	AQ058829	RPCI11-52I4.TJ RPCI11	2.76e+00
c	526	14	11.8	408	28	AQ220024	HS_3251_B2_D09_MR CIT	2.76e+00
c	527	14	11.8	408	19	N39452	yy49h02.s1 Homo sapien	2.76e+00
c	528	14	11.8	409	5	R10923	yf38a04.r1 Homo sapien	2.76e+00
	529	14	11.8	410	19	H83249	yq49c09.r1 Homo sapien	2.76e+00
c	530	14	11.8	410	16	H29313	ym60d01.s1 Homo sapien	2.76e+00
c	531	14	11.8	410	23	AA119321	mp59f06.r1 Soares 2NbM	2.76e+00
	532	14	11.8	412	23	W84857	zh53b09.s1 Soares feta	2.76e+00
	533	14	11.8	412	29	AQ253020	HS_2046_A1_C03_T7 CIT	2.76e+00
	534	14	11.8	412	10	AA651747	ns67d01.r1 NCI_CGAP_Pr	2.76e+00
c	535	14	11.8	412	15	R75828	yi59d10.r1 Homo sapien	2.76e+00
c	536	14	11.8	412	24	AA318934	EST21077 Adrenal gland	2.76e+00
	537	14	11.8	413	19	N30703	yw68d05.s1 Homo sapien	2.76e+00
c	538	14	11.8	413	27	AQ065101	HS_2263_A1_E01_MF CIT	2.76e+00
c	539	14	11.8	413	17	AI115158	ui47c12.x1 Sugano mous	2.76e+00
c	540	14	11.8	413	26	AQ023113	HS_2182_B1_C06_MF CIT	2.76e+00
c	541	14	11.8	414	27	AQ032580	HS_2197_B2_H11_MF CIT	2.76e+00
	542	14	11.8	415	11	AA696304	GM07410.5prime GM Dros	2.76e+00
	543	14	11.8	415	22	AA127253	zn89b09.r1 Stratagene	2.76e+00
c	544	14	11.8	415	14	AA982208	vw61e08.s1 Soares mous	2.76e+00
c	545	14	11.8	415	26	B49563	RPCI11-1A11.TVB RPCI11	2.76e+00
	546	14	11.8	416	28	AQ187015	HS_3122_B2_F12_MR CIT	2.76e+00
c	547	14	11.8	416	21	W51987	xc49e04.s1 Soares sene	2.76e+00
c	548	14	11.8	416	28	AQ092441	HS_3004_A2_C05_T7 CIT	2.76e+00
	549	14	11.8	417	18	AI091125	qa46d05.s1 Soares_NhHM	2.76e+00
c	550	14	11.8	417	19	H56413	yt87h02.r1 Homo sapien	2.76e+00
	551	14	11.8	418	28	AQ214074	HS_2246_B2_H07_MR CIT	2.76e+00
c	552	14	11.8	418	16	H54523	yq91d02.r1 Homo sapien	2.76e+00

	553	14	11.8	419	11	N37614	18841 Lambda-PRL2 Arab	2.76e+00
	554	14	11.8	419	29	AQ145965	HS_2224_B2_F07_MF CIT	2.76e+00
c	555	14	11.8	419	17	AI016135	ot35g05.s1 Soares_test	2.76e+00
	556	14	11.8	420	29	AQ143512	HS_3067_B1_B05_MR CIT	2.76e+00
c	557	14	11.8	420	27	AQ032271	HS_2194_B2_H05_MF CIT	2.76e+00
c	558	14	11.8	420	13	AA158453	zo59d12.s1 Stratagene	2.76e+00
c	559	14	11.8	420	5	T99862	ye68b02.r1 Homo sapien	2.76e+00
	560	14	11.8	421	8	AA158116	zo43g04.s1 Stratagene	2.76e+00
	561	14	11.8	422	26	FR0003452	F.rubripes GSS sequenc	2.76e+00
	562	14	11.8	422	8	AA133933	zo16f11.r1 Stratagene	2.76e+00
c	563	14	11.8	423	28	AQ093820	HS_3023_B2_E10_T7 CIT	2.76e+00
c	564	14	11.8	423	18	AI088689	qa12a01.x1 NCI_CGAP_Br	2.76e+00
	565	14	11.8	424	16	R78247	yi81a09.r1 Homo sapien	2.76e+00
	566	14	11.8	424	18	AI052220	oz14d07.x1 Soares_feta	2.76e+00
c	567	14	11.8	424	13	AA178896	zp38d07.r1 Stratagene	2.76e+00
	568	14	11.8	425	15	AI026523	TENU0735 T. cruzi epim	2.76e+00
	569	14	11.8	425	27	AQ002148	CIT-HSP-2284A4.TF CIT-	2.76e+00
c	570	14	11.8	425	16	H53926	yq87h02.r1 Homo sapien	2.76e+00
c	571	14	11.8	426	11	AA075025	zm83a06.r1 Stratagene	2.76e+00
	572	14	11.8	427	18	AI074508	oy59d08.x1 NCI_CGAP_Br	2.76e+00
	573	14	11.8	428	12	AA762244	vw58h08.r1 Soares mous	2.76e+00
	574	14	11.8	430	28	AQ164149	HS_2166_A1_D09_MR CIT	2.76e+00
c	575	14	11.8	430	18	AI147704	qb43b10.x1 NCI_CGAP_Br	2.76e+00
c	576	14	11.8	430	27	B85480	RPCI11-18E1.TPB RPCI11	2.76e+00
c	577	14	11.8	432	3	AQ139285	HS_3079_A2_C07_MF CIT	2.76e+00
c	578	14	11.8	432	16	H50941	yp84a01.r1 Homo sapien	2.76e+00
	579	14	11.8	433	10	T04693	740 Lambda-PRL2 Arabid	2.76e+00
	580	14	11.8	433	28	AQ211417	HS_3234_B2_A06_MR CIT	2.76e+00
c	581	14	11.8	433	28	AQ100926	HS_3060_B2_C12_MR CIT	2.76e+00
c	582	14	11.8	433	16	H50131	yo27a01.r1 Homo sapien	2.76e+00
c	583	14	11.8	433	17	AI075521	TENU3051 T. cruzi epim	2.76e+00
c	584	14	11.8	434	10	AA664186	ac05a06.s1 Stratagene	2.76e+00
c	585	14	11.8	434	12	AA775052	ac76f02.s1 Stratagene	2.76e+00
	586	14	11.8	435	8	AA562423	vl36h09.r1 Stratagene	2.76e+00
c	587	14	11.8	435	12	AA390473	LD08717.5prime LD Dros	2.76e+00
c	588	14	11.8	435	28	AQ113870	CIT-HSP-2375C2.TF CIT-	2.76e+00
	589	14	11.8	436	27	AQ008792	RPCI11-22L16.TVB RPCI1	2.76e+00
c	590	14	11.8	436	16	R76589	yi60b12.s1 Homo sapien	2.76e+00
c	591	14	11.8	436	6	C20127	Rice cDNA, partial seq	2.76e+00
	592	14	11.8	437	28	AQ217783	HS_2007_A1_B02_MR CIT	2.76e+00
c	593	14	11.8	437	8	AA133934	zo16f11.s1 Stratagene	2.76e+00
c	594	14	11.8	437	26	FR0007023	F.rubripes GSS sequenc	2.76e+00
	595	14	11.8	438	5	T09042	EST06934 Homo sapiens	2.76e+00
	596	14	11.8	438	10	AA427841	zw49e10.r1 Soares tota	2.76e+00
	597	14	11.8	438	26	B48215	RPCI11-6D6.TV RPCI11 H	2.76e+00
c	598	14	11.8	439	20	W01247	za40d10.r1 Soares feta	2.76e+00
c	599	14	11.8	439	16	H46778	yo14a06.r1 Homo sapien	2.76e+00
	600	14	11.8	440	19	H70646	yu19a03.r1 Homo sapien	2.76e+00
c	601	14	11.8	440	22	AA198257	mv47g07.r1 Soares mous	2.76e+00
c	602	14	11.8	440	26	B32897	HS-1016-A1-E12-MR.abi	2.76e+00
c	603	14	11.8	440	8	AA292004	zt28b08.r1 Soares ovar	2.76e+00

c	604	14	11.8	442	15	AU008876	Schizosaccharomyces po	2.76e+00
c	605	14	11.8	442	12	AA729418	nx36c05.s1 NCI_CGAP_GC	2.76e+00
c	606	14	11.8	442	29	AQ244269	HS_3204_B2_F09_MR CIT	2.76e+00
c	607	14	11.8	442	5	M79731	wEST00268 Caenorhabdit	2.76e+00
c	608	14	11.8	443	14	AA922533	om64d03.s1 NCI_CGAP_GC	2.76e+00
c	609	14	11.8	443	5	T99786	ye68e04.s1 Homo sapien	2.76e+00
c	610	14	11.8	443	29	AQ144432	HS_3084_A1_C07_MR CIT	2.76e+00
c	611	14	11.8	443	10	AA639881	nq85g05.s1 NCI_CGAP_Co	2.76e+00
c	612	14	11.8	444	21	W15238	zc19e02.s1 Soares para	2.76e+00
c	613	14	11.8	444	14	AA975738	oq64c07.s1 NCI_CGAP_Ki	2.76e+00
c	614	14	11.8	444	27	AQ074868	CIT-HSP-2368D13.TR CIT	2.76e+00
	615	14	11.8	445	26	AQ018645	CIT-HSP-2308J22.TR CIT	2.76e+00
c	616	14	11.8	445	18	AI056720	oy53f12.x1 NCI_CGAP_Br	2.76e+00
c	617	14	11.8	445	13	AA206103	zq54f12.s1 Stratagene	2.76e+00
c	618	14	11.8	446	5	T66735	ya49f02.s1 Homo sapien	2.76e+00
c	619	14	11.8	446	12	AA762476	vv80c06.r1 Stratagene	2.76e+00
c	620	14	11.8	446	16	H01342	yi99g07.s1 Homo sapien	2.76e+00
	621	14	11.8	447	7	AA457592	aa92f11.s1 Stratagene	2.76e+00
	622	14	11.8	447	11	AA692872	vr59g05.s1 Knowles Sol	2.76e+00
c	623	14	11.8	447	27	AQ053566	RPCI11-51D22.TJ RPCI11	2.76e+00
	624	14	11.8	448	29	AQ252101	HS_3080_A1_E06_MR CIT	2.76e+00
	625	14	11.8	448	10	AA634550	zu76h02.s1 Soares test	2.76e+00
c	626	14	11.8	448	9	AA553892	nk95b03.s1 NCI_CGAP_Co	2.76e+00
c	627	14	11.8	448	10	AA428407	zw57f08.s1 Soares tota	2.76e+00
c	628	14	11.8	448	19	N30250	yw64q04.s1 Homo sapien	2.76e+00
c	629	14	11.8	449	19	H94915	yu57d02.r1 Homo sapien	2.76e+00
c	630	14	11.8	450	26	B41932	HS-1054-B1-F12-MF.abi	2.76e+00
c	631	14	11.8	451	10	AA603666	np21h06.s1 NCI_CGAP_GC	2.76e+00
	632	14	11.8	452	28	AQ174260	HS_3206_A2_H09_MR CIT	2.76e+00
c	633	14	11.8	452	17	AI075001	oq97e10.x1 NCI_CGAP_Co	2.76e+00
	634	14	11.8	453	27	B60477	CIT-HSP-2008M15.TRB CI	2.76e+00
	635	14	11.8	453	9	AA513950	nf53b01.s1 NCI_CGAP_Al	2.76e+00
	636	14	11.8	453	8	AA398513	zt75c02.s1 Soares test	2.76e+00
c	637	14	11.8	453	12	AA792286	vo01d11.r1 Stratagene	2.76e+00
c	638	14	11.8	454	28	AQ177109	HS_3120_A1_C04_MR CIT	2.76e+00
c	639	14	11.8	454	27	FR0034146	Fugu rubripes GSS sequ	2.76e+00
c	640	14	11.8	455	28	AQ224407	HS_2005_B2_A08_T7 CIT	2.76e+00
c	641	14	11.8	455	16	H18136	yn47c06.r1 Homo sapien	2.76e+00
c	642	14	11.8	455	12	AA757420	zg36b12.s1 Soares pine	2.76e+00
c	643	14	11.8	456	11	AA681464	vr36h10.r1 Barstead mo	2.76e+00
	644	14	11.8	457	5	T69726	yd13c05.s1 Homo sapien	2.76e+00
c	645	14	11.8	457	29	AQ234955	HS_2056_A2_A05_T7 CIT	2.76e+00
c	646	14	11.8	459	10	T12954	1020 Lambda-PRL1 Arabi	2.76e+00
	647	14	11.8	460	8	C28978	Rice cDNA, partial seq	2.76e+00
	648	14	11.8	460	16	R83399	yp82b08.r1 Homo sapien	2.76e+00
c	649	14	11.8	462	27	B69381	CIT-HSP-2054A10.TF CIT	2.76e+00
c	650	14	11.8	462	17	AI088644	qb14h07.x1 Soares_preg	2.76e+00
c	651	14	11.8	462	11	T44319	7582 Lambda-PRL2 Arabi	2.76e+00
	652	14	11.8	463	29	AQ233277	HS_2259_B1_D02_T7 CIT	2.76e+00
c	653	14	11.8	463	15	AI057778	TENU1869 T. cruzi epim	2.76e+00
c	654	14	11.8	463	5	T67737	yc27g02.s1 Homo sapien	2.76e+00

c	655	14	11.8	463	23	AA260588	mx76d10.r1 Soares mous	2.76e+00
c	656	14	11.8	464	7	AA466533	ve12e10.r1 Soares mous	2.76e+00
	657	14	11.8	465	28	AQ224302	HS_2003_A2_H03_MR CIT	2.76e+00
c	658	14	11.8	465	18	AI139479	qc25a03.x1 Soares_preg	2.76e+00
	659	14	11.8	466	21	W87984	mf69b08.r1 Soares mous	2.76e+00
	660	14	11.8	466	21	W79425	zd88d03.s1 Soares feta	2.76e+00
	661	14	11.8	466	17	AI076991	BSBmMPSZ10H7SK Brugia	2.76e+00
c	662	14	11.8	466	12	AA791921	vs61h02.r1 Stratagene	2.76e+00
c	663	14	11.8	466	28	AQ176995	HS_3207_A1_H05_MR CIT	2.76e+00
	664	14	11.8	467	29	AQ225548	HS_2007_B2_H11_MR CIT	2.76e+00
c	665	14	11.8	467	18	AI086037	oy70h06.x1 NCI_CGAP_CL	2.76e+00
c	666	14	11.8	467	5	T72661	yd19e11.s1 Homo sapien	2.76e+00
c	667	14	11.8	467	13	AA188883	zp77f11.r1 Stratagene	2.76e+00
c	668	14	11.8	468	5	T74506	yc83a12.r1 Homo sapien	2.76e+00
c	669	14	11.8	468	26	B57348	CIT-HSP-2009F18.TF CIT	2.76e+00
c	670	14	11.8	469	6	AA436227	zv23h01.r1 Soares NhHM	2.76e+00
c	671	14	11.8	469	10	AA406047	zu67d11.s1 Soares test	2.76e+00
c	672	14	11.8	469	14	AA923295	om51c06.s1 NCI_CGAP_GC	2.76e+00
c	673	14	11.8	471	28	AQ121587	HS_3079_B2_F04_MF CIT	2.76e+00
	674	14	11.8	472	10	AA400143	zu69d06.s1 Soares test	2.76e+00
	675	14	11.8	473	29	AQ245189	HS_2008_A1_H02_T7 CIT	2.76e+00
c	676	14	11.8	473	9	N60236	TgESTxyl1a05.r1 TgRH T	2.76e+00
c	677	14	11.8	473	5	R07498	ye97f02.s1 Homo sapien	2.76e+00
c	678	14	11.8	473	10	AA643550	nq78d05.s1 NCI_CGAP_Co	2.76e+00
c	679	14	11.8	474	28	AQ194734	RPCI11-59E9.TJ RPCI11	2.76e+00
	680	14	11.8	475	28	AQ134744	HS_3052_B1_D09_MF CIT	2.76e+00
	681	14	11.8	475	19	N24302	yx13g07.r1 Homo sapien	2.76e+00
c	682	14	11.8	475	10	AA593084	nm97h12.s1 NCI_CGAP_Co	2.76e+00
c	683	14	11.8	476	9	AA592622	vo20h11.r1 Barstead mo	2.76e+00
	684	14	11.8	477	27	B87619	RPCI11-29M17.TV RPCI11	2.76e+00
	685	14	11.8	477	26	B52826	CIT-HSP-2007C3.TR CIT-	2.76e+00
	686	14	11.8	477	24	AA306667	EST177816 Jurkat T-cel	2.76e+00
	687	14	11.8	477	26	B48217	RPCI11-6D8.TV RPCI11 H	2.76e+00
c	688	14	11.8	477	28	AQ121940	HS_3090_A1_E07_MF CIT	2.76e+00
c	689	14	11.8	478	26	B47916	RPCI11-3G9.TP RPCI11 H	2.76e+00
c	690	14	11.8	478	27	AQ040458	CIT-HSP-2324C18.TF CIT	2.76e+00
c	691	14	11.8	478	16	H02101	yj38b01.r1 Homo sapien	2.76e+00
c	692	14	11.8	479	17	AI115227	ui47g04.x1 Sugano mous	2.76e+00
c	693	14	11.8	479	15	AA913822	ol39b12.s1 Soares_NFL_	2.76e+00
	694	14	11.8	480	28	AQ224067	HS_2011_A1_A10_MR CIT	2.76e+00
c	695	14	11.8	480	15	AA936148	om57e02.s1 NCI_CGAP_GC	2.76e+00
c	696	14	11.8	480	27	B87651	RPCI11-29P10.TP RPCI11	2.76e+00
c	697	14	11.8	480	16	R67293	yh08c12.r2 Homo sapien	2.76e+00
c	698	14	11.8	480	22	AA024406	ze74e09.s1 Soares feta	2.76e+00
	699	14	11.8	481	14	AA836900	of35f11.s1 NCI_CGAP_Ki	2.76e+00
c	700	14	11.8	481	5	T56182	yb79g04.s1 Homo sapien	2.76e+00
	701	14	11.8	482	28	AQ192518	HS_2251_A2_H11_MR CIT	2.76e+00
c	702	14	11.8	482	28	AQ116505	HS_2191_A2_G04_MR CIT	2.76e+00
	703	14	11.8	483	28	AQ221336	HS_2008_A1_G08_MR CIT	2.76e+00
	704	14	11.8	483	12	AA753679	97BS0170 Rice Immature	2.76e+00
c	705	14	11.8	483	28	AQ133361	HS_3045_A1_C06_MR CIT	2.76e+00

	706	14	11.8	484	28	AQ123258	HS_3090_B1_C03_T7 CIT	2.76e+00
c	707	14	11.8	484	17	AI075499	TENU3092 T. cruzi epim	2.76e+00
c	708	14	11.8	484	15	AA932095	oo62f01.s1 NCI_CGAP_Lu	2.76e+00
c	709	14	11.8	485	29	AQ151506	HS_2218_B2_G03_MR CIT	2.76e+00
c	710	14	11.8	485	5	R36977	yf53h07.s1 Homo sapien	2.76e+00
c	711	14	11.8	485	5	R33090	yh77d06.s1 Homo sapien	2.76e+00
c	712	14	11.8	486	8	AA483389	ne74d10.s1 NCI_CGAP_Ew	2.76e+00
c	713	14	11.8	486	26	B73183	RPCI11-11M13.TV RPCI11	2.76e+00
c	714	14	11.8	487	5	R37629	yf61a01.s1 Homo sapien	2.76e+00
	715	14	11.8	488	27	AQ074302	CIT-HSP-2382L21.TR CIT	2.76e+00
	716	14	11.8	488	28	AQ171134	HS_3076_A2_G12_MF CIT	2.76e+00
c	717	14	11.8	488	10	AA593075	nm97g12.s1 NCI_CGAP_Co	2.76e+00
c	718	14	11.8	489	13	AA840399	vx92b09.r1 Soares 2NbM	2.76e+00
c	719	14	11.8	489	6	AA387287	vc24h09.r1 Ko mouse em	2.76e+00
c	720	14	11.8	489	11	AA700766	zi65a03.s1 Soares feta	2.76e+00
	721	14	11.8	490	27	B74548	CIT-HSP-2028N16.TR CIT	2.76e+00
	722	14	11.8	490	19	N20869	yx52e10.s1 Homo sapien	2.76e+00
c	723	14	11.8	490	16	H49971	yo26a01.s1 Homo sapien	2.76e+00
c	724	14	11.8	490	11	AA705190	zj96d11.s1 Soares feta	2.76e+00
c	725	14	11.8	490	13	AA869201	vq48h01.r1 Barstead bo	2.76e+00
c	726	14	11.8	491	17	AI129794	qc45d08.x1 Soares_preg	2.76e+00
c	727	14	11.8	491	6	AA432191	zw71h05.s1 Soares test	2.76e+00
c	728	14	11.8	491	5	T66304	yc78c11.s1 Homo sapien	2.76e+00
c	729	14	11.8	492	29	AQ229457	HS_2023_A2_D05_T7 CIT	2.76e+00
c	730	14	11.8	493	5	T70257	yd26b05.s1 Homo sapien	2.76e+00
c	731	14	11.8	493	5	R19554	yg26h02.r1 Homo sapien	2.76e+00
	732	14	11.8	494	13	AA504662	ab02a09.r1 Stratagene	2.76e+00
	733	14	11.8	494	27	AQ054860	CIT-HSP-2344H14.TF CIT	2.76e+00
c	734	14	11.8	495	21	AA043287	zk49h04.s1 Soares preg	2.76e+00
	735	14	11.8	496	11	N96541	21216 CD4-14 Arabidops	2.76e+00
c	736	14	11.8	496	14	AA899585	UI-R-E0-cy-e-01-0-UI.s	2.76e+00
c	737	14	11.8	496	13	AA190514	zp86g04.s1 Stratagene	2.76e+00
c	738	14	11.8	496	12	AA755833	vu04c12.r1 Soares mous	2.76e+00
	739	14	11.8	497	28	AQ215978	HS_3238_B2_H01_T7 CIT	2.76e+00
c	740	14	11.8	497	29	AQ237175	RPCI11-65P21.TJ RPCI11	2.76e+00
	741	14	11.8	498	29	AQ232887	HS_2028_B2_G08_MR CIT	2.76e+00
c	742	14	11.8	498	29	AQ245994	HS_2059_A2_G08_T7 CIT	2.76e+00
	743	14	11.8	499	23	AA123843	mp93f03.r1 Soares 2NbM	2.76e+00
c	744	14	11.8	499	28	AQ207331	HS_3240_B1_D05_T7 CIT	2.76e+00
	745	14	11.8	500	28	AQ217425	HS_3245_B1_G03_T7 CIT	2.76e+00
	746	14	11.8	500	10	AA625342	af68f11.r1 Soares NhHM	2.76e+00
	747	14	11.8	500	28	AQ218859	HS_3248_B1_F01_T7 CIT	2.76e+00
c	748	14	11.8	500	16	R55601	yg88g06.r1 Homo sapien	2.76e+00
c	749	14	11.8	501	29	AQ149859	HS_3189_B2_E09_MR CIT	2.76e+00
c	750	14	11.8	501	9	AA554102	nk34h03.s1 NCI_CGAP_Co	2.76e+00
	751	14	11.8	503	27	AQ065183	HS_2183_A2_C02_MR CIT	2.76e+00
	752	14	11.8	505	12	AA790443	vw04f04.r1 Soares mous	2.76e+00
	753	14	11.8	506	5	T84052	yd76f09.s1 Homo sapien	2.76e+00
	754	14	11.8	506	11	N64998	20038 Lambda-PRL2 Arab	2.76e+00
c	755	14	11.8	506	11	AA677651	ae29g08.s1 Soares NbHF	2.76e+00
c	756	14	11.8	507	9	AA577431	nm96g12.s1 NCI_CGAP_Co	2.76e+00

c	757	14	11.8	507	9	AA571431	vm05e06.r1 Knowles Sol	2.76e+00
c	758	14	11.8	508	6	AA011714	zi03e03.r1 Soares feta	2.76e+00
c	759	14	11.8	508	11	AA075534	zm88b03.s1 Stratagene	2.76e+00
	760	14	11.8	509	27	AQ008069	RPCI11-20A8.TVB RPCI11	2.76e+00
	761	14	11.8	509	15	AA931911	oo62a09.s1 NCI_CGAP_Lu	2.76e+00
	762	14	11.8	509	8	AA470032	zu10h08.r1 Soares test	2.76e+00
c	763	14	11.8	509	11	AA695664	GM03221.5prime GM Dros	2.76e+00
	764	14	11.8	510	27	AQ013271	RPCI11-23B8.TKBR RPCI1	2.76e+00
c	765	14	11.8	510	27	B65670	CIT-HSP-2022K19.TR CIT	2.76e+00
c	766	14	11.8	510	28	AQ190519	HS_3227_A2_F04_MR CIT	2.76e+00
c	767	14	11.8	510	9	AA577752	nn23e09.s1 NCI_CGAP_Ga	2.76e+00
	768	14	11.8	511	17	AI071420	UI-R-C2-nb-f-09-0-UI.s	2.76e+00
	769	14	11.8	512	16	R56075	yg91f01.s1 Homo sapien	2.76e+00
	770	14	11.8	512	27	B57936	CIT-HSP-2011k19.TR CIT	2.76e+00
c	771	14	11.8	513	13	AA167678	sq39e07.r1 Stratagene	2.76e+00
	772	14	11.8	515	22	AA029648	ze94a12.r1 Soares feta	2.76e+00
c	773	14	11.8	515	6	AA387283	vc24g09.r1 Ko mouse em	2.76e+00
c	774	14	11.8	517	8	AA243477	za15f08.s1 NCI_CGAP_GC	2.76e+00
c	775	14	11.8	517	11	AA670452	ae62g12.s1 Stratagene	2.76e+00
c	776	14	11.8	517	8	AA552969	nk83b03.s1 NCI_CGAP_Sc	2.76e+00
c	777	14	11.8	517	10	AA599893	ag32f08.s1 Jia bone ma	2.76e+00
	778	14	11.8	518	17	AA907205	ol03f01.s1 NCI_CGAP_Lu	2.76e+00
c	779	14	11.8	518	6	AA412220	zu10d08.s1 Soares test	2.76e+00
c	780	14	11.8	519	28	AQ173649	HS_3204_B1_B10_T7 CIT	2.76e+00
c	781	14	11.8	519	6	AA387279	vc24f09.r1 Ko mouse em	2.76e+00
c	782	14	11.8	520	7	AA449048	zx07g05.r1 Soares tota	2.76e+00
	783	14	11.8	521	5	T79576	yd71a03.s1 Homo sapien	2.76e+00
c	784	14	11.8	521	15	AA902530	oj58c08.s1 NCI_CGAP_Ki	2.76e+00
c	785	14	11.8	524	28	AQ179916	HS_3203_B1_B06_T7 CIT	2.76e+00
c	786	14	11.8	524	13	AA613885	no81g12.s1 NCI_CGAP_AA	2.76e+00
c	787	14	11.8	525	26	B38729	HS-1048-A2-B10-MF.abi	2.76e+00
	788	14	11.8	527	28	AQ176054	HS_3209_A1_E01_T7 CIT	2.76e+00
c	789	14	11.8	527	27	FR0029028	Fugu rubripes GSS sequ	2.76e+00
c	790	14	11.8	528	28	AQ195170	RPCI11-66K5.TJ RPCI11	2.76e+00
	791	14	11.8	529	11	AA658071	nu17h01.s1 NCI_CGAP_Pr	2.76e+00
c	792	14	11.8	529	10	AA427425	zw54a08.s1 Soares tota	2.76e+00
	793	14	11.8	530	14	AA848828	EST191589 Normalized r	2.76e+00
	794	14	11.8	530	11	N96567	21242 CD4-14 Arabidops	2.76e+00
c	795	14	11.8	530	17	AA934255	ka04h12.s1 Onchocerca	2.76e+00
c	796	14	11.8	530	14	AA950314	LD28833.5prime LD Dros	2.76e+00
c	797	14	11.8	532	28	AQ189863	HS_3219_A2_C02_T7 CIT	2.76e+00
c	798	14	11.8	533	28	AQ174841	HS_3211_B2_F08_T7 CIT	2.76e+00
c	799	14	11.8	534	29	AQ143769	HS_3075_B1_E06_MF CIT	2.76e+00
c	800	14	11.8	534	29	AQ229864	HS_3027_A2_F09_MR CIT	2.76e+00
c	801	14	11.8	535	13	AA216527	zq95d07.s1 Stratagene	2.76e+00
c	802	14	11.8	535	11	H36506	14174 Lambda-PRL2 Arab	2.76e+00
c	803	14	11.8	537	14	AA915489	vz31e12.r1 Soares 2NbM	2.76e+00
c	804	14	11.8	537	29	AQ141617	HS_3164_A1_D08_T7 CIT	2.76e+00
c	805	14	11.8	539	10	AA588407	nm93g03.s1 NCI_CGAP_Co	2.76e+00
c	806	14	11.8	540	11	AA075581	zm88h06.s1 Stratagene	2.76e+00
c	807	14	11.8	540	18	AI096489	qa03d07.x1 NCI_CGAP_Br	2.76e+00

c	808	14	11.8	541	13	AA176242	zp08e07.s1 Stratagene	2.76e+00
	809	14	11.8	542	26	B49243	RPCI11-5J11.TP RPCI11	2.76e+00
c	810	14	11.8	542	6	AA156869	zl20g08.s1 Soares preg	2.76e+00
c	811	14	11.8	543	26	B29479	F25D14TFB IGF Arabidop	2.76e+00
	812	14	11.8	544	28	AQ213639	HS_3250_A1_A06_T7 CIT	2.76e+00
c	813	14	11.8	545	8	AA483552	ne69g02.s1 NCI_CGAP_Al	2.76e+00
	814	14	11.8	546	11	AA707680	zj29h03.s1 Soares feta	2.76e+00
	815	14	11.8	547	28	AQ171502	HS_3088_B1_C11_MF CIT	2.76e+00
	816	14	11.8	548	23	AA072103	mm40h05.r1 Stratagene	2.76e+00
c	817	14	11.8	549	26	B24867	F22I22TF IGF Arabidops	2.76e+00
c	818	14	11.8	549	11	AA700472	zj74e09.s1 Soares feta	2.76e+00
	819	14	11.8	550	27	AQ045451	RPCI11-33A7.TJ RPCI11	2.76e+00
c	820	14	11.8	551	10	AA619649	v101e12.r1 Knowles Sol	2.76e+00
	821	14	11.8	554	27	AQ046684	RPCI11-35H16.TJ RPCI11	2.76e+00
	822	14	11.8	556	8	AA397662	zt87c10.r1 Soares test	2.76e+00
c	823	14	11.8	556	15	AI063481	GH03320.5prime GH Dros	2.76e+00
c	824	14	11.8	557	5	T63058	yc15h11.s1 Homo sapien	2.76e+00
c	825	14	11.8	557	27	B65878	CIT-HSP-2023G10.TF CIT	2.76e+00
c	826	14	11.8	558	13	AA599322	ag25b10.s1 Jia bone ma	2.76e+00
c	827	14	11.8	559	10	AA477706	zu44d09.s1 Soares ovar	2.76e+00
	828	14	11.8	562	10	AA424430	zv82a11.r1 Soares tota	2.76e+00
	829	14	11.8	562	22	AA191965	rs06h01.r1 Sommer Pris	2.76e+00
c	830	14	11.8	562	21	N95605	zb66h11.s1 Soares feta	2.76e+00
	831	14	11.8	566	15	AA911716	ok88g09.s1 NCI_CGAP_Lu	2.76e+00
	832	14	11.8	568	17	AI106853	GH06093.3prime GH Dros	2.76e+00
c	833	14	11.8	573	10	AA621824	nq24g11.s1 NCI_CGAP_Co	2.76e+00
c	834	14	11.8	573	27	B92275	CIT-HSP-2167M18.TF CIT	2.76e+00
	835	14	11.8	574	26	FR0017997	F.rubripes GSS sequenc	2.76e+00
	836	14	11.8	574	26	FR0024390	F.rubripes GSS sequenc	2.76e+00
c	837	14	11.8	574	17	AI077088	TENU3235 T. cruzi epim	2.76e+00
c	838	14	11.8	575	16	H43252	yp05f08.r1 Homo sapien	2.76e+00
c	839	14	11.8	576	28	AQ179954	HS_3203_B1_G05_T7 CIT	2.76e+00
c	840	14	11.8	577	6	AA034038	zi05d02.s1 Soares feta	2.76e+00
c	841	14	11.8	580	14	AA830687	oo54g03.s1 NCI_CGAP_GC	2.76e+00
c	842	14	11.8	582	18	AI094703	oy63h11.s1 NCI_CGAP_Br	2.76e+00
c	843	14	11.8	583	8	AA544526	vk39a06.r1 Soares mous	2.76e+00
	844	14	11.8	584	17	AU020539	Mus musculus 8-cell em	2.76e+00
	845	14	11.8	585	6	AA402762	zt63c12.r1 Soares test	2.76e+00
	846	14	11.8	585	27	AQ039110	CIT-HSP-2335C1.TR CIT-	2.76e+00
	847	14	11.8	585	20	HSF4G5	H.sapiens mRNA for exp	2.76e+00
	848	14	11.8	586	16	H04995	yl74c12.r1 Homo sapien	2.76e+00
c	849	14	11.8	587	15	AI064230	GH04561.5prime GH Dros	2.76e+00
c	850	14	11.8	589	17	AU020569	Mus musculus 8-cell em	2.76e+00
	851	14	11.8	590	28	AQ159343	nbxb0013N15r CUGI Rice	2.76e+00
	852	14	11.8	591	17	AI110022	GH09448.5prime GH Dros	2.76e+00
c	853	14	11.8	594	9	AA577672	nn22e12.s1 NCI_CGAP_Co	2.76e+00
c	854	14	11.8	595	24	AA286641	vb75c11.r1 Soares mous	2.76e+00
c	855	14	11.8	599	27	FR0034359	Fugu rubripes GSS sequ	2.76e+00
	856	14	11.8	603	10	AA650663	30777 Lambda-PRL2 Arab	2.76e+00
	857	14	11.8	606	28	AQ193525	CIT-HSP-2381N21.TR CIT	2.76e+00
	858	14	11.8	607	26	B26691	T16A4TF TAMU Arabidops	2.76e+00

c	859	14	11.8	608	27	AQ074284	100 pGEM4 PstI Rice Et	2.76e+00
c	860	14	11.8	610	11	H36137	14659 Lambda-PRL2 Arab	2.76e+00
c	861	14	11.8	610	13	AA629750	ae63b11.s1 Stratagene	2.76e+00
	862	14	11.8	612	26	B56278	CIT-HSP-2006I11.TR CIT	2.76e+00
c	863	14	11.8	612	11	C84131	Dictyostelium discoide	2.76e+00
c	864	14	11.8	614	28	AQ157608	nbxb0009D18r CUGI Rice	2.76e+00
c	865	14	11.8	616	21	AA049263	mj49g08.r1 Soares mous	2.76e+00
c	866	14	11.8	616	28	AQ115789	RPCI11-55F16.TK RPCI11	2.76e+00
c	867	14	11.8	618	19	N28825	yx71g09.r1 Homo sapien	2.76e+00
	868	14	11.8	619	26	FR0024385	F.rubripes GSS sequenc	2.76e+00
	869	14	11.8	628	19	N24441	yx26d11.r1 Homo sapien	2.76e+00
c	870	14	11.8	628	5	R47227	ph2e8_19/1TV Homo sapi	2.76e+00
c	871	14	11.8	630	27	AQ075265	CIT-HSP-2369M6.TR CIT-	2.76e+00
	872	14	11.8	631	29	AQ234077	HS_2057_A1_A02_MR CIT	2.76e+00
c	873	14	11.8	633	8	AA489098	aa56h08.s1 NCI_CGAP_GC	2.76e+00
c	874	14	11.8	639	23	AA059473	zf66f03.s1 Soares reti	2.76e+00
	875	14	11.8	644	28	AQ201231	RPCI11-45I20.TJ RPCI11	2.76e+00
c	876	14	11.8	650	23	AA096877	mo02b02.r1 Stratagene	2.76e+00
c	877	14	11.8	650	28	AQ159069	nbxb0012J24r CUGI Rice	2.76e+00
	878	14	11.8	651	27	AQ078818	CIT-HSP-2367F19.TR CIT	2.76e+00
c	879	14	11.8	651	10	AA660478	00364 MtRHE Medicago t	2.76e+00
	880	14	11.8	652	10	AA660835	00729 MtRHE Medicago t	2.76e+00
c	881	14	11.8	654	27	AQ045909	RPCI11-35N3.TK RPCI11	2.76e+00
c	882	14	11.8	655	27	AQ034618	CIT-HSP-2329H17.TR CIT	2.76e+00
	883	14	11.8	658	29	AQ235048	HS_2046_A2_E11_MR CIT	2.76e+00
	884	14	11.8	664	10	AA660762	00654 MtRHE Medicago t	2.76e+00
	885	14	11.8	672	23	AA198680	mu15d08.r1 Soares 2NbM	2.76e+00
c	886	14	11.8	676	17	AI116617	ue27h11.y1 Sugano mous	2.76e+00
c	887	14	11.8	676	17	AI107632	GH05416.5prime GH Dros	2.76e+00
	888	14	11.8	685	27	B70324	CIT-HSP-2028O8.TFB CIT	2.76e+00
c	889	14	11.8	685	13	AA883898	am26h06.s1 Soares NFL	2.76e+00
c	890	14	11.8	687	12	AA263715	LD06910.5prime LD Dros	2.76e+00
	891	14	11.8	688	27	AQ074160	132 pUC8 PstI Rice Eti	2.76e+00
c	892	14	11.8	689	8	AA550890	1306m3 gmbPfHB3.1, G.	2.76e+00
	893	14	11.8	690	26	B71928	RPCI11-13G3.TV RPCI11	2.76e+00
c	894	14	11.8	691	11	C84141	Dictyostelium discoide	2.76e+00
	895	14	11.8	698	9	HSZ99431	Homo sapiens mRNA; exp	2.76e+00
	896	14	11.8	698	19	N32782	yw91g03.s1 Homo sapien	2.76e+00
	897	14	11.8	698	27	AQ078372	CIT-HSP-2365E16.TF CIT	2.76e+00
c	898	14	11.8	702	27	AQ056056	CIT-HSP-2338L20.TR CIT	2.76e+00
	899	14	11.8	707	26	AQ018949	CIT-HSP-2307L8.TR CIT-	2.76e+00
	900	14	11.8	708	6	AA402711	zu48d09.r1 Soares ovar	2.76e+00
c	901	14	11.8	708	26	AG006108	Homo sapiens genomic D	2.76e+00
c	902	14	11.8	708	28	AG014097	Homo sapiens genomic D	2.76e+00
	903	14	11.8	710	28	AQ112427	CIT-HSP-2376P4.TR CIT-	2.76e+00
	904	14	11.8	711	22	AA155377	mt64a11.r1 Soares mous	2.76e+00
	905	14	11.8	723	28	AQ157853	nbxb0010O11r CUGI Rice	2.76e+00
	906	14	11.8	728	27	AQ054414	CIT-HSP-2341C8.TF CIT-	2.76e+00
	907	14	11.8	739	10	AA651697	nn47c04.r1 NCI_CGAP_Ki	2.76e+00
	908	14	11.8	740	28	AQ158660	nbxb0011N14f CUGI Rice	2.76e+00
	909	14	11.8	740	27	AQ079854	CIT-HSP-2367N7.TF CIT-	2.76e+00

	910	14	11.8	740	26	B03808	cSRL-187G4-u cSRL flow	2.76e+00
c	911	14	11.8	742	20	W26172	23f6 Human retina cDNA	2.76e+00
	912	14	11.8	744	6	AA131249	zl31f02.r1 Soares preg	2.76e+00
c	913	14	11.8	754	29	AQ254252	CpG0727B CpIOWAgDNA1 C	2.76e+00
	914	14	11.8	758	27	B65771	CIT-HSP-2023A18.TR CIT	2.76e+00
c	915	14	11.8	760	27	AQ048907	cLM-10h1-u cLM Giardia	2.76e+00
c	916	14	11.8	763	27	AQ050330	nbxb0003dD02r CUGI Ric	2.76e+00
	917	14	11.8	772	18	AU035510	Mus musculus cDNA, clo	2.76e+00
	918	14	11.8	773	26	B01206	cSRL-128e10-u cSRL flo	2.76e+00
c	919	14	11.8	779	17	AI119338	uf03d04.y1 Sugano mous	2.76e+00
c	920	14	11.8	788	27	B65630	CIT-HSP-2022I15.TF CIT	2.76e+00
c	921	14	11.8	791	27	B81467	CIT-HSP-2059M16.TFB CI	2.76e+00
c	922	14	11.8	801	12	AA263288	LD06379.5prime LD Dros	2.76e+00
c	923	14	11.8	814	29	AF046725	Mus musculus clone OST	2.76e+00
	924	14	11.8	832	20	W28836	52d2 Human retina cDNA	2.76e+00
c	925	14	11.8	838	26	B20571	T20C11-T7 TAMU Arabido	2.76e+00
	926	14	11.8	846	26	B18559	F13P20-Sp6 IGF Arabido	2.76e+00
	927	14	11.8	863	26	B20136	T9M12-T7 TAMU Arabidop	2.76e+00
c	928	14	11.8	891	18	AU035291	Mus musculus cDNA, clo	2.76e+00
c	929	14	11.8	918	7	C24329	Dictyostelium discoide	2.76e+00
	930	14	11.8	956	26	B13104	T8E16-Sp6 TAMU Arabido	2.76e+00
	931	14	11.8	1012	26	B08553	T31F5-Sp6 TAMU Arabido	2.76e+00
c	932	14	11.8	1325	26	B12001	T15I16-Sp6.1 TAMU Arab	2.76e+00
c	933	13	10.9	132	7	HUM0C12C04	Human HepG2 3'-directe	5.70e+01
	934	13	10.9	135	5	T57375	yb51d05.s1 Homo sapien	5.70e+01
c	935	13	10.9	190	5	T55693	yb39g11.r1 Homo sapien	5.70e+01
c	936	13	10.9	260	19	N21940	SMNHADA020052SK Schist	5.70e+01
c	937	13	10.9	266	24	AA313192	EST18507 Liver, subtra	5.70e+01
c	938	13	10.9	267	24	AA322463	EST25101 Cerebellum II	5.70e+01
c	939	13	10.9	279	23	AA250496	mz59h07.r1 Soares mous	5.70e+01
	940	13	10.9	299	25	AA342280	EST47841 Fetal spleen	5.70e+01
	941	13	10.9	303	5	T74880	yc58c05.s1 Homo sapien	5.70e+01
	942	13	10.9	304	5	R22444	yh23g07.s1 Homo sapien	5.70e+01
	943	13	10.9	305	6	AA383068	EST96401 Testis I Homo	5.70e+01
	944	13	10.9	310	5	T56815	ya83a04.s2 Homo sapien	5.70e+01
	945	13	10.9	315	5	T61120	yc45f12.s1 Homo sapien	5.70e+01
c	946	13	10.9	319	18	AI086611	or76e06.x1 Soares_sene	5.70e+01
	947	13	10.9	323	17	AI125705	qd95b06.x1 Soares_test	5.70e+01
c	948	13	10.9	324	5	R10530	yf31a07.r1 Homo sapien	5.70e+01
c	949	13	10.9	326	16	M78370	EST00518 Homo sapiens	5.70e+01
c	950	13	10.9	330	6	C20715	HUMGS0004740, Human Ge	5.70e+01
	951	13	10.9	336	14	AA848793	EST191554 Normalized r	5.70e+01
c	952	13	10.9	342	24	AA313934	EST185791 Colon carcin	5.70e+01
c	953	13	10.9	348	27	AQ088927	HS_3004_B1_D12_MR CIT	5.70e+01
	954	13	10.9	352	5	T55858	yb82e05.s1 Homo sapien	5.70e+01
	955	13	10.9	360	9	C65873	C.elegans cDNA clone y	5.70e+01
c	956	13	10.9	369	6	AA133251	zl16b11.r1 Soares preg	5.70e+01
c	957	13	10.9	371	9	C64664	C.elegans cDNA clone y	5.70e+01
c	958	13	10.9	373	24	AA313117	EST18464 Liver, subtra	5.70e+01
	959	13	10.9	374	16	R96757	yq55h10.s1 Homo sapien	5.70e+01
	960	13	10.9	375	5	T67408	yc26a09.s1 Homo sapien	5.70e+01

c	961	13	10.9	383	7	MUS86D12	Mouse embryonal carcin 5.70e+01
	962	13	10.9	387	5	T47711	yb16e11.s1 Homo sapien 5.70e+01
	963	13	10.9	395	5	T48276	yb44e06.s1 Homo sapien 5.70e+01
c	964	13	10.9	402	8	AA482095	zv42c09.s1 Soares ovar 5.70e+01
c	965	13	10.9	403	13	AA824023	vq93h08.r1 Knowles Sol 5.70e+01
	966	13	10.9	409	5	T53216	ya86d05.s1 Homo sapien 5.70e+01
c	967	13	10.9	411	27	AQ024243	HS_2182_A1_H01_MF CIT 5.70e+01
	968	13	10.9	428	5	T54787	yb42b11.s1 Homo sapien 5.70e+01
	969	13	10.9	432	5	T74194	yc60f10.s1 Homo sapien 5.70e+01
	970	13	10.9	433	5	T73434	yc35c05.s1 Homo sapien 5.70e+01
	971	13	10.9	440	22	AA197128	zq10g01.r1 Stratagene 5.70e+01
c	972	13	10.9	444	6	AA027958	zk05d12.r1 Soares preg 5.70e+01
c	973	13	10.9	446	8	AA226524	nc26c10.r1 NCI_CGAP_Pr 5.70e+01
	974	13	10.9	447	5	T73454	yc35e09.s1 Homo sapien 5.70e+01
	975	13	10.9	453	5	T71859	yc64f08.s1 Homo sapien 5.70e+01
	976	13	10.9	455	19	H83352	yv82f01.r1 Homo sapien 5.70e+01
	977	13	10.9	462	16	H02975	yj46h06.r1 Homo sapien 5.70e+01
c	978	13	10.9	468	27	AQ002049	CIT-HSP-2288E17.TR CIT 5.70e+01
	979	13	10.9	472	21	AA039014	mi99b01.r1 Soares mous 5.70e+01
c	980	13	10.9	476	28	AQ124496	HS_3145_B1_D12_T7 CIT 5.70e+01
c	981	13	10.9	483	17	AI069580	TENU2640 T. cruzi epim 5.70e+01
	982	13	10.9	490	14	AA878819	of87a02.s1 NCI_CGAP_Li 5.70e+01
	983	13	10.9	494	18	AI139674	qc20a06.x1 Soares_feta 5.70e+01
c	984	13	10.9	494	17	AA948090	ok22g10.s1 Soares_NSF_ 5.70e+01
c	985	13	10.9	497	28	AQ233565	HS_2045_B2_C08_MR CIT 5.70e+01
	986	13	10.9	498	5	T53877	yb85f12.s1 Homo sapien 5.70e+01
	987	13	10.9	504	17	AI127447	qb75d02.x1 Soares_feta 5.70e+01
	988	13	10.9	509	5	T69118	yc32a10.s1 Homo sapien 5.70e+01
	989	13	10.9	514	5	T70918	yc49g05.s1 Homo sapien 5.70e+01
	990	13	10.9	523	18	AI143331	qb71c06.x1 Soares_feta 5.70e+01
c	991	13	10.9	530	10	AA611247	vo51a02.r1 Barstead mo 5.70e+01
c	992	13	10.9	534	27	AQ058679	CIT-HSP-2344D13.TR CIT 5.70e+01
	993	13	10.9	538	18	AI187111	qe38d12.s1 Soares_feta 5.70e+01
c	994	13	10.9	540	28	AQ274429	HS_3059_B1_H03_T7 CIT 5.70e+01
	995	13	10.9	545	27	B70948	CIT-HSP-2063G21.TF CIT 5.70e+01
c	996	13	10.9	547	18	AI122776	qa48g03.x1 Soares_NhHM 5.70e+01
	997	13	10.9	570	5	T68294	yc41c09.s1 Homo sapien 5.70e+01
c	998	13	10.9	593	27	AQ013370	RPCI11-23F8.TKBP RPCI1 5.70e+01
	999	13	10.9	670	22	AA129470	zn78a01.s1 Stratagene 5.70e+01
c	1000	13	10.9	684	27	AQ051498	nbxb0002bD01r CUGI Ric 5.70e+01

ALIGNMENTS

```
RESULT    1
LOCUS       AA479058       404 bp    mRNA            EST       08-AUG-1997
DEFINITION  zv18g02.s1 Soares NhHMPu S1 Homo sapiens cDNA clone 754034 3', mRNA
            sequence.
ACCESSION   AA479058
NID         g2207614
KEYWORDS    EST.
```



```
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 404)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Kucaba,T., Leahy,M., Le,N., Lennon,G., Marra,M., Martin,J.,
            Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-Merck EST Project 1997
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -41m13 fwd. ET from Amersham
            High quality sequence stop: 279.
FEATURES              Location/Qualifiers
     source           1. .404
                      /organism="Homo sapiens"
                      /note="Organ: mixed (see below); Vector: pT7T3D-Pac
                      (Pharmacia) with a modified polylinker; Site_1: Not I;
                      Site_2: Eco RI; Equal amounts of plasmid DNA from three
                      normalized libraries (melanocyte 2NbHM, pregnant uterus
                      NbHPU, and fetal heart NbHH19W) were mixed, and ss circles
                      were made in vitro. Following HAP purification, this DNA
                      was used as tracer in a subtractive hybridization
                      reaction. The driver was PCR-amplified cDNAs from pools of
                      5,000 clones made from the same 3 libraries. The pools
                      consisted of I.M.A.G.E. clones 260232-265223,
                      340488-345479, and 484488-489479."
                      /db_xref="taxon:9606"
                      /clone="754034"
                      /clone_lib="Soares NhHMPu S1"
                      /tissue_type="Pooled human melanocyte, fetal heart, and
                      pregnant uterus"
                      /lab_host="DH10B"
     mRNA             complement(<1. .>404)
                      /db_xref="GDB:5976908"
BASE COUNT      114 a      74 c      94 g     122 t
ORIGIN
```

```
  Query Match          100.0%;  Score 119;  DB 8;  Length 404;
  Best Local Similarity  99.2%;  Pred. No. 2.13e-254;
  Matches   119;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;
```



```
Db      4 TTTGCTTTTTAGGGAGCTTTTCGGCGCTCCCATTTATTCCTTATACAGGCTCAGTTTTGT 63
          |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Cp    120 TTTGCTTTTTANGGAGCTTTTCGGCGCTCCCATTTATTCCTTATACAGGCTCAGTTTTGT 61

Db     64 CTAGTGTATAGGGACAGCATTTCTGAGAGAAAGCTTATTCATTTCTCAAGAGTAAAGATC 123
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     60 CTAGTGTATAGGGACAGCATTTCTGAGAGAAAGCTTATTCATTTCTCAAGAGTAAAGATC 1
```

```
RESULT    2
LOCUS       C21276         117 bp    DNA             EST       01-OCT-1996
DEFINITION  HUMGS0002342, Human Gene Signature, 3'-directed cDNA sequence, mRNA
            sequence.
ACCESSION   C21276
NID         g1622386
KEYWORDS    EST; Gene Signature; GS; EST(expressed sequence tag); BodyMap; gene
            expression.
SOURCE      One or more human adult tissue.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 117)
  AUTHORS   Okubo,K.
  TITLE     BodyMap; human gene expression database
  JOURNAL   Unpublished (1996)
REFERENCE   2  (bases 1 to 117)
  AUTHORS   Okubo,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-SEP-1996) to the DDBJ/EMBL/GenBank databases. Kousaku
            Okubo, Osaka University, Institute for Molecular and Cellular Bio;
            1-3,Yamada-oka, Suita, Osaka Pref. 565, Japan
            (E-mail:kousaku@imcb.osaka-u.ac.jp, Tel:06-877-5111(ex.3315),
            Fax:06-877-1922)
COMMENT     We are not submitting the same cDNA sequence redundantly to DDBJ
            since 1993. For the abundance information of clones with this
            sequence in this library and as well as in other 3'-directed
            libraries, see ' http://www.imcb.osaka-u.ac.jp/bodymap'.  The
            sequences of the clones represented by this GS sequences is also
            found there.
FEATURES              Location/Qualifiers
     source           1. .117
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
BASE COUNT       42 a      24 c      22 g      28 t       1 others
ORIGIN
```

```
  Query Match          97.5%;  Score 116;  DB 6;  Length 117;
  Best Local Similarity 100.0%;  Pred. No. 2.17e-246;
  Matches   117;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```



```
Db      1 GATCTTTACTCTTGAGAAATGAATAAGCTTTCTCTCAGAAATGCTGTCCCTATACACTAG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      1 GATCTTTACTCTTGAGAAATGAATAAGCTTTCTCTCAGAAATGCTGTCCCTATACACTAG 60

Db     61 ACAAAACTGAGCCTGTATAAGGAATAAATGGGAGCGCCGAAAAGCTCCNTAAAAAGC 117
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     61 ACAAAACTGAGCCTGTATAAGGAATAAATGGGAGCGCCGAAAAGCTCCNTAAAAAGC 117
```

```
RESULT    3
LOCUS       AA832233       247 bp    mRNA            EST       20-FEB-1998
DEFINITION  oc98b06.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1357715
            similar to contains Alu repetitive element;contains element MER35
            repetitive element ;, mRNA sequence.
ACCESSION   AA832233
NID         g2905332
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 247)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
            Ph.D., Gerald Marti, M.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 193.
FEATURES              Location/Qualifiers
     source           1. .247
                      /organism="Homo sapiens"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                      was prepared from human tonsillar cells enriched for
                      germinal center B cells by flow sorting (CD20+, IgD-),
                      provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
```

(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was primed with a Not I - oligo(dT) primer [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTT-3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."
/db_xref="taxon:9606"
/clone="IMAGE:1357715"
/clone_lib="NCI_CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"

BASE COUNT 55 a 49 c 47 g 96 t
ORIGIN

Query Match 15.1%; Score 18; DB 13; Length 247;
Best Local Similarity 100.0%; Pred. No. 1.90e-06;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 216 GAAATGAATAAGCTTTCT 233
||||||||||||||||||
Qy 16 GAAATGAATAAGCTTTCT 33

RESULT 4
LOCUS AA605560 430 bp mRNA EST 29-SEP-1997
DEFINITION 30596 Lambda-PRL2 Arabidopsis thaliana cDNA clone 205D13XP 3', mRNA sequence.
ACCESSION AA605560
NID g2446088
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryotae; Viridiplantae; Charophyta/Embryophyta group; Embryophyta; vascular plants; seed plants; Magnoliophyta; Magnoliopsida; Capparales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 430)
AUTHORS Newman,T., deBruijn,F.J., Green,P., Keegstra,K., Kende,H., McIntosh,L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomashow,M., Retzel,E. and Somerville,C.
TITLE Genes galore: a summary of methods for accessing results from large-scale partial sequencing of anonymous Arabidopsis cDNA clones
JOURNAL Plant Physiol. 106, 1241-1255 (1994)
MEDLINE 95148729
COMMENT
Contact: Thomas Newman
MSU-DOE Plant Research Laboratory
Michigan State University
MSU-DOE-PRL, Michigan State University,Plant Biology Bldg.,E. Lansing,Mi

Tel: 517-353-0854
Fax: 517-353-9168
Email: 22313tcn@ibm.cl.msu.edu
The sequence entry for this EST has been reverse complimented and is being submitted in the sense orientation.
Seq primer: M13 -21 dye primer.
FEATURES Location/Qualifiers
source 1. .430
/organism="Arabidopsis thaliana"
/strain="var columbia"
/note="Vector: lambda Zip-Lox; Site_1: Sal; Site_2: Not; Lambda PRL2 is a cDNA library derived from equal quantities of 4 pools of mRNA. The mRNA sources were 1) 7 day germinated etiolated seedlings; 2) tissue culture grown roots; 3) staged plants half with 24 hour light cycle, half on 16 hr light, 8 hour dark- rosettes; 4) same plants as 3 but aerial tissue (stems, flowers and siliques. The vector is BRL's lambda Zip-Lox. The cDNA inserts were directionally cloned with Sal-Not arms using oligo dT primed cDNA."
/db_xref="taxon:3702"
/clone="205D13XP"
/clone_lib="Lambda-PRL2"
BASE COUNT 110 a 106 c 90 g 107 t 17 others
ORIGIN

Query Match 15.1%; Score 18; DB 9; Length 430;
Best Local Similarity 100.0%; Pred. No. 1.90e-06;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 285 TGAGAGAAAGCTTATTCA 302
||||||||||||||||||
Cp 37 TGAGAGAAAGCTTATTCA 20

RESULT 5
LOCUS W89200 455 bp mRNA EST 07-MAY-1997
DEFINITION zh69f06.s1 Soares fetal liver spleen 1NFLS S1 Homo sapiens cDNA clone 417347 3' similar to gb:M24902 PROSTATIC ACID PHOSPHATASE PRECURSOR (HUMAN);, mRNA sequence.
ACCESSION W89200
NID g1404512
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 455)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,

Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
TITLE WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT
Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 922 Std Error: 0.00
Seq primer: mob.REGA+ET
High quality sequence stop: 321.
FEATURES Location/Qualifiers
source 1. .455
/organism="Homo sapiens"
/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site_1: Pac I; Site_2: Eco RI; This is a subtracted version of the original Soares fetal liver spleen 1NFLS library. 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."
/db_xref="taxon:9606"
/clone="417347"
/clone_lib="Soares fetal liver spleen 1NFLS S1"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
mRNA complement(<1. .>455)
/db_xref="GDB:1325817"
BASE COUNT 146 a 87 c 87 g 134 t 1 others
ORIGIN

Query Match 15.1%; Score 18; DB 6; Length 455;
Best Local Similarity 100.0%; Pred. No. 1.90e-06;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 166 TCCCATTTATTCCTTATA 183
||||||||||||||||||
Cp 93 TCCCATTTATTCCTTATA 76

RESULT 6
LOCUS AA805195 473 bp mRNA EST 19-FEB-1998
DEFINITION oa89h05.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1319481 similar to contains L1.t3 L1 repetitive element ;, mRNA sequence.
ACCESSION AA805195
NID g2876347
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 473)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman, Ph.D., Gerald Marti, M.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 290 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 441.
FEATURES Location/Qualifiers
source 1. .473
/organism="Homo sapiens"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from human tonsillar cells enriched for germinal center B cells by flow sorting (CD20+, IgD-), provided by Dr. Louis M. Staudt (NCI), Dr. David Allman (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was primed with a Not I - oligo(dT) primer [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTT-3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."
/db_xref="taxon:9606"
/clone="IMAGE:1319481"

```
                     /clone_lib="NCI_CGAP_GCB1"
                     /tissue_type="germinal center B cell"
                     /lab_host="DH10B"
BASE COUNT      108 a      97 c      74 g     194 t
ORIGIN

  Query Match           15.1%;  Score 18;  DB 13;  Length 473;
  Best Local Similarity 100.0%;  Pred. No. 1.90e-06;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     267 GAAATGAATAAGCTTTCT 284
           ||||||||||||||||||
Qy      16 GAAATGAATAAGCTTTCT 33


RESULT    7
LOCUS       ATTS2628      299 bp    RNA             EST       14-MAR-1994
DEFINITION  A. thaliana transcribed sequence; clone VBVSB10; Similar to
            Light-induced protein CPRF-2 ; Petroselinum crispum, mRNA sequence.
ACCESSION   Z30876
NID         g460509
KEYWORDS    EST; expressed sequence tag; partial cDNA sequence.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryotae; mitochondrial eukaryotes; Viridiplantae;
            Charophyta/Embryophyta group; Embryophyta; Magnoliophyta;
            Magnoliopsida; Capparales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 299)
  AUTHORS   Despres,T., Amselem,J., Chiapello,H., Caboche,M. and Hofte,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAR-1994) CNRS, GDR-1003 ACS, INRA, laboratoire de
            Biologie Moleculaire, BP 27, 31326 Castanet-Tolosan cedex, France.
            E-mail:gdr-avp@toulouse.inra.fr. On behalf of: INRA Versailles,
            Laboratoire de Biologie Cellulaire, Route de Saint-Cyr, 78026
            Versailles Cedex, France. E-mail:thierry@versailles.inra.fr
REFERENCE   2  (bases 1 to 299)
  AUTHORS   CNRS.
  TITLE     The Arabidopsis thaliana transcribed genome: the GDR cDNA program
  JOURNAL   Unpublished
COMMENT     Cloning vector: pBluescript (Stratagene);
            Physiological conditions: in vitro-grown etiolated seedlings,
                                      5 days old.
            full automatic.
            similarity detected by BlastX against PIR entry S16321.
FEATURES             Location/Qualifiers
     source          1. .299
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /clone="VBVSB10"
                     /tissue_type="whole seedlings of A. thaliana ecotype
                     Columbia"
```



```
                     /clone_lib="Versailles-VB"
BASE COUNT       75 a      73 c      70 g      80 t       1 others
ORIGIN

  Query Match           14.3%;  Score 17;  DB 20;  Length 299;
  Best Local Similarity 100.0%;  Pred. No. 8.61e-05;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     201 CATTTCTCAAGAGTAAA 217
           |||||||||||||||||
Cp      21 CATTTCTCAAGAGTAAA 5


RESULT    8
LOCUS       AA390332      385 bp    mRNA            EST       14-JAN-1998
DEFINITION  LD03274.5prime LD Drosophila melanogaster embryo BlueScript
            Drosophila melanogaster cDNA clone LD03274 5prime, mRNA sequence.
ACCESSION   AA390332
NID         g2043327
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryotae; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea;
            Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 385)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: httpd://fruitfly.berkeley.edu/EST, est@fruitfly.berkeley.edu
            Plate: 32  row: G  column: 2
            High quality sequence stop: 284.
FEATURES             Location/Qualifiers
     source          1. .385
                     /organism="Drosophila melanogaster"
                     /note="Organ: embryo; Vector: BlueScript SK; Site_1:
                     EcoRI; Site_2: XhoI; Constructed using Stratagene ZAP-cDNA
                     Synthesis kit. Oligo dT-primed and directionally cloned at
                     EcoRI and XhoI in BlueScript SK(+/-)"
                     /db_xref="BDGP_EST:BDcln002725"
                     /db_xref="taxon:7227"
                     /clone="LD03274"
                     /clone_lib="LD Drosophila melanogaster embryo BlueScript"
                     /sex="male and female"
```



```
                     /dev_stage="0 to 24 hours mixed stage embryonic"
                     /lab_host="SOLR"
BASE COUNT      106 a      78 c      74 g     127 t
ORIGIN

  Query Match           14.3%;  Score 17;  DB 12;  Length 385;
  Best Local Similarity 100.0%;  Pred. No. 8.61e-05;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     328 ACACTAGACAAAACTGA 344
           |||||||||||||||||
Qy      54 ACACTAGACAAAACTGA 70


RESULT    9
LOCUS       AQ116088      406 bp    DNA             GSS       28-AUG-1998
DEFINITION  RPCI11-57B18.TK RPCI11 Homo sapiens genomic clone R-57B18, genomic
            survey sequence.
ACCESSION   AQ116088
NID         g3492209
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 406)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
            Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
  TITLE     Use of human BAC End Sequences for Sequence-Ready Map Building
  JOURNAL   Unpublished (1998)
COMMENT
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            For clone availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .406
                     /organism="Homo sapiens"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
                     /db_xref="taxon:9606"
                     /clone="R-57B18"
                     /clone_lib="RPCI11"
```



```
                     /sex="Male"
                     /cell_type="Lymphocytes"
BASE COUNT      126 a      63 c      71 g     146 t
ORIGIN

  Query Match           14.3%;  Score 17;  DB 28;  Length 406;
  Best Local Similarity 100.0%;  Pred. No. 8.61e-05;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     258 AGAAATGAATAAGCTTT 274
           |||||||||||||||||
Qy      15 AGAAATGAATAAGCTTT 31


RESULT   10
LOCUS       AA793171      422 bp    mRNA            EST       09-FEB-1998
DEFINITION  vp21g10.r1 Soares mouse mammary gland NbMMG Mus musculus cDNA clone
            1069314 5', mRNA sequence.
ACCESSION   AA793171
NID         g2856126
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 422)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:591674
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 385.
FEATURES             Location/Qualifiers
     source          1. .422
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /note="Organ: mammary gland; Vector: pT7T3D-Pac
```

```
                     (Pharmacia) with a modified polylinker; Site_1: Not I;
                     Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                     oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATGGTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
                     constructed and normalized by Bento Soares and M.Fatima
                     Bonaldo."
                     /db_xref="taxon:10090"
                     /clone="1069314"
                     /clone_lib="Soares mouse mammary gland NbMMG"
                     /sex="male"
                     /tissue_type="mammary gland"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
BASE COUNT      142 a     67 c     70 g    143 t
ORIGIN

  Query Match           14.3%;  Score 17;  DB 12;  Length 422;
  Best Local Similarity 100.0%;  Pred. No. 8.61e-05;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    268 ACTCTTGAGAAATGAAT 284
          |||||||||||||||||
Qy      8 ACTCTTGAGAAATGAAT 24


RESULT   11
LOCUS       AA288616      426 bp    mRNA            EST       14-APR-1997
DEFINITION  vb23d10.r1 Soares mouse 3NbMS Mus musculus cDNA clone 749779 5',
            mRNA sequence.
ACCESSION   AA288616
NID         g1937667
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae;
            Murinae; Mus.
REFERENCE   1  (bases 1 to 426)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
```



```
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:458763
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 406.
FEATURES              Location/Qualifiers
     source           1. .426
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                      was primed with a Not I - oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCGCTGTTTTTTTTTTTTTTTTTTTTT
                      3']; double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Not I and cloned into the Not I
                      and Eco RI sites of the modified pT7T3 vector. RNA
                      provided by Dr. Bertrand Jordan. Library went through
                      three rounds of normalization, and was constructed by
                      Bento Soares and M.Fatima Bonaldo."
                      /db_xref="taxon:10090"
                      /clone="749779"
                      /clone_lib="Soares mouse 3NbMS"
                      /sex="male"
                      /tissue_type="Spleen"
                      /dev_stage="4 weeks"
                      /lab_host="DH10B"
     mRNA             <1. .>426
BASE COUNT      142 a     71 c     69 g    144 t
ORIGIN

  Query Match           14.3%;  Score 17;  DB 24;  Length 426;
  Best Local Similarity 100.0%;  Pred. No. 8.61e-05;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    102 ACTCTTGAGAAATGAAT 118
          |||||||||||||||||
Qy      8 ACTCTTGAGAAATGAAT 24


RESULT   12
LOCUS       AQ252689      523 bp    DNA             GSS       08-OCT-1998
DEFINITION  HS_3202_B1_E08_T7 CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3202 Col=15 Row=J, genomic survey
            sequence.
ACCESSION   AQ252689
```



```
NID         g3723603
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 523)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Furlong,J.,
            Shaker,R., Schmidt,S., Traicoff,R. and Hood,L.E.
  TITLE     Construction of a Characterized Clone Resource for Genomic
            Sequencing
  JOURNAL   Unpublished (1998)
COMMENT
            Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3202  row: J  column: 15
            Class: BAC ends
            High quality sequence stop: 523.
FEATURES              Location/Qualifiers
     source           1. .523
                      /organism="Homo sapiens"
                      /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                      E-Coli DH10B"
                      /db_xref="taxon:9606"
                      /clone="Plate=3202 Col=15 Row=J"
                      /clone_lib="CIT Approved Human Genomic Sperm Library D"
                      /sex="male"
BASE COUNT      142 a    100 c    105 g    172 t      4 others
ORIGIN

  Query Match           14.3%;  Score 17;  DB 29;  Length 523;
  Best Local Similarity 100.0%;  Pred. No. 8.61e-05;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    256 AGAAATGAATAAGCTTT 272
          |||||||||||||||||
Qy     15 AGAAATGAATAAGCTTT 31


RESULT   13
LOCUS       AA978637      541 bp    mRNA            EST       26-MAY-1998
DEFINITION  LD32174.5prime LD Drosophila melanogaster embryo pOT2 Drosophila
            melanogaster cDNA clone LD32174 5prime, mRNA sequence.
ACCESSION   AA978637
NID         g3156053
```



```
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea;
            Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 541)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: httpd://fruitfly.berkeley.edu/EST, est@fruitfly.berkeley.edu
            Plate: 321  row: G  column: 2
            High quality sequence stop: 381.
FEATURES              Location/Qualifiers
     source           1. .541
                      /organism="Drosophila melanogaster"
                      /note="Organ: embryo; Vector: pOT2; Site_1: EcoRI; Site_2:
                      XhoI; Sized fractionated cDNAs were directly ligated into
                      pOT2."
                      /db_xref="taxon:7227"
                      /clone="LD32174"
                      /clone_lib="LD Drosophila melanogaster embryo pOT2"
                      /sex="male and female"
                      /dev_stage="0 to 24 hours mixed stage embryonic"
                      /lab_host="XL1 Blue"
BASE COUNT      186 a    104 c    122 g    129 t
ORIGIN

  Query Match           14.3%;  Score 17;  DB 14;  Length 541;
  Best Local Similarity 100.0%;  Pred. No. 8.61e-05;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    369 ACACTAGACAAAACTGA 385
          |||||||||||||||||
Qy     54 ACACTAGACAAAACTGA 70


RESULT   14
LOCUS       AA270045      546 bp    mRNA            EST       26-MAR-1997
DEFINITION  va57f08.r1 Soares mouse 3NME12 5 Mus musculus cDNA clone 735495 5',
            mRNA sequence.
ACCESSION   AA270045
NID         g1908383
KEYWORDS    EST.
```

```
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Mus.
REFERENCE   1  (bases 1 to 546)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:452543
            Seq primer: -28ml3 rev2 ET from Amersham
            High quality sequence stop: 491.
FEATURES              Location/Qualifiers
     source           1. .546
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                      was primed with a Not I - oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCCTTATTTTTTTTTTTTTTTTT 3'],
                      on total  mouse RNA (provided by Minoru Ko, Wayne State
                      Univ.); double-stranded cDNA was ligated to Eco RI
                      adaptors (Pharmacia), digested with Not I and cloned into
                      the Not I and Eco RI sites of the modified pT7T3 vector.
                      Library went through one round of normalization, and was
                      constructed by Bento Soares and M. Fatima Bonaldo."
                      /db_xref="taxon:10090"
                      /clone="735495"
                      /clone_lib="Soares mouse 3NME12 5"
                      /sex="unknown"
                      /dev_stage="12.5dpc total fetus"
                      /lab_host="DH10B"
     mRNA             <1. .>546
BASE COUNT      189 a     89 c     84 g    184 t
ORIGIN

  Query Match           14.3%;  Score 17;  DB 24;  Length 546;
```



```
  Best Local Similarity 100.0%;  Pred. No. 8.61e-05;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    229 ACTCTTGAGAAATGAAT 245
          |||||||||||||||||
Qy      8 ACTCTTGAGAAATGAAT 24


RESULT   15
LOCUS       AA505495      258 bp    mRNA            EST       20-AUG-1997
DEFINITION  nh96c11.s1 NCI_CGAP_Br2 Homo sapiens cDNA clone IMAGE:966356, mRNA
            sequence.
ACCESSION   AA505495
NID         g2241632
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 258)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 1179   Std Error: 0.00
            Seq primer: -40ml3 fwd. ET from Amersham
            High quality sequence stop: 157.
FEATURES              Location/Qualifiers
     source           1. .258
                      /organism="Homo sapiens"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; 1st strand cDNA was prepared from pooled bulk
                      breast tumor tissue, and was then primed with a Not I -
                      oligo(dT) primer. Double-stranded cDNA was ligated to Eco
                      RI adaptors (Pharmacia), digested with Not I and cloned
                      into the Not I and Eco RI sites of the modified pT7T3
                      vector.  This library is the normalized version of
```



```
                      NCI_CGAP_Br1.1. Library was constructed by Bento Soares
                      and M. Fatima Bonaldo."
                      /db_xref="taxon:9606"
                      /clone="IMAGE:966356"
                      /clone_lib="NCI_CGAP_Br2"
                      /sex="female, pooled"
                      /tissue_type="breast"
                      /lab_host="DH10B"
     mRNA             <1. .>258
BASE COUNT       81 a     52 c     76 g     49 t
ORIGIN

  Query Match           13.4%;  Score 16;  DB 8;  Length 258;
  Best Local Similarity 100.0%;  Pred. No. 3.31e-03;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    235 CTCAGAAATGCTGTCC 250
          ||||||||||||||||
Qy     34 CTCAGAAATGCTGTCC 49


RESULT   16
LOCUS       W36278        312 bp    mRNA            EST       15-MAY-1996
DEFINITION  HFBEST-55 Human fetal brain QBoqin2 Homo sapiens cDNA, mRNA
            sequence.
ACCESSION   W36278
NID         g1318020
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 312)
  AUTHORS   Pan,M., Yuan,J., Zhou,Y. and Qiang,B.
  TITLE     Human fetal brain cDNAs. Pan,M. et al
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Boqin Qiang
            Inst. of Basic Medical Sciences
            Peking Union Medical College
            5 Dong Dan San Tiao, Beijing 100005, China
            Seq primer: M13 Forward and Reverse Primer.
FEATURES              Location/Qualifiers
     source           1. .312
                      /organism="Homo sapiens"
                      /note="Vector: pBluescript SK; Total RNA was isolated from
                      8-month and 3-month human fetal cerebrum and cerebellum.
                      Then Differential Display RT-PCR was conducted. The cDNA
                      PCR product of differentially expressional signifance was
                      cloned and sequenced by employing pBluescript SK as
                      vector."
```



```
                      /db_xref="taxon:9606"
                      /clone_lib="Human fetal brain QBoqin2"
                      /tissue_type="cerebrum and cerebellum"
                      /dev_stage="fetal, 8-month and 3-month"
                      /lab_host="E.coli DH5a"
     mRNA             <1. .>312
BASE COUNT       63 a     87 c     61 g    101 t
ORIGIN

  Query Match           13.4%;  Score 16;  DB 20;  Length 312;
  Best Local Similarity 100.0%;  Pred. No. 3.31e-03;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     78 GGACAGCATTTCTGAG 93
          ||||||||||||||||
Cp     49 GGACAGCATTTCTGAG 34


RESULT   17
LOCUS       AQ197563      328 bp    DNA             GSS       15-SEP-1998
DEFINITION  CIT-HSP-2383H11.TR CIT-HSP Homo sapiens genomic clone 2383H11,
            genomic survey sequence.
ACCESSION   AQ197563
NID         g3604925
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 328)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
            Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
            Venter,J.C.
  TITLE     Use of a random human BAC End Sequence Database for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1998)
COMMENT     Other_GSSs: CIT-HSP-2383H11.TF
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13 Reverse
            Class: BAC ends.
FEATURES              Location/Qualifiers
     source           1. .328
```

```
                     /organism="Homo sapiens"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII"
                     /db_xref="taxon:9606"
                     /clone="2383H11"
                     /clone_lib="CIT-HSP"
                     /sex="Male"
                     /cell_type="Sperm"
BASE COUNT       107 a      75 c      54 g      92 t
ORIGIN

  Query Match           13.4%;  Score 16;  DB 28;  Length 328;
  Best Local Similarity 100.0%;  Pred. No. 3.31e-03;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    185 ACACTAGACAAAACTG 200
          ||||||||||||||||
Qy     54 ACACTAGACAAAACTG 69


RESULT   18
LOCUS       AA649647      382 bp    mRNA            EST       13-NOV-1997
DEFINITION  ns46f02.s1 NCI_CGAP_Alv1 Homo sapiens cDNA clone IMAGE:1186683,
            mRNA sequence.
ACCESSION   AA649647
NID         g2576975
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 382)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Lee Helman, M.D., Michael R. Emmert-Buck, M.D.,
            Ph.D.
             cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 628   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham.
```



```
FEATURES             Location/Qualifiers
     source          1. .382
                     /organism="Homo sapiens"
                     /note="Vector: pAMP10; mRNA made from alveolar
                     rhabdomyosarcoma. cDNA made by oligo-dT priming.
                     Non-directionally cloned. Size-selected on agarose gel,
                     average insert  size 600 bp. Reference:  Krizman et al.
                     (1996) Cancer Research 56:5380-5383."
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1186683"
                     /clone_lib="NCI_CGAP_Alv1"
                     /tissue_type="alveolar rhabdomyosarcoma"
                     /lab_host="DH10B"
BASE COUNT       84 a      100 c      72 g      126 t
ORIGIN

  Query Match           13.4%;  Score 16;  DB 10;  Length 382;
  Best Local Similarity 100.0%;  Pred. No. 3.31e-03;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    193 GGACAGCATTTCTGAG 208
          ||||||||||||||||
Cp     49 GGACAGCATTTCTGAG 34


RESULT   19
LOCUS       N49999        388 bp    mRNA            EST       14-FEB-1996
DEFINITION  yv24c06.r1 Homo sapiens cDNA clone 243658 5' similar to SW:CL4_RAT
            Q09167 INSULIN-INDUCED GROWTH RESPONSE PROTEIN CL-4 ;.
ACCESSION   N49999
NID         g1191165
KEYWORDS    EST.
SOURCE      human clone=243658 primer=T7 library=Soares fetal liver spleen
            1NFLS vector=pT7T3D (Pharmacia) with a modified polylinker
            host=DH10B (ampicillin resistant) Rsite1=Pac I Rsite2=Eco RI Liver
            and spleen from a 20 week-post conception male fetus. 1st strand
            cDNA was primed with a Pac I - oligo(dT) primer [5'
            AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTT 3'], double-stranded
            cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac
            I and cloned into the Pac I and Eco RI sites of the modified pT7T3
            vector.  Library went through one round of normalization. Library
            constructed by Bento Soares and M.Fatima Bonaldo.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 388)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
```



```
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 270
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES             Location/Qualifiers
     source          1. .388
                     /organism="Homo sapiens"
                     /clone="243658"
     mRNA            <1. .>388
BASE COUNT      110 a      88 c     103 g      83 t      4 others
ORIGIN

  Query Match           13.4%;  Score 16;  DB 19;  Length 388;
  Best Local Similarity 100.0%;  Pred. No. 3.31e-03;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     42 AAGGAATAAATGGGAG 57
          ||||||||||||||||
Qy     79 AAGGAATAAATGGGAG 94


RESULT   20
LOCUS       AI015026      390 bp    mRNA            EST       27-AUG-1998
DEFINITION  ot30e06.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1618306
            3', mRNA sequence.
ACCESSION   AI015026
NID         g3229362
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 390)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
```



```
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 822   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 376.
FEATURES             Location/Qualifiers
     source          1. .390
                     /organism="Homo sapiens"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from mRNA obtained from Clontech
                     Laboratories, Inc., and primed with  a Not I - oligo(dT)
                     primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3']
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization to Cot5, and was
                     constructed  by Bento Soares and M. Fatima Bonaldo."
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1618306"
                     /clone_lib="Soares_testis_NHT"
                     /sex="male"
                     /lab_host="DH10B"
BASE COUNT      121 a      79 c     100 g      90 t
ORIGIN

  Query Match           13.4%;  Score 16;  DB 17;  Length 390;
  Best Local Similarity 100.0%;  Pred. No. 3.31e-03;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    236 CTCAGAAATGCTGTCC 251
          ||||||||||||||||
Qy     34 CTCAGAAATGCTGTCC 49


RESULT   21
LOCUS       AU015448      390 bp    mRNA            EST       30-JUL-1998
DEFINITION  Mus musculus 2-cell embryo cDNA 3'-end sequence, clone J0712B03,
            mRNA sequence.
ACCESSION   AU015448
NID         g3370239
KEYWORDS    EST; EST(expressed sequence tag).
```

```
SOURCE      Mus musculus (strain:C57BL/6J) 2-cell stage embryo cDNA to mRNA,
            clone:J0712B03.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 390)
  AUTHORS   Doi,H. and Ko,M.S.
  TITLE     Systematic analyses of genes expressed in 2-cell stage mouse
            embryos (The ERATO/Doi Project at Wayne State University)
  JOURNAL   Published Only in DataBase (1998) In press
REFERENCE   2  (bases 1 to 390)
  AUTHORS   Doi,H. and Ko,M.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUN-1998) to the DDBJ/EMBL/GenBank databases.
            Hirofumi Doi, Japan Science and Technology Corporation (JST), Doi
            Bioasymmetry Project, ERATO; WBG Marive East 12F, 2-6 Nakase,
            Mihama-ku, Chiba, Chiba 261-7112, Japan (E-mail:hd@bioa.jst.go.jp,
            Tel:+81-43-299-1351, Fax:+81-43-297-7530)
FEATURES              Location/Qualifiers
     source           1. .390
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="J0712B03"
                      /dev_stage="2-cell stage embryo"
BASE COUNT      113 a    102 c     71 g    103 t      1 others
ORIGIN

  Query Match           13.4%;  Score 16;  DB 17;  Length 390;
  Best Local Similarity  94.1%;  Pred. No. 3.31e-03;
  Matches   16;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db    277 GAAAAGCTCCTTAAAAA 293
          ||||||||||  ||||||
Qy     99 GAAAAGCTCCNTAAAAA 115


RESULT   22
LOCUS       AA457655      396 bp    mRNA            EST       06-JUN-1997
DEFINITION  aa91h02.s1 Stratagene fetal retina 937202 Homo sapiens cDNA clone
            838707 3' similar to contains Alu repetitive element;contains
            element MSR1 repetitive element ;, mRNA sequence.
ACCESSION   AA457655
NID         g2180375
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 396)
```



```
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J.,
            Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-Merck EST Project 1997
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -41ml3 fwd. ET from Amersham
            High quality sequence stop: 394.
FEATURES              Location/Qualifiers
     source           1. .396
                      /organism="Homo sapiens"
                      /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                      XhoI; Cloned unidirectionally.  Primer: Oligo dT. Pooled
                      retinal tissue. Average insert size: 1.0 kb; Uni-ZAP XR
                      Vector; -5' adaptor sequence: 5' GAATTCGGCACGAG 3' -3'
                      adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTTT 3'"
                      /db_xref="taxon:9606"
                      /clone="838707"
                      /clone_lib="Stratagene fetal retina 937202"
                      /sex="mixed"
                      /lab_host="SOLR (kanamycin resistant)"
     mRNA             complement(<1. .>396)
BASE COUNT       95 a     88 c    134 g     79 t
ORIGIN

  Query Match           13.4%;  Score 16;  DB 7;  Length 396;
  Best Local Similarity 100.0%;  Pred. No. 3.31e-03;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    380 GCTCAGTTTGTCTAG 395
          ||||||||||||||||
Cp     72 GCTCAGTTTGTCTAG 57


RESULT   23
LOCUS       H65048        402 bp    mRNA            EST       18-OCT-1995
DEFINITION  yr69c09.s1 Homo sapiens cDNA clone 210544 3'.
ACCESSION   H65048
NID         g1023788
KEYWORDS    EST.
SOURCE      human clone=210544 primer=Promega -21ml3 library=Soares fetal liver
```



```
            spleen 1NFLS vector=pT7T3D (Pharmacia) with a modified polylinker
            host=DH10B (ampicillin resistant) Rsitel=Pac I Rsite2=Eco RI Liver
            and spleen from a 20 week-post conception male fetus. 1st strand
            cDNA was primed with a Pac I - oligo(dT) primer [5'
            AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTT 3'], double-stranded
            cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac
            I and cloned into the Pac I and Eco RI sites of the modified pT7T3
            vector.  Library went through one round of normalization. Library
            constructed by Bento Soares and M.Fatima Bonaldo.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 402)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 301
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES              Location/Qualifiers
     source           1. .402
                      /organism="Homo sapiens"
                      /clone="210544"
     mRNA             <1. .>402
BASE COUNT      128 a     75 c    106 g     92 t      1 others
ORIGIN

  Query Match           13.4%;  Score 16;  DB 19;  Length 402;
  Best Local Similarity 100.0%;  Pred. No. 3.31e-03;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    224 CTCAGAAATGCTGTCC 239
          ||||||||||||||||
Qy     34 CTCAGAAATGCTGTCC 49
```



```
RESULT   24
LOCUS       H09691        403 bp    mRNA            EST       23-JUN-1995
DEFINITION  y199d07.r1 Homo sapiens cDNA clone 46538 5'.
ACCESSION   H09691
NID         g874513
KEYWORDS    EST.
SOURCE      human clone=46538 library=Soares infant brain 1NIB vector=Lafmid BA
            host=DH10B (ampicillin resistant) primer=M13RP1 Rsite1=Not I
            Rsite2=Hind III Whole brain from a 73 days post natal female. 1st
            strand cDNA was primed with a Not I - oligo(dT) primer [5'
            AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTT 3']; double-stranded
            cDNA was ligated to Hind III adaptors (Pharmacia), digested with
            Not I and directionally cloned into the Not I and Hind III sites of
            the Lafmid BA vector. Library went through one round of
            normalization. Library constructed by Bento Soares and M.Fatima
            Bonaldo.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 403)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            GDB: G00-419-079
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 275
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES              Location/Qualifiers
     source           1. .403
                      /organism="Homo sapiens"
                      /clone="46538"
BASE COUNT       87 a     88 c     94 g    121 t     13 others
ORIGIN

  Query Match           13.4%;  Score 16;  DB 16;  Length 403;
  Best Local Similarity 100.0%;  Pred. No. 3.31e-03;
```

```
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    297 GGACAGCATTTCTGAG 312
          ||||||||||||||||
Cp     49 GGACAGCATTTCTGAG 34


RESULT   25
LOCUS       AQ152046       408 bp    DNA              GSS       07-OCT-1998
DEFINITION  HS_3104_A2_H12_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3104 Col=24 Row=O, genomic survey
            sequence.
ACCESSION   AQ152046
NID         g3541014
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 408)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Furlong,J.,
            Shaker,R., Schmidt,S., Traicoff,R. and Hood,L.E.
  TITLE     Construction of a Characterized Clone Resource for Genomic
            Sequencing
  JOURNAL   Unpublished (1998)
COMMENT
            Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3104  row: O  column: 24
            Class: BAC ends
            High quality sequence stop: 408.
FEATURES             Location/Qualifiers
     source          1. .408
                     /organism="Homo sapiens"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
                     /db_xref="taxon:9606"
                     /clone="Plate=3104 Col=24 Row=O"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
BASE COUNT      133 a     68 c     54 g    153 t
ORIGIN

  Query Match           13.4%;  Score 16;  DB 29;  Length 408;
  Best Local Similarity 100.0%;  Pred. No. 3.31e-03;
```



```
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     40 TATAGGGACAGCATTT 55
          ||||||||||||||||
Cp     54 TATAGGGACAGCATTT 39


RESULT   26
LOCUS       AA598625       412 bp    mRNA             EST       05-MAR-1998
DEFINITION  ae39h02.s1 Gessler Wilms tumor Homo sapiens cDNA clone 898227 3',
            mRNA sequence.
ACCESSION   AA598625
NID         g2432208
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 412)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
            Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
            Theising,B., White,T., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 2648   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 379.
FEATURES             Location/Qualifiers
     source          1. .412
                     /organism="Homo sapiens"
                     /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; RNA
                     was prepared from a pool of 6 anonymous Wilms' tumor RNAs.
                      RNA was prepared by acid-phenol, followed by one round of
                     oligo dT selection.  cDNA library preparation was with
                     the BRL/Life Tech. Superscript Plasmid system.  An
                     oligo-dT NotI primer for first strand synthesis generated
                     gcggccgccc(t)n at the 3' end of the clones.  A 5' SalI
                     adaptor was used  with sequence 5'-gtcgacccacgcgtccg-3'.
                     Resulting cDNAs were size selected (average size 2 kb),
                     NotI digested, and ligated into NotI/SalI-cut pSPORT1.
                     Library was constructed by Dr. Manfred Gessler."
```



```
                     /db_xref="taxon:9606"
                     /clone="898227"
                     /clone_lib="Gessler Wilms tumor"
                     /sex="pooled (6)"
                     /lab_host="DH10B"
BASE COUNT      130 a     86 c    106 g     90 t
ORIGIN

  Query Match           13.4%;  Score 16;  DB 13;  Length 412;
  Best Local Similarity 100.0%;  Pred. No. 3.31e-03;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    228 CTCAGAAATGCTGTCC 243
          ||||||||||||||||
Qy     34 CTCAGAAATGCTGTCC 49


RESULT   27
LOCUS       H91292         414 bp    mRNA             EST       28-NOV-1995
DEFINITION  yu99h08.s1 Homo sapiens cDNA clone 241407 3'.
ACCESSION   H91292
NID         g1081722
KEYWORDS    EST.
SOURCE      human clone=241407 primer=Promega -21m13 library=Soares fetal liver
            spleen 1NFLS vector=pT7T3D (Pharmacia) with a modified polylinker
            host=DH10B (ampicillin resistant) Rsite1=Pac I Rsite2=Eco RI Liver
            and spleen from a 20 week-post conception male fetus. 1st strand
            cDNA was primed with a Pac I - oligo(dT) primer [5'
            AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTT 3'], double-stranded
            cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac
            I and cloned into the Pac I and Eco RI sites of the modified pT7T3
            vector.  Library went through one round of normalization. Library
            constructed by Bento Soares and M.Fatima Bonaldo.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 414)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
```



```
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 348
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES             Location/Qualifiers
     source          1. .414
                     /organism="Homo sapiens"
                     /clone="241407"
     mRNA            <1. .>414
BASE COUNT      131 a     79 c    113 g     90 t      1 others
ORIGIN

  Query Match           13.4%;  Score 16;  DB 19;  Length 414;
  Best Local Similarity 100.0%;  Pred. No. 3.31e-03;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    231 CTCAGAAATGCTGTCC 246
          ||||||||||||||||
Qy     34 CTCAGAAATGCTGTCC 49


RESULT   28
LOCUS       T44484         418 bp    mRNA             EST       07-JAN-1998
DEFINITION  7747 Lambda-PRL2 Arabidopsis thaliana cDNA clone 125K24T7, mRNA
            sequence.
ACCESSION   T44484
NID         g2759281
KEYWORDS    EST.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryotae; Viridiplantae; Charophyta/Embryophyta group;
            Embryophyta; Tracheophyta; seed plants; Magnoliophyta;
            eudicotyledons; Rosidae; Capparales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 418)
  AUTHORS   Newman,T., deBruijn,F.J., Green,P., Keegstra,K., Kende,H.,
            McIntosh,L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomashow,M.,
            Retzel,E. and Somerville,C.
  TITLE     Genes galore: a summary of methods for accessing results from
            large-scale partial sequencing of anonymous Arabidopsis cDNA clones
  JOURNAL   Plant Physiol. 106, 1241-1255 (1994)
  MEDLINE   95148729
COMMENT     On Jan 8, 1998 this sequence version replaced gi:948777.

            Contact: Thomas Newman
            MSU-DOE Plant Research Laboratory
            Michigan State University
            MSU-DOE-PRL, Michigan State University,Plant Biology Bldg.,E.
            Lansing,Mi
```

```
            Tel: 517-353-0854
            Fax: 517-353-9168
            Email: 22313tcn@ibm.cl.msu.edu
            Seq primer: T7.
FEATURES              Location/Qualifiers
     source           1. .418
                      /organism="Arabidopsis thaliana"
                      /strain="var columbia"
                      /note="Vector: lambda Zip-Lox; Site_1: Sal; Site_2: Not;
                      Lambda PRL2 is a cDNA library derived from equal
                      quantities of 4 pools of mRNA.  The mRNA sources were 1) 7
                      day germinated etiolated seedlings; 2) tissue culture
                      grown roots; 3) staged plants half with 24 hour light
                      cycle, half on 16 hr light, 8 hour dark- rosettes;  4)
                      same plants as 3 but aerial tissue (stems, flowers and
                      siliques.  The vector is BRL's lambda Zip-Lox.  The cDNA
                      inserts were directionally cloned with Sal-Not arms using
                      oligo dT primed cDNA."
                      /db_xref="taxon:3702"
                      /clone="125K24T7"
                      /clone_lib="Lambda-PRL2"
BASE COUNT      129 a      81 c      95 g      97 t      16 others
ORIGIN

  Query Match           13.4%;  Score 16;  DB 11;  Length 418;
  Best Local Similarity  94.1%;  Pred. No. 3.31e-03;
  Matches   16;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db    372 CTTTTTAGGGAGCTTTT 388
          ||||||| |||||||||
Cp    116 CTTTTTANGGAGCTTTT 100


RESULT   29
LOCUS       AQ204013      421 bp    DNA             GSS       16-SEP-1998
DEFINITION  HS_3106_B2_D03_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3106 Col=6 Row=H, genomic survey
            sequence.
ACCESSION   AQ204013
NID         g3614583
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 421)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Furlong,J.,
            Shaker,R., Schmidt,S., Traicoff,R. and Hood,L.E.
  TITLE     Construction of a Characterized Clone Resource for Genomic
            Sequencing
  JOURNAL   Unpublished (1998)
```



```
COMMENT
            Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3106  row: H  column: 6
            Class: BAC ends
            High quality sequence stop: 421.
FEATURES              Location/Qualifiers
     source           1. .421
                      /organism="Homo sapiens"
                      /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                      E-Coli DH10B"
                      /db_xref="taxon:9606"
                      /clone="Plate=3106 Col=6 Row=H"
                      /clone_lib="CIT Approved Human Genomic Sperm Library D"
                      /sex="male"
BASE COUNT      110 a      91 c      103 g      117 t
ORIGIN

  Query Match           13.4%;  Score 16;  DB 28;  Length 421;
  Best Local Similarity 100.0%;  Pred. No. 3.31e-03;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    200 GGGACAGCATTTCTGA 215
          ||||||||||||||||
Cp     50 GGGACAGCATTTCTGA 35


RESULT   30
LOCUS       B44253        421 bp    DNA             GSS       20-OCT-1997
DEFINITION  HS-1059-A1-D02-MF.abi CIT Human Genomic Sperm Library C Homo
            sapiens genomic clone Plate=CT 781 Col=3 Row=G, genomic survey
            sequence.
ACCESSION   B44253
NID         g2549087
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 421)
  AUTHORS   Mahairas,G.G., Zackrone,K.D., Smith,T., Tipton,S., Schmidt,S.,
            Traicoff,R., Abajian,C., Blanchard,A., West,A. and Hood,L.E.
  TITLE     Construction of a Characterized Clone Resource for Genomic
            Sequencing: Generation and Preliminary Analysis of 20,000 Sequence
            Tagged Connectors
```



```
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Mahairas GG, Zackrone KD, Hood L
            University of Washington
            Seattle, WA 98195, USA
            Tel: (206) 616-8744
            Fax: (206) 685-7301
            Email: kzackron@u.washington.edu
            Sequence Tagged Connector
            Plate: CT 781  row: G  column: 3
            Class: BAC ends
            High quality sequence stop: 421.
FEATURES              Location/Qualifiers
     source           1. .421
                      /organism="Homo sapiens"
                      /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                      E-Coli DH10B"
                      /db_xref="taxon:9606"
                      /clone="Plate=CT 781 Col=3 Row=G"
                      /clone_lib="CIT Human Genomic Sperm Library C"
                      /sex="M"
BASE COUNT      116 a      75 c      121 g      108 t      1 others
ORIGIN

  Query Match           13.4%;  Score 16;  DB 26;  Length 421;
  Best Local Similarity 100.0%;  Pred. No. 3.31e-03;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    131 GAGAGAAAGCTTATTC 146
          ||||||||||||||||
Cp     36 GAGAGAAAGCTTATTC 21


RESULT   31
LOCUS       R20601        430 bp    mRNA            EST       18-APR-1995
DEFINITION  yg49f06.r1 Homo sapiens cDNA clone 36044 5'.
ACCESSION   R20601
NID         g775382
KEYWORDS    EST.
SOURCE      human clone=36044 library=Soares infant brain 1NIB vector=Lafmid BA
            host=DH10B (ampicillin resistant) primer=M13RP1 RsiteI=Not I
            RsiteJ=Hind III Whole brain from a 73 days post natal female. 1st
            strand cDNA was primed with a Not I - oligo(dT) primer [5'
            AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTT 3']; double-stranded
            cDNA was ligated to Hind III adaptors (Pharmacia), digested with
            Not I and directionally cloned into the Not I and Hind III sites of
            the Lafmid BA vector. Library went through one round of
            normalization. Library constructed by Bento Soares and M.Fatima
            Bonaldo.
  ORGANISM  Homo sapiens
            Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
```



```
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 430)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            GDB: G00-408-545
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 329
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES              Location/Qualifiers
     source           1. .430
                      /organism="Homo sapiens"
                      /clone="36044"
BASE COUNT       89 a      99 c      99 g      137 t      6 others
ORIGIN

  Query Match           13.4%;  Score 16;  DB 5;  Length 430;
  Best Local Similarity 100.0%;  Pred. No. 3.31e-03;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    257 GGACAGCATTTCTGAG 272
          ||||||||||||||||
Cp     49 GGACAGCATTTCTGAG 34


RESULT   32
LOCUS       AQ079945      431 bp    DNA             GSS       19-AUG-1998
DEFINITION  CIT-HSP-2369K11.TR CIT-HSP Homo sapiens genomic clone 2369K11,
            genomic survey sequence.
ACCESSION   AQ079945
NID         g3441129
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 431)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
```

```
            Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
            Venter,J.C.
  TITLE     Use of a random human BAC End Sequence Database for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1998)
COMMENT     Other_GSSs: CIT-HSP-2369K11.TF
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13 Reverse
            Class: BAC ends.
FEATURES            Location/Qualifiers
     source         1. .431
                    /organism="Homo sapiens"
                    /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                    HindIII"
                    /db_xref="taxon:9606"
                    /clone="2369K11"
                    /clone_lib="CIT-HSP"
                    /sex="Male"
                    /cell_type="Sperm"
BASE COUNT      124 a    105 c     91 g    111 t
ORIGIN

  Query Match          13.4%;  Score 16;  DB 27;  Length 431;
  Best Local Similarity 100.0%;  Pred. No. 3.31e-03;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    351 CATTTCTCAAGAGTAA 366
          ||||||||||||||||
Cp     21 CATTTCTCAAGAGTAA 6


RESULT   33
LOCUS       B45942         436 bp    DNA             GSS       20-OCT-1997
DEFINITION  HS-1062-B1-G08-MR.abi CIT Human Genomic Sperm Library C Homo
            sapiens genomic clone Plate=CT 784 Col=15 Row=N, genomic survey
            sequence.
ACCESSION   B45942
NID         g2550776
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
```



```
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 436)
  AUTHORS   Mahairas,G.G., Zackrone,K.D., Smith,T., Tipton,S., Schmidt,S.,
            Traicoff,R., Abajian,C., Blanchard,A., West,A. and Hood,L.E.
  TITLE     Construction of a Characterized Clone Resource for Genomic
            Sequencing: Generation and Preliminary Analysis of 20,000 Sequence
            Tagged Connectors
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Mahairas GG, Zackrone KD, Hood L
            University of Washington
            Seattle, WA 98195, USA
            Tel: (206) 616-8744
            Fax: (206) 685-7301
            Email: kzackron@u.washington.edu
            Sequence Tagged Connector
            Plate: CT 784  row: N  column: 15
            Class: BAC ends
            High quality sequence stop: 436.
FEATURES            Location/Qualifiers
     source         1. .436
                    /organism="Homo sapiens"
                    /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                    E-Coli DH10B"
                    /db_xref="taxon:9606"
                    /clone="Plate=CT 784 Col=15 Row=N"
                    /clone_lib="CIT Human Genomic Sperm Library C"
                    /sex="M"
BASE COUNT      186 a     68 c     69 g    112 t      1 others
ORIGIN

  Query Match          13.4%;  Score 16;  DB 26;  Length 436;
  Best Local Similarity 100.0%;  Pred. No. 3.31e-03;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    217 TTTCTCTCAGAAATGC 232
          ||||||||||||||||
Qy     29 TTTCTCTCAGAAATGC 44


RESULT   34
LOCUS       AA246496       442 bp    mRNA            EST       14-JAN-1998
DEFINITION  LD05301.5prime LD Drosophila melanogaster embryo BlueScript
            Drosophila melanogaster cDNA clone LD05301 5prime, mRNA sequence.
ACCESSION   AA246496
NID         g1877825
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryotae; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea;
```



```
            Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 442)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: httpd://fruitfly.berkeley.edu/EST, est@fruitfly.berkeley.edu
            Plate: 53  row: A  column: 1
            High quality sequence stop: 315.
FEATURES            Location/Qualifiers
     source         1. .442
                    /organism="Drosophila melanogaster"
                    /note="Organ: embryo; Vector: BlueScript SK; Site_1:
                    EcoRI; Site_2: XhoI: Constructed using Stratagene ZAP-cDNA
                    Synthesis kit. Oligo dT-primed and directionally cloned at
                    EcoRI and XhoI in BlueScript SK(+/-)"
                    /db_xref="taxon:7227"
                    /clone="LD05301"
                    /clone_lib="LD Drosophila melanogaster embryo BlueScript"
                    /sex="male and female"
                    /dev_stage="0 to 24 hours mixed stage embryonic"
                    /lab_host="SOLR"
BASE COUNT      151 a     76 c    100 g    115 t
ORIGIN

  Query Match          13.4%;  Score 16;  DB 12;  Length 442;
  Best Local Similarity 100.0%;  Pred. No. 3.31e-03;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    362 ACACTAGACAAAACTG 377
          ||||||||||||||||
Qy     54 ACACTAGACAAAACTG 69


RESULT   35
LOCUS       AA747849       445 bp    mRNA            EST       22-JAN-1998
DEFINITION  nx31e07.s1 NCI_CGAP_GC4 Homo sapiens cDNA clone IMAGE:1257732, mRNA
            sequence.
ACCESSION   AA747849
NID         g2787807
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
```



```
REFERENCE   1  (bases 1 to 445)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 668   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 445.
FEATURES            Location/Qualifiers
     source         1. .445
                    /organism="Homo sapiens"
                    /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                    polylinker; 1st strand cDNA was prepared from 3 pooled
                    germ cell tumors, and was then primed with a Not I -
                    oligo(dT) primer. Double-stranded cDNA was ligated to Eco
                    RI adaptors (Pharmacia), digested with Not I and cloned
                    into the Not I and Eco RI sites of the modified pT7T3
                    vector.  Library is normalized. Library was constructed by
                    Bento Soares and M. Fatima Bonaldo."
                    /db_xref="taxon:9606"
                    /clone="IMAGE:1257732"
                    /clone_lib="NCI_CGAP_GC4"
                    /tissue_type="pooled germ cell tumors"
                    /lab_host="DH10B"
BASE COUNT      143 a     84 c    124 g     94 t
ORIGIN

  Query Match          13.4%;  Score 16;  DB 12;  Length 445;
  Best Local Similarity 100.0%;  Pred. No. 3.31e-03;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    237 CTCAGAAATGCTGTCC 252
          ||||||||||||||||
Qy     34 CTCAGAAATGCTGTCC 49


RESULT   36
LOCUS       AA055645       448 bp    mRNA            EST       17-SEP-1996
```

```
DEFINITION  z175c08.s1 Stratagene colon (#937204) Homo sapiens cDNA clone
            510446 3', mRNA sequence.
ACCESSION   AA055645
NID         g1547984
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 448)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
            Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40M13 fwd. from Amersham
            High quality sequence stop: 418.
FEATURES              Location/Qualifiers
     source           1. .448
                      /organism="Homo sapiens"
                      /note="Organ: colon; Vector: pBluescript SK-; Site_1:
                      EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                      Oligo dT. T-84 colonic epithelial cell line.  Average
                      insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor
                      sequence: 5' GAATTCGGCACGAG 3' -3' adaptor sequence: 5'
                      CTCGAGTTTTTTTTTTTTTTT 3'"
                      /db_xref="taxon:9606"
                      /clone="510446"
                      /clone_lib="Stratagene colon (#937204)"
                      /lab_host="SOLR cells (kanamycin resistant)"
     mRNA             complement(<1. .>448)
BASE COUNT      163 a      84 c      67 g     132 t       2 others
ORIGIN

  Query Match           13.4%;  Score 16;  DB 21;  Length 448;
  Best Local Similarity 100.0%;  Pred. No. 3.31e-03;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    423 ACACTAGACAAAACTG 438
          ||||||||||||||||
```



```
Qy     54 ACACTAGACAAAACTG 69


RESULT   37
LOCUS       T50662        455 bp    mRNA            EST       08-FEB-1995
DEFINITION  yb31b02.s1 Homo sapiens cDNA clone 72747 3'.
ACCESSION   T50662
NID         g652522
KEYWORDS    EST.
SOURCE      human clone=72747 library=Stratagene fetal spleen (#937205)
            vector=pBluescript SK- host=SOLR cells (kanamycin resistant)
            primer=-21m13 Rsite1=EcoRI Rsite2=XhoI Pooled fetal spleens. Cloned
            unidirectionally.  Primer: Oligo dT. Average insert size: 1.0 kb;
            Uni-ZAP XR Vector; 5' adaptor sequence: 5'-GAATTCGGCACGAG-3'; 3'
            adaptor sequence: 5'-CTCGAGTTTTTTTTTTTTTTTT-3'.
  ORGANISM  Homo sapiens
            Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 455)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
            Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High qality sequence stops: 309
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES              Location/Qualifiers
     source           1. .455
                      /organism="Homo sapiens"
                      /clone="72747"
BASE COUNT      134 a      86 c     128 g     105 t       2 others
ORIGIN

  Query Match           13.4%;  Score 16;  DB 5;  Length 455;
  Best Local Similarity 100.0%;  Pred. No. 3.31e-03;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    236 CTCAGAAATGCTGTCC 251
          ||||||||||||||||
Qy     34 CTCAGAAATGCTGTCC 49
```



```
RESULT   38
LOCUS       N33238        456 bp    mRNA            EST       10-JAN-1996
DEFINITION  yy07e06.s1 Homo sapiens cDNA clone 270562 3'.
ACCESSION   N33238
NID         g1153637
KEYWORDS    EST.
SOURCE      human clone=270562 primer=m13 -40 forward library=Soares melanocyte
            2NbHM vector=pT7T3D (Pharmacia) with a modified polylinker
            host=DH10B (ampicillin resistant) Rsite1=Not I Rsite2=Eco RI Male.
            1st strand cDNA was primed with a Not I - oligo(dT) primer
            [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCAGTTTTTTTTTTTTTTTTT-3'],
            double-stranded cDNA was size selected, ligated to Eco RI adapters
            (Pharmacia), digested with Not I and cloned into the Not I and Eco
            RI sites of a modified pT7T3 vector (Pharmacia). Library
            constructed by Bento Soares and M.Fatima Bonaldo. RNA from normal
            foreskin melanocytes (FS374) was kindly provided by Dr. Anthony P.
            Albino.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 456)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 373
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES              Location/Qualifiers
     source           1. .456
                      /organism="Homo sapiens"
                      /clone="270562"
     mRNA             <1. .>456
BASE COUNT      145 a      89 c     124 g      98 t
ORIGIN
```



```
  Query Match           13.4%;  Score 16;  DB 19;  Length 456;
  Best Local Similarity 100.0%;  Pred. No. 3.31e-03;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    236 CTCAGAAATGCTGTCC 251
          ||||||||||||||||
Qy     34 CTCAGAAATGCTGTCC 49


RESULT   39
LOCUS       B55083        456 bp    DNA             GSS       19-JUN-1998
DEFINITION  CIT-HSP-386A19.TF CIT-HSP Homo sapiens genomic clone 386A19,
            genomic survey sequence.
ACCESSION   B55083
NID         g2609417
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 456)
  AUTHORS   Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K.,
            Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
            Simon,M. and Venter,J.C.
  TITLE     Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: CIT-HSP-386A19.TR
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: M13-21
            Class: BAC ends.
FEATURES              Location/Qualifiers
     source           1. .456
                      /organism="Homo sapiens"
                      /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                      HindIII"
                      /db_xref="taxon:9606"
                      /clone="386A19"
                      /clone_lib="CIT-HSP"
                      /sex="Male"
                      /cell_type="Sperm"
```

```
BASE COUNT      100 a    116 c     89 g    150 t      1 others
ORIGIN

  Query Match           13.4%;  Score 16;  DB 26;  Length 456;
  Best Local Similarity 100.0%;  Pred. No. 3.3le-03;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    247 TTATTCATTTCTCAAG 262
          ||||||||||||||||
Cp     26 TTATTCATTTCTCAAG 11


RESULT   40
LOCUS       N26980        459 bp    mRNA            EST       29-DEC-1995
DEFINITION  yx18h06.r1 Homo sapiens cDNA clone 262139 5'.
ACCESSION   N26980
NID         g1141328
KEYWORDS    EST.
SOURCE      human clone=262139 primer=T7 library=Soares melanocyte 2NbHM
            vector=pT7T3D (Pharmacia) with a modified polylinker host=DH10B
            (ampicillin resistant) Rsite1=Not I Rsite2=Eco RI Male. 1st strand
            cDNA was primed with a Not I - oligo(dT) primer
            [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCAGTTTTTTTTTTTTTTTTTT-3'],
            double-stranded cDNA was size selected, ligated to Eco RI adapters
            (Pharmacia), digested with Not I and cloned into the Not I and Eco
            RI sites of a modified pT7T3 vector (Pharmacia). Library
            constructed by Bento Soares and M.Fatima Bonaldo. RNA from normal
            foreskin melanocytes (FS374) was kindly provided by Dr. Anthony P.
            Albino.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 459)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 393
```



```
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES              Location/Qualifiers
     source           1. .459
                      /organism="Homo sapiens"
                      /clone="262139"
     mRNA             <1. .>459
BASE COUNT      101 a    124 c    106 g    127 t      1 others
ORIGIN

  Query Match           13.4%;  Score 16;  DB 19;  Length 459;
  Best Local Similarity 100.0%;  Pred. No. 3.3le-03;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     27 GGACAGCATTTCTGAG 42
          ||||||||||||||||
Cp     49 GGACAGCATTTCTGAG 34


RESULT   41
LOCUS       R18159        472 bp    mRNA            EST       14-APR-1995
DEFINITION  yf97e11.r1 Homo sapiens cDNA clone 30422 5'.
ACCESSION   R18159
NID         g771769
KEYWORDS    EST.
SOURCE      human clone=30422 library=Soares infant brain 1NIB vector=Lafmid BA
            host=DH10B (ampicillin resistant) primer=M13RP1 Rsite1=Not I
            Rsite2=Hind III Whole brain from a 73 days post natal female. 1st
            strand cDNA was primed with a Not I - oligo(dT) primer [5'
            AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTTT 3']; double-stranded
            cDNA was ligated to Hind III adaptors (Pharmacia), digested with
            Not I and directionally cloned into the Not I and Hind III sites of
            the Lafmid BA vector. Library went through one round of
            normalization. Library constructed by Bento Soares and M.Fatima
            Bonaldo.
  ORGANISM  Homo sapiens
            Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 472)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            GDB: G00-402-769
            Contact: Wilson RK
            WashU-Merck EST Project
```



```
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 351
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES              Location/Qualifiers
     source           1. .472
                      /organism="Homo sapiens"
                      /clone="30422"
BASE COUNT       99 a    105 c    116 g    150 t      2 others
ORIGIN

  Query Match           13.4%;  Score 16;  DB 5;  Length 472;
  Best Local Similarity 100.0%;  Pred. No. 3.3le-03;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    287 GGACAGCATTTCTGAG 302
          ||||||||||||||||
Cp     49 GGACAGCATTTCTGAG 34


RESULT   42
LOCUS       N20136        475 bp    mRNA            EST       15-DEC-1995
DEFINITION  yx29f06.s1 Homo sapiens cDNA clone 263171 3'.
ACCESSION   N20136
NID         g1124803
KEYWORDS    EST.
SOURCE      human clone=263171 primer=ml3 -40 forward library=Soares melanocyte
            2NbHM vector=pT7T3D (Pharmacia) with a modified polylinker
            host=DH10B (ampicillin resistant) Rsite1=Not I Rsite2=Eco RI Male.
            1st strand cDNA was primed with a Not I - oligo(dT) primer
            [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCAGTTTTTTTTTTTTTTTTT-3'],
            double-stranded cDNA was size selected, ligated to Eco RI adapters
            (Pharmacia), digested with Not I and cloned into the Not I and Eco
            RI sites of a modified pT7T3 vector (Pharmacia). Library
            constructed by Bento Soares and M.Fatima Bonaldo. RNA from normal
            foreskin melanocytes (FS374) was kindly provided by Dr. Anthony P.
            Albino.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 475)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
```



```
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 313
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES              Location/Qualifiers
     source           1. .475
                      /organism="Homo sapiens"
                      /clone="263171"
     mRNA             <1. .>475
BASE COUNT      145 a     92 c    130 g    105 t      3 others
ORIGIN

  Query Match           13.4%;  Score 16;  DB 19;  Length 475;
  Best Local Similarity 100.0%;  Pred. No. 3.3le-03;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    237 CTCAGAAATGCTGTCC 252
          ||||||||||||||||
Qy     34 CTCAGAAATGCTGTCC 49


RESULT   43
LOCUS       H08237        480 bp    mRNA            EST       23-JUN-1995
DEFINITION  y192h07.s1 Homo sapiens cDNA clone 45726 3'.
ACCESSION   H08237
NID         g873059
KEYWORDS    EST.
SOURCE      human clone=45726 library=Soares infant brain 1NIB vector=Lafmid BA
            host=DH10B (ampicillin resistant) primer=Promega -21ml3 Rsite1=Not
            I Rsite2=Hind III Whole brain from a 73 days post natal female. 1st
            strand cDNA was primed with a Not I - oligo(dT) primer [5'
            AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTTT 3']; double-stranded
            cDNA was ligated to Hind III adaptors (Pharmacia), digested with
            Not I and directionally cloned into the Not I and Hind III sites of
            the Lafmid BA vector. Library went through one round of
            normalization. Library constructed by Bento Soares and M.Fatima
            Bonaldo.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
```

```
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 480)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            GDB: G00-418-267
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 313
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES              Location/Qualifiers
     source           1. .480
                      /organism="Homo sapiens"
                      /clone="45726"
BASE COUNT      135 a      85 c     131 g     121 t       8 others
ORIGIN

  Query Match           13.4%;  Score 16;  DB 16;  Length 480;
  Best Local Similarity 100.0%;  Pred. No. 3.31e-03;
  Matches    16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    251 CTCAGAAATGCTGTCC 266
          ||||||||||||||||
Qy     34 CTCAGAAATGCTGTCC 49


RESULT   44
LOCUS       AA097379      488 bp    mRNA            EST       25-OCT-1996
DEFINITION  mk16g10.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone 493122
            5', mRNA sequence.
ACCESSION   AA097379
NID         g1643071
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
```



```
            Mus.
REFERENCE   1  (bases 1 to 488)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:296570
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 482.
FEATURES              Location/Qualifiers
     source           1. .488
                      /organism="Mus musculus"
                      /note="Vector: pT7T3D (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                      was primed with a Not I - oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCATTTTTTTTTTTTTTTTTT 3'],
                      double-stranded cDNA was size selected, ligated to Eco RI
                      adapters (Pharmacia), digested with Not I and cloned into
                      the Not I and Eco RI sites of a modified pT7T3 vector
                      (Pharmacia). Library went through one round of
                      normalization to a Cot = 5. Library constructed by Bento
                      Soares and M.Fatima Bonaldo. RNA was kindly provided by
                      Dr. Minoru Ko (Wayne State University)."
                      /db_xref="taxon:10090"
                      /clone="493122"
                      /clone_lib="Soares mouse p3NMF19.5"
                      /dev_stage="19.5 dpc total fetus"
                      /lab_host="DH10B (ampicillin resistant)"
     mRNA             <1. .>488
BASE COUNT      136 a      85 c      93 g     174 t
ORIGIN

  Query Match           13.4%;  Score 16;  DB 22;  Length 488;
  Best Local Similarity 100.0%;  Pred. No. 3.31e-03;
  Matches    16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     12 TTTCTCTCAGAAATGC 27
          ||||||||||||||||
```



```
Qy     29 TTTCTCTCAGAAATGC 44


RESULT   45
LOCUS       AI156664      489 bp    mRNA            EST       30-SEP-1998
DEFINITION  ue53e05.r1 Soares mouse mammary gland NMLMG Mus musculus cDNA clone
            1494848 5', mRNA sequence.
ACCESSION   AI156664
NID         g3685133
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 489)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:932452
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 489.
FEATURES              Location/Qualifiers
     source           1. .489
                      /organism="Mus musculus"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; 1st strand cDNA was prepared from mammary
                      gland tissue from a lactating female, and was then primed
                      with a Not I - oligo(dT) primer. Double-stranded  cDNA was
                      ligated to Eco RI adaptors (Pharmacia), digested with Not
                      I  and cloned into the Not I and Eco RI sites of the
                      modified pT7T3 vector.  Library is normalized.  Library
                      was constructed by Bento Soares and M. Fatima Bonaldo. "
                      /db_xref="taxon:10090"
                      /clone="1494848"
                      /clone_lib="Soares mouse mammary gland NMLMG"
                      /sex="female (lactating)"
                      /tissue_type="mammary gland"
```



```
                      /lab_host="DH10B"
BASE COUNT      134 a      89 c      94 g     172 t
ORIGIN

  Query Match           13.4%;  Score 16;  DB 18;  Length 489;
  Best Local Similarity 100.0%;  Pred. No. 3.31e-03;
  Matches    16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     20 TTTCTCTCAGAAATGC 35
          ||||||||||||||||
Qy     29 TTTCTCTCAGAAATGC 44


RESULT   46
LOCUS       H48298        489 bp    mRNA            EST       14-SEP-1995
DEFINITION  yq77b02.r1 Homo sapiens cDNA clone 201771 5' similar to SP:B47112
            B47112 GROWTH RESPONSE PROTEIN HRS, INSULIN-INDUCED - ;.
ACCESSION   H48298
NID         g986685
KEYWORDS    EST.
SOURCE      human clone=201771 primer=M13RP1 library=Soares fetal liver spleen
            1NFLS vector=pT7T3D (Pharmacia) with a modified polylinker
            host=DH10B (ampicillin resistant) Rsite1=Pac I Rsite2=Eco RI Liver
            and spleen from a 20 week-post conception male fetus. 1st strand
            cDNA was primed with a Pac I - oligo(dT) primer [5'
            AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTT 3'], double-stranded
            cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac
            I and cloned into the Pac I and Eco RI sites of the modified pT7T3
            vector.  Library went through one round of normalization. Library
            constructed by Bento Soares and M.Fatima Bonaldo.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 489)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
```

```
            High quality sequence stops: 345
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES              Location/Qualifiers
     source           1. .489
                      /organism="Homo sapiens"
                      /clone="201771"
     mRNA             <1. .>489
BASE COUNT      138 a    102 c    124 g    118 t      7 others
ORIGIN

  Query Match           13.4%;  Score 16;  DB 16;  Length 489;
  Best Local Similarity 100.0%;  Pred. No. 3.31e-03;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    112 AAGGAATAAATGGGAG 127
          ||||||||||||||||
Qy     79 AAGGAATAAATGGGAG 94


RESULT   47
LOCUS       W76347          490 bp    mRNA            EST       16-OCT-1996
DEFINITION  sd63a09.r1 Soares fetal heart NbHH19W Homo sapiens cDNA clone
            345304 5', mRNA sequence.
ACCESSION   W76347
NID         g1386797
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 490)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1533   Std Error: 0.00
```



```
            Seq primer: mob.REGA+ET
            High quality sequence stop: 427.
FEATURES              Location/Qualifiers
     source           1. .490
                      /organism="Homo sapiens"
                      /note="Organ: heart; Vector: pT7T3D (Pharmacia) with a
                      modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                      strand cDNA was primed with a Not I - oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCATCTTTTTTTTTTTTTTTTT 3'),
                      double-stranded cDNA was size selected, ligated to Eco RI
                      adapters (Pharmacia), digested with Not I and cloned into
                      the Not I and Eco RI sites of a modified pT7T3 vector
                      (Pharmacia). Library went through one round of
                      normalization to a Cot = 5. Library constructed by
                      M.Fatima Bonaldo. This library was constructed from the
                      same fetus as the fetal lung library, Soares fetal lung
                      NbHL19W."
                      /db_xref="taxon:9606"
                      /clone="345304"
                      /clone_lib="Soares fetal heart NbHH19W"
                      /sex="unknown"
                      /dev_stage="19 weeks"
                      /lab_host="DH10B (ampicillin resistant)"
     mRNA             <1. .>490
BASE COUNT      105 a    125 c    110 g    150 t
ORIGIN

  Query Match           13.4%;  Score 16;  DB 21;  Length 490;
  Best Local Similarity 100.0%;  Pred. No. 3.31e-03;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    123 GGACAGCATTTCTGAG 138
          ||||||||||||||||
Cp     49 GGACAGCATTTCTGAG 34


RESULT   48
LOCUS       AA630644        493 bp    mRNA            EST       06-MAR-1998
DEFINITION  ac11h02.s1 Stratagene HeLa cell s3 937216 Homo sapiens cDNA clone
            856179 3', mRNA sequence.
ACCESSION   AA630644
NID         g2553255
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 493)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
            Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
```



```
            Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1160   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 471.
FEATURES              Location/Qualifiers
     source           1. .493
                      /organism="Homo sapiens"
                      /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                      XhoI; Cloned unidirectionally.  Primer: Oligo dT. HeLa S3
                      epithelioid carcinoma cells grown to semi-confluency
                      without induction. Average insert size: 1.5 kb; Uni-ZAP XR
                      Vector. -5' adaptor sequence: 5' GAATTCGGCACGAG 3' -3'
                      adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTTT 3'"
                      /db_xref="taxon:9606"
                      /clone="856179"
                      /clone_lib="Stratagene HeLa cell s3 937216"
                      /sex="female"
                      /dev_stage="HeLa S3 cell line"
                      /lab_host="SOLR (kanamycin resistant)"
BASE COUNT      159 a    102 c    129 g    103 t
ORIGIN

  Query Match           13.4%;  Score 16;  DB 13;  Length 493;
  Best Local Similarity 100.0%;  Pred. No. 3.31e-03;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    234 CTCAGAAATGCTGTCC 249
          ||||||||||||||||
Qy     34 CTCAGAAATGCTGTCC 49


RESULT   49
LOCUS       AA868504        507 bp    mRNA            EST       16-MAR-1998
DEFINITION  ax43b03.s1 Soares testis NHT Homo sapiens cDNA clone IMAGE:1408685
            3', mRNA sequence.
ACCESSION   AA868504
NID         g2963949
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
```



```
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 507)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 470.
FEATURES              Location/Qualifiers
     source           1. .507
                      /organism="Homo sapiens"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                      was prepared from mRNA obtained from Clontech
                      Laboratories, Inc., and primed with  a Not I - oligo(dT)
                      primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTTT 3']. 
                      Double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Not I and cloned into the Not I
                      and Eco RI sites of the modified pT7T3 vector. Library
                      went through one round of normalization to Cot5, and was
                      constructed  by Bento Soares and M. Fatima Bonaldo."
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1408685"
                      /clone_lib="Soares testis NHT"
                      /sex="male"
                      /lab_host="DH10B"
BASE COUNT      161 a    106 c    131 g    109 t
ORIGIN

  Query Match           13.4%;  Score 16;  DB 13;  Length 507;
  Best Local Similarity 100.0%;  Pred. No. 3.31e-03;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    236 CTCAGAAATGCTGTCC 251
          ||||||||||||||||
Qy     34 CTCAGAAATGCTGTCC 49
```

```
RESULT   50
LOCUS       AQ143721       543 bp    DNA             GSS       23-SEP-1998
DEFINITION  HS_3075_B1_A03_MF CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3075 Col=5 Row=B, genomic survey
            sequence.
ACCESSION   AQ143721
NID         g3534374
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 543)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Furlong,J.,
            Shaker,R., Schmidt,S., Traicoff,R. and Hood,L.E.
  TITLE     Construction of a Characterized Clone Resource for Genomic
            Sequencing
  JOURNAL   Unpublished (1998)
COMMENT
            Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3075  row: B  column: 5
            Class: BAC ends
            High quality sequence stop: 543.
FEATURES             Location/Qualifiers
     source          1. .543
                     /organism="Homo sapiens"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
                     /db_xref="taxon:9606"
                     /clone="Plate=3075 Col=5 Row=B"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
BASE COUNT      187 a     98 c     81 g    166 t     11 others
ORIGIN

  Query Match           13.4%;  Score 16;  DB 29;  Length 543;
  Best Local Similarity 100.0%;  Pred. No. 3.31e-03;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    309 CCTTATACAGGCTCAG 324
          ||||||||||||||||
Cp     82 CCTTATACAGGCTCAG 67
```



```
Search completed: Tue Dec 22 21:35:16 1998
Job time : 314 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Tue Dec 22 21:21:48 1998; MasPar time 210.77 Seconds
1254.714 Million cell updates/sec

Tabular output not generated.

Title: >US-08-530-112A-2058
Description: (1-120) from US08530112A.seq
Perfect Score: 119
N.A. Sequence: 1 GATCTTTACTCTTGAGAAAT.....AAAGCTCCNTAAAAAGCAAA 120
Comp: CTAGAAATGAGAACTCTTTA.....TTTCGAGGNATTTTTCGTTT

Scoring table: TABLE jmetric
Gap 60

Rmatch STD : Dbase 0; Query 0

Searched: 567134 seqs, 1101898692 bases x 2

Post-processing: Minimum Match 0%
Listing first 1000 summaries

Database: embl55
1:em_ba 2:em_fun 3:em_htg 4:em_hum1 5:em_hum2 6:em_in
7:em_om 8:em_or 9:em_ov 10:em_pat 11:em_ph 12:em_pl
13:em_ro 14:em_vi
Database: genbank107
15:gb_bal 16:gb_ba2 17:gb_htg 18:gb_in 19:gb_om 20:gb_ov
21:gb_pat 22:gb_ph 23:gb_pl1 24:gb_pl2 25:gb_pr1
26:gb_pr2 27:gb_pr3 28:gb_ro 29:gb_st 30:gb_sts 31:gb_sy
32:gb_un 33:gb_vi

Statistics: Mean 7.476; Variance 2.548; scale 2.934

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	118	99.2	6849	25	S76771	TPO=thrombopoietin [hu	1.80e-111
2	118	99.2	7666	26	HUMTA	Human gene for thrombo	1.80e-111
3	70	58.8	1721	10	E11961	Human cDNA encoding th	4.93e-55
4	69	58.0	1691	26	D32047	Homo sapiens mRNA for	6.87e-54
5	41	34.5	1795	25	HUMMLCMPL	Human c-mpl ligand (ML	6.60e-23
6	41	34.5	1795	21	I33525	Sequence 1 from patent	6.60e-23
7	30	25.2	4506	21	E12183	Human thrombopoietin g	1.07e-11
8	30	25.2	4506	10	E11962	Human genomic DNA sequ	1.07e-11
9	30	25.2	4506	21	E12215	Human thrombopoietin g	1.07e-11
c 10	20	16.8	39353	18	CEF11A1	Caenorhabditis elegans	1.43e-02
c 11	18	15.1	455	30	G37636	SHGC-58226 Human Homo	5.92e-01
12	18	15.1	44797	18	CELB0416	Caenorhabditis elegans	5.92e-01
13	18	15.1	128559	26	AC004677	Homo Sapiens Chromosom	5.92e-01
c 14	17	14.3	876	28	MMAT01	Mus musculus aspartate	3.49e+00
c 15	17	14.3	876	28	MUSASPT01	Mouse mitochondrial as	3.49e+00
c 16	17	14.3	22073	18	CELK06H7	C. elegans cosmid K06H	3.49e+00
c 17	17	14.3	31197	17	CEBA2	*** SEQUENCING IN PROG	3.49e+00
18	17	14.3	32043	18	CEC06C3	Caenorhabditis elegans	3.49e+00
c 19	17	14.3	33477	18	CET07C4	Caenorhabditis elegans	3.49e+00
c 20	17	14.3	73805	24	AC004135	Genomic sequence for A	3.49e+00
21	17	14.3	85779	17	AC005085	*** SEQUENCING IN PROG	3.49e+00
22	17	14.3	93845	24	ATAC004684	Arabidopsis thaliana c	3.49e+00
c 23	17	14.3	104793	26	AF007262	Human Xq28 cosmid 3C3	3.49e+00
c 24	17	14.3	123468	17	HS232N11	Human DNA sequence ***	3.49e+00
c 25	17	14.3	132698	17	AC005051	*** SEQUENCING IN PROG	3.49e+00
c 26	17	14.3	133798	17	AC004950	*** SEQUENCING IN PROG	3.49e+00
c 27	17	14.3	161453	27	AC004993	Homo sapiens PAC clone	3.49e+00
c 28	17	14.3	170128	17	AC005059	*** SEQUENCING IN PROG	3.49e+00
29	17	14.3	191908	17	CEY51B9	Caenorhabditis elegans	3.49e+00
c 30	17	14.3	224064	17	AC005055	*** SEQUENCING IN PROG	3.49e+00
31	17	14.3	329751	17	HS407F11	Human DNA sequence ***	3.49e+00
32	16	13.4	448	30	G36497	SHGC-53095 Human Homo	1.92e+01
c 33	16	13.4	623	25	HS3MS630	H.sapiens minisatellit	1.92e+01
c 34	16	13.4	805	25	HSU85248	Human cholesteryl este	1.92e+01
35	16	13.4	1264	20	CHKHLPR	Chicken H1 histone gen	1.92e+01
c 36	16	13.4	1656	25	HUMCETP1	Human cholesteryl este	1.92e+01
37	16	13.4	2610	16	AF052209	Streptococcus pneumoni	1.92e+01
c 38	16	13.4	2683	23	SPBYR1	Fission yeast byrl gen	1.92e+01
c 39	16	13.4	3500	27	HSCETPS1	Human cholesteryl este	1.92e+01
c 40	16	13.4	4429	23	YSCPOL2RPB	Yeast (S.cerevisiae) R	1.92e+01
41	16	13.4	4700	24	AF017056	Arabidopsis thaliana b	1.92e+01
c 42	16	13.4	5672	15	MTRIBC	M.thermoautotrophicum	1.92e+01

c 43	16	13.4	5887	26	AF027656	Homo sapiens cholester	1.92e+01
c 44	16	13.4	8372	20	CHKY	Gallus gallus ovalbumi	1.92e+01
c 45	16	13.4	19574	18	PFSC03016	Plasmodium falciparum	1.92e+01
46	16	13.4	27777	18	CEF43D2	Caenorhabditis elegans	1.92e+01
c 47	16	13.4	29117	28	AC001230	Genomic sequence from	1.92e+01
c 48	16	13.4	30722	18	CELT19A5	Caenorhabditis elegans	1.92e+01
c 49	16	13.4	31201	23	SPAC1D4	S.pombe chromosome I c	1.92e+01
50	16	13.4	32359	27	HSN120B6B	Human DNA sequence fro	1.92e+01
51	16	13.4	32626	26	AC002295	Genomic sequence from	1.92e+01
c 52	16	13.4	35234	25	HSU82213	Human chromosome 10 cl	1.92e+01
c 53	16	13.4	36987	26	U73639	Human Chromosome 11 Co	1.92e+01
54	16	13.4	37512	18	CET09F5	Caenorhabditis elegans	1.92e+01
c 55	16	13.4	48265	26	AC000388	Genomic sequence from	1.92e+01
56	16	13.4	51050	18	AC004340	Drosophila melanogaste	1.92e+01
57	16	13.4	67571	25	HUM1D9	Human BAC 1D9.	1.92e+01
58	16	13.4	83549	26	AC004752	Homo sapiens chromosom	1.92e+01
59	16	13.4	90665	24	ATAC005496	Arabidopsis thaliana c	1.92e+01
c 60	16	13.4	93163	25	HSAC000121	Human BAC clone RG249A	1.92e+01
c 61	16	13.4	93762	17	AC004335	*** SEQUENCING IN PROG	1.92e+01
62	16	13.4	99083	18	CEY37A1B	Caenorhabditis elegans	1.92e+01
63	16	13.4	99207	17	AC002423	*** SEQUENCING IN PROG	1.92e+01
64	16	13.4	100000	26	AP000020	Homo sapiens genomic D	1.92e+01
c 65	16	13.4	123336	26	AC003075	Human PAC clone DJ0658	1.92e+01
66	16	13.4	126270	26	HS211D12	Human DNA sequence fro	1.92e+01
c 67	16	13.4	128915	26	AC002418	Human Chromosome X, co	1.92e+01
68	16	13.4	135259	17	AC005018	*** SEQUENCING IN PROG	1.92e+01
c 69	16	13.4	136299	26	HS359N14	Human DNA sequence fro	1.92e+01
70	16	13.4	140060	17	AC005101	*** SEQUENCING IN PROG	1.92e+01
71	16	13.4	140901	17	HS888M10	Human DNA sequence ***	1.92e+01
72	16	13.4	145414	26	HS78F24	Homo sapiens DNA seque	1.92e+01
c 73	16	13.4	152593	23	ATAC002332	Arabidopsis thaliana c	1.92e+01
74	16	13.4	156490	26	AC004052	Homo sapiens chromosom	1.92e+01
c 75	16	13.4	162485	17	AC004840	*** SEQUENCING IN PROG	1.92e+01
76	16	13.4	172657	26	AC004386	Homo Sapiens Chromosom	1.92e+01
c 77	16	13.4	179707	17	HSAC001234	*** SEQUENCING IN PROG	1.92e+01
c 78	16	13.4	198060	17	AC004831	*** SEQUENCING IN PROG	1.92e+01
79	16	13.4	198456	17	AC003115	*** SEQUENCING IN PROG	1.92e+01
c 80	16	13.4	220000	17	AC005798	*** SEQUENCING IN PROG	1.92e+01
81	16	13.4	221273	17	CEY40H4	Caenorhabditis elegans	1.92e+01
82	16	13.4	243820	17	CEY75B12	Caenorhabditis elegans	1.92e+01
83	16	13.4	316170	17	CEY37A1	Caenorhabditis elegans	1.92e+01
c 84	15	12.6	149	25	HUM63N	Human (clone: pHyTN1/6	9.83e+01
c 85	15	12.6	190	25	HUMTNL11	Human DNA flanking 5'	9.83e+01
c 86	15	12.6	227	30	HUMUT8169B	Human STS UT8169, 3' p	9.83e+01
87	15	12.6	250	30	G14782	human STS SHGC-13542.	9.83e+01
88	15	12.6	272	25	HS191E7R	H.sapiens CpG island D	9.83e+01
c 89	15	12.6	370	28	MUSASSA07	Mouse argininosuccinat	9.83e+01
c 90	15	12.6	380	30	HUM4S1232	Human chromosome 4 (cl	9.83e+01
c 91	15	12.6	380	30	G34013	human STS SHGC-51046,	9.83e+01
92	15	12.6	424	30	G30319	human STS SHGC-36760,	9.83e+01
c 93	15	12.6	452	30	G28248	human STS SHGC-34296.	9.83e+01

c 94	15	12.6	456	25	HSARS1	Human ARS1 sequence, a	9.83e+01
95	15	12.6	651	26	HSCD88	H.sapiens mRNA for CD8	9.83e+01
96	15	12.6	1176	28	CGU31241	Cricetulus griseus int	9.83e+01
c 97	15	12.6	1199	26	HUMHTTP	Human mRNA for alpha-t	9.83e+01
c 98	15	12.6	1199	30	G28587	human STS SHGC-35358,	9.83e+01
99	15	12.6	1246	20	GGTROP3	G.gallus tropomyosin g	9.83e+01
c 100	15	12.6	1336	23	BNAFAD3A	Brasicca napus omega-3	9.83e+01
c 101	15	12.6	1350	23	ATHFADO	Arabidopsis thaliana c	9.83e+01
c 102	15	12.6	1409	23	ATHFADIER2	Arabidopsis thaliana m	9.83e+01
c 103	15	12.6	1522	23	PSU56697	Pisum sativum pyruvate	9.83e+01
c 104	15	12.6	1568	26	AF030424	Homo sapiens histone a	9.83e+01
105	15	12.6	1658	33	OLU57507	Ovine lentivirus envel	9.83e+01
106	15	12.6	1658	33	OLU57509	Ovine lentivirus envel	9.83e+01
107	15	12.6	1658	33	OLU57508	Ovine lentivirus envel	9.83e+01
108	15	12.6	1671	23	AF043610	Brassica oleracea var.	9.83e+01
109	15	12.6	1828	25	HUMCS1PA	Human cleavage signal	9.83e+01
110	15	12.6	1900	21	I02065	Sequence 1 from Patent	9.83e+01
c 111	15	12.6	1943	24	AB011523	Chlorella vulgaris mit	9.83e+01
c 112	15	12.6	1989	19	BTU48360	Bos taurus keratan sul	9.83e+01
113	15	12.6	2000	21	I19467	Sequence 1 from patent	9.83e+01
c 114	15	12.6	2040	26	HSDSC2X1	H.sapiens DSC2 gene, p	9.83e+01
115	15	12.6	2070	15	VIBNEUP	V.proteolyticus neutra	9.83e+01
116	15	12.6	2076	33	AF015180	Caprine arthritis-ence	9.83e+01
117	15	12.6	2293	21	A11978	Synthetic nucleotide s	9.83e+01
118	15	12.6	2294	30	G27040	human STS SHGC-31374.	9.83e+01
119	15	12.6	2294	26	HUMUKA	Homo sapiens mRNA for	9.83e+01
120	15	12.6	2294	21	A09202	Artificial mRNA for ur	9.83e+01
121	15	12.6	2296	21	A35395	H.sapiens u-PA cDNA se	9.83e+01
122	15	12.6	2300	21	A21571	pro-Urokinase gene.	9.83e+01
123	15	12.6	2300	21	A04029	Synthetic H.sapiens pr	9.83e+01
124	15	12.6	2301	21	E01467	DNA encoding human pro	9.83e+01
125	15	12.6	2303	21	I08092	Sequence 1 from Patent	9.83e+01
126	15	12.6	2303	21	I07013	Sequence 4 from Patent	9.83e+01
127	15	12.6	2304	25	HUMUKPM	Human pro-urokinase mR	9.83e+01
128	15	12.6	2304	21	I03932	Sequence 6 from Patent	9.83e+01
129	15	12.6	2304	21	I04632	Sequence 1 from Patent	9.83e+01
130	15	12.6	2304	21	E00178	cDNA encoding human ur	9.83e+01
131	15	12.6	2377	21	I08106	Sequence 1 from Patent	9.83e+01
c 132	15	12.6	2427	23	ATHPAB2A	Arabidopsis thaliana p	9.83e+01
133	15	12.6	2500	28	RATGAP01	Rat GAP-43 gene, exon	9.83e+01
134	15	12.6	2629	25	HUMUK	Human urokinase gene,	9.83e+01
135	15	12.6	2895	33	CEACAEV63	Caprine arthritis ence	9.83e+01
136	15	12.6	3052	23	SCYBR198C	S.cerevisiae chromosom	9.83e+01
137	15	12.6	3128	28	MMU47024	Mus musculus maternal-	9.83e+01
138	15	12.6	3156	28	MUSMX	Mouse DNA, MX-related	9.83e+01
139	15	12.6	3234	23	SCU31217	Saccharomyces cerevisi	9.83e+01
140	15	12.6	3255	15	RPFUSNUS	R.provazekii fusA, trp	9.83e+01
141	15	12.6	3291	28	RATKV1AA	Rat potassium channel-	9.83e+01
c 142	15	12.6	3564	28	MMTABBY	Mus musculus Tabby gen	9.83e+01
c 143	15	12.6	3649	23	SCYOL145C	S.cerevisiae chromosom	9.83e+01
c 144	15	12.6	4019	18	AF033001	Cerebratulus lacteus n	9.83e+01

c	145	15	12.6	4250	23	ATHFAD3	Arabidopsis thaliana f 9.83e+01
	146	15	12.6	4605	23	SCU69264	Saccharomyces cerevisi 9.83e+01
c	147	15	12.6	5042	26	AB011152	Homo sapiens mRNA for 9.83e+01
	148	15	12.6	5206	23	CLAF000309	Colletotrichum lindemu 9.83e+01
	149	15	12.6	5211	28	MMT1	M.musculus T1 gene. 9.83e+01
	150	15	12.6	6431	27	HSLIS10	Human platelet activat 9.83e+01
c	151	15	12.6	6757	16	AF012886	Buchnera aphidicola UD 9.83e+01
	152	15	12.6	6822	15	BSPDMC	B. stearothermophilus 9.83e+01
	153	15	12.6	7038	28	RATACACA	Rat acetyl-coenzyme A 9.83e+01
c	154	15	12.6	7127	23	KLLAC12	Kluyveromyces lactis L 9.83e+01
	155	15	12.6	7177	28	MMU40811	Mus musculus orphan G- 9.83e+01
	156	15	12.6	7258	25	HSUPA	H.sapiens uPA gene. 9.83e+01
c	157	15	12.6	7476	19	AF036962	Bos taurus corneal ker 9.83e+01
	158	15	12.6	9189	33	CEAVCG	Caprine arthritis ence 9.83e+01
c	159	15	12.6	9281	16	MBDR3S1	Mycobacterium bovis de 9.83e+01
	160	15	12.6	10519	15	U67524	Methanococcus jannasch 9.83e+01
	161	15	12.6	10895	15	ECAE000420	Escherichia coli K-12 9.83e+01
	162	15	12.6	11995	15	MMU47274	Methanosarcina mazeii 9.83e+01
c	163	15	12.6	12412	21	I86264	Sequence 18 from paten 9.83e+01
c	164	15	12.6	13001	23	SCXVLORFS	S.cerevisiae chromosom 9.83e+01
	165	15	12.6	14369	16	AE001221	Treponema pallidum sec 9.83e+01
	166	15	12.6	15228	18	PTU20449	Paramecium tetraurelia 9.83e+01
	167	15	12.6	22253	26	HSA1280	Human DNA sequence fro 9.83e+01
c	168	15	12.6	24177	15	AE000785	Borrelia burgdorferi p 9.83e+01
	169	15	12.6	24608	18	CEC11G10	Caenorhabditis elegans 9.83e+01
c	170	15	12.6	26435	18	CELC06E8	Caenorhabditis elegans 9.83e+01
c	171	15	12.6	28835	26	HS444G9	Human DNA sequence fro 9.83e+01
	172	15	12.6	29286	23	SC9952X	S.cerevisiae chromosom 9.83e+01
	173	15	12.6	30182	18	CEF25B3	Caenorhabditis elegans 9.83e+01
	174	15	12.6	30940	18	CELF41D9	Caenorhabditis elegans 9.83e+01
	175	15	12.6	30981	23	SCCHRIIFG	S.cerevisiae RIM, MSI1 9.83e+01
	176	15	12.6	31278	18	CEB0198	Caenorhabditis elegans 9.83e+01
	177	15	12.6	32437	16	MTCY336	Mycobacterium tubercul 9.83e+01
	178	15	12.6	32458	18	CEF49E11	Caenorhabditis elegans 9.83e+01
c	179	15	12.6	33950	18	CEF37B12	Caenorhabditis elegans 9.83e+01
	180	15	12.6	34683	26	HSU09822	Human cosmid clone 26h 9.83e+01
	181	15	12.6	35187	16	MTCY441	Mycobacterium tubercul 9.83e+01
c	182	15	12.6	37385	18	CELF29C4	Caenorhabditis elegans 9.83e+01
	183	15	12.6	37447	18	CEF11A5	Caenorhabditis elegans 9.83e+01
c	184	15	12.6	37531	17	AC002051	*** SEQUENCING IN PROG 9.83e+01
	185	15	12.6	37799	18	CELC05E11	Caenorhabditis elegans 9.83e+01
c	186	15	12.6	37799	18	CELC05E11	Caenorhabditis elegans 9.83e+01
c	187	15	12.6	38041	25	HUMU129A10	Human cosmid U129A10. 9.83e+01
c	188	15	12.6	38500	16	MTCY39	Mycobacterium tubercul 9.83e+01
c	189	15	12.6	40124	18	CEF09E8	Caenorhabditis elegans 9.83e+01
c	190	15	12.6	40301	18	CELD1044	Caenorhabditis elegans 9.83e+01
	191	15	12.6	40780	18	CEM03C11	Caenorhabditis elegans 9.83e+01
c	192	15	12.6	41622	17	HS399E4	Human DNA sequence *** 9.83e+01
c	193	15	12.6	42061	15	MSGY223	Mycobacterium tubercul 9.83e+01
c	194	15	12.6	43843	26	AC002054	Homo sapiens Chromosom 9.83e+01
c	195	15	12.6	46419	25	HSAC000058	Human BAC clone RG111F 9.83e+01

c	196	15	12.6	47167	27	HS757P12	Human DNA sequence fro 9.83e+01
c	197	15	12.6	50511	26	AC005214	Homo sapiens chromosom 9.83e+01
c	198	15	12.6	52777	26	AC000053	Homo sapiens Chromosom 9.83e+01
c	199	15	12.6	66084	24	AB015468	Arabidopsis thaliana g 9.83e+01
c	200	15	12.6	68548	25	HSU70984	Human Xp22 cosmids U10 9.83e+01
c	201	15	12.6	72722	18	AC004373	Drosophila melanogaste 9.83e+01
c	202	15	12.6	74155	26	AC004656	Homo sapiens Xp22 BAC 9.83e+01
	203	15	12.6	77362	17	HS335H7	Human DNA sequence *** 9.83e+01
c	204	15	12.6	78661	26	AC005350	Homo sapiens chromosom 9.83e+01
	205	15	12.6	80514	26	AC004454	Homo sapiens PAC clone 9.83e+01
	206	15	12.6	82032	26	AC002458	Human BAC clone RG098M 9.83e+01
	207	15	12.6	83698	23	AB010072	Arabidopsis thaliana g 9.83e+01
c	208	15	12.6	88543	24	ATAC004680	Arabidopsis thaliana c 9.83e+01
c	209	15	12.6	88882	26	HS448E20	Human DNA sequence fro 9.83e+01
	210	15	12.6	90019	23	TM021B04	Arabidopsis thaliana B 9.83e+01
c	211	15	12.6	91406	17	HS301K23	Human DNA sequence *** 9.83e+01
	212	15	12.6	91863	24	ATAC004667	Arabidopsis thaliana c 9.83e+01
	213	15	12.6	94733	26	HS232P20	Homo sapiens DNA seque 9.83e+01
	214	15	12.6	97129	26	HS292H14	Human DNA sequence fro 9.83e+01
c	215	15	12.6	98067	26	HS435A7	Homo sapiens DNA seque 9.83e+01
	216	15	12.6	100000	26	AP000042	Homo sapiens genomic D 9.83e+01
c	217	15	12.6	100000	26	AP000055	Homo sapiens genomic D 9.83e+01
	218	15	12.6	107183	23	AC002983	A. thaliana BAC T3F12 9.83e+01
c	219	15	12.6	108879	27	AC004804	Homo sapiens 12p13.3 P 9.83e+01
	220	15	12.6	110908	15	SYCSLRD	Synechocystis sp. PCC6 9.83e+01
c	221	15	12.6	111398	26	HS381E2	Human DNA sequence fro 9.83e+01
c	222	15	12.6	113704	25	HS267P19	Human DNA sequence fro 9.83e+01
c	223	15	12.6	114197	17	HS786C22	Human DNA sequence *** 9.83e+01
c	224	15	12.6	114736	18	PFMAL3P3	Plasmodium falciparum 9.83e+01
c	225	15	12.6	118591	17	AC005024	*** SEQUENCING IN PROG 9.83e+01
	226	15	12.6	122325	27	HS1044O17	Human DNA sequence *** 9.83e+01
c	227	15	12.6	122571	26	AF064859	Homo sapiens chromosom 9.83e+01
c	228	15	12.6	125251	17	HS816K17	Human DNA sequence *** 9.83e+01
c	229	15	12.6	126807	17	HS391O22	Human DNA sequence *** 9.83e+01
	230	15	12.6	127757	26	AC004468	Homo sapiens Xp22 PAC 9.83e+01
c	231	15	12.6	129271	24	AC005687	Arabidopsis thaliana B 9.83e+01
c	232	15	12.6	134137	27	HS3E5	Homo sapiens DNA seque 9.83e+01
c	233	15	12.6	134641	17	AC004876	*** SEQUENCING IN PROG 9.83e+01
	234	15	12.6	135505	27	AC004477	Homo sapiens chromosom 9.83e+01
	235	15	12.6	135886	26	HSXIST2	Human Xq13 3' end of P 9.83e+01
c	236	15	12.6	136150	27	AC002081	Homo sapiens BAC clone 9.83e+01
	237	15	12.6	136561	17	HS430A16	Human DNA sequence *** 9.83e+01
	238	15	12.6	137065	17	HS29M10	Human DNA sequence *** 9.83e+01
c	239	15	12.6	137246	26	HS49J10	Human DNA sequence fro 9.83e+01
c	240	15	12.6	140820	17	HS439F8	Human DNA sequence *** 9.83e+01
	241	15	12.6	141460	17	AC004829	*** SEQUENCING IN PROG 9.83e+01
	242	15	12.6	141475	26	AC002980	Homo sapiens Xp22 BAC 9.83e+01
c	243	15	12.6	141589	26	HS117P20	Homo sapiens DNA seque 9.83e+01
c	244	15	12.6	141878	26	AC000004	Genomic sequence from 9.83e+01
c	245	15	12.6	156197	25	HS97K10	Human DNA sequence fro 9.83e+01
c	246	15	12.6	157379	26	AC004997	Homo sapiens PAC clone 9.83e+01

	247	15	12.6	157412	17	HS376D21	Human DNA sequence *** 9.83e+01
c	248	15	12.6	158257	27	AC005585	Homo sapiens chromosom 9.83e+01
	249	15	12.6	159311	17	HS167A19	Human DNA sequence *** 9.83e+01
c	250	15	12.6	162455	17	AC005536	*** SEQUENCING IN PROG 9.83e+01
	251	15	12.6	162531	26	AC005249	Homo sapiens BAC clone 9.83e+01
	252	15	12.6	162822	27	AC005751	Homo sapiens chromosom 9.83e+01
c	253	15	12.6	163216	26	HUAC002289	Homo sapiens Chromosom 9.83e+01
	254	15	12.6	165507	17	AC004903	*** SEQUENCING IN PROG 9.83e+01
	255	15	12.6	166219	26	AC003084	Human BAC clone RG084D 9.83e+01
c	256	15	12.6	166818	17	AC005233	*** SEQUENCING IN PROG 9.83e+01
	257	15	12.6	167880	27	HUU91323	Human Chromosome 16 BA 9.83e+01
	258	15	12.6	168636	17	AC002060	*** SEQUENCING IN PROG 9.83e+01
	259	15	12.6	169693	26	HS57A13	Human DNA sequence fro 9.83e+01
	260	15	12.6	170658	26	AC004072	Human Chromosome X clo 9.83e+01
c	261	15	12.6	170952	26	HS501N12	Homo sapiens DNA seque 9.83e+01
c	262	15	12.6	175745	25	HS24608	Human DNA sequence fro 9.83e+01
	263	15	12.6	178660	26	HSG1150	Homo sapiens chromosom 9.83e+01
c	264	15	12.6	180114	17	HS52D1	Human DNA sequence *** 9.83e+01
	265	15	12.6	184263	27	AC004805	Homo sapiens Xp22 bins 9.83e+01
	266	15	12.6	185364	17	AC005477	*** SEQUENCING IN PROG 9.83e+01
c	267	15	12.6	191028	23	PPU38804	Porphyra purpurea chlo 9.83e+01
	268	15	12.6	192335	17	AC005042	*** SEQUENCING IN PROG 9.83e+01
c	269	15	12.6	192335	17	AC005042	*** SEQUENCING IN PROG 9.83e+01
	270	15	12.6	193869	17	AC005043	*** SEQUENCING IN PROG 9.83e+01
c	271	15	12.6	196559	17	HS399M14	Human DNA sequence *** 9.83e+01
	272	15	12.6	198552	23	ATAP22	Arabidopsis thaliana D 9.83e+01
	273	15	12.6	202233	26	AC003664	Homo sapiens chromosom 9.83e+01
c	274	15	12.6	210364	17	AC005537	*** SEQUENCING IN PROG 9.83e+01
c	275	15	12.6	218470	15	BSUB0013	Bacillus subtilis comp 9.83e+01
	276	15	12.6	243037	17	AC000353	*** SEQUENCING IN PROG 9.83e+01
c	277	15	12.6	245265	17	HS581F12	Human DNA sequence *** 9.83e+01
	278	15	12.6	282700	15	BACJH642	Bacillus subtilis DNA, 9.83e+01
c	279	15	12.6	305444	17	CET67H2	Caenorhabditis elegans 9.83e+01
	280	15	12.6	330836	17	HS78D13	Human DNA sequence *** 9.83e+01
c	281	14	11.8	18	21	I25836	Sequence 16 from paten 4.62e+02
c	282	14	11.8	60	21	E04052	Synthetic DNA sequence 4.62e+02
c	283	14	11.8	71	27	HSPDEA22	Human phosphodiesteras 4.62e+02
	284	14	11.8	129	25	HUMMHCFANC	Human MHC class I mRNA 4.62e+02
	285	14	11.8	165	25	HUMOAT10	Human ornithine aminot 4.62e+02
	286	14	11.8	205	25	HS37E1F	H.sapiens CpG island D 4.62e+02
	287	14	11.8	227	28	RATAPOBPI	Rat apolipoprotein B P 4.62e+02
	288	14	11.8	249	28	RATSATELLB	Rattus norvegicus poly 4.62e+02
	289	14	11.8	250	30	G15338	human STS SHGC-16030. 4.62e+02
c	290	14	11.8	250	30	G15383	human STS SHGC-16115. 4.62e+02
c	291	14	11.8	256	21	I02003	Sequence 3 from Patent 4.62e+02
	292	14	11.8	257	30	G04903	human STS WI-5359. 4.62e+02
	293	14	11.8	265	26	HS4COL5A18	Homo sapiens type IV c 4.62e+02
c	294	14	11.8	273	30	HSA133WF1	H.sapiens DNA segment 4.62e+02
c	295	14	11.8	297	33	HIVU41711	Human immunodeficiency 4.62e+02
c	296	14	11.8	308	30	HSB317ZB9	H.sapiens DNA segment 4.62e+02
	297	14	11.8	317	25	HUMTEC12A	Homo sapiens (clone te 4.62e+02

c	298	14	11.8	317	30	G05317	human STS WI-6109. 4.62e+02
c	299	14	11.8	319	30	HS309XB1	H.sapiens DNA segment 4.62e+02
c	300	14	11.8	334	19	ECVHL81	E.caballus DNA segment 4.62e+02
c	301	14	11.8	337	25	HUMCD1C1	Human cortical thymocy 4.62e+02
	302	14	11.8	355	30	HUM4STS865	Human Chromosome 4 (cl 4.62e+02
c	303	14	11.8	378	28	RATTHYP	Rat spleen parathymosi 4.62e+02
c	304	14	11.8	378	25	HSC7EXN1	H.sapiens C7 gene, exo 4.62e+02
c	305	14	11.8	378	28	RATPARA	Rattus norvegicus para 4.62e+02
	306	14	11.8	383	30	HSA052WH1	H.sapiens (D881822) DN 4.62e+02
	307	14	11.8	394	30	HUM4STS808	Human Chromosome 4 (cl 4.62e+02
c	308	14	11.8	448	28	AF007418	Rattus norvegicus micr 4.62e+02
	309	14	11.8	450	18	U91361	Nabis sp. Damsel.bug.3 4.62e+02
	310	14	11.8	454	19	PIGAROMA	Sus scrofa cytochrome 4.62e+02
	311	14	11.8	454	19	SS1P450S6	Sus scrofa type III cy 4.62e+02
	312	14	11.8	459	30	AF026448	Poeciliopsis occidenta 4.62e+02
	313	14	11.8	466	30	G37557	SHGC-57988 Human Homo 4.62e+02
c	314	14	11.8	522	21	I13451	Sequence 4 from patent 4.62e+02
c	315	14	11.8	528	21	E07528	DNA sequences coding h 4.62e+02
c	316	14	11.8	549	30	HUMUT592	Human chromosome 1 STS 4.62e+02
	317	14	11.8	569	30	HUMUT5425	Human STS UT5425, sequ 4.62e+02
c	318	14	11.8	594	30	AF026450	Poeciliopsis occidenta 4.62e+02
c	319	14	11.8	600	15	TMDYRA	T.maritima dyrA gene. 4.62e+02
c	320	14	11.8	608	30	G39788	Z10372 Zebrafish AB Da 4.62e+02
c	321	14	11.8	629	30	AF026447	Poeciliopsis occidenta 4.62e+02
c	322	14	11.8	639	26	AG000453	Homo sapiens genomic D 4.62e+02
c	323	14	11.8	645	26	AG000452	Homo sapiens genomic D 4.62e+02
	324	14	11.8	649	26	AG000431	Homo sapiens genomic D 4.62e+02
	325	14	11.8	654	26	AG000430	Homo sapiens genomic D 4.62e+02
	326	14	11.8	677	30	HUMUT6950	Human STS UT6950, sequ 4.62e+02
	327	14	11.8	725	23	DCGLYR	D. carota mRNA for a g 4.62e+02
c	328	14	11.8	760	25	HSHIP	H.sapiens HIP mRNA. 4.62e+02
	329	14	11.8	767	25	HSUATM18	Human ataxia-telangiec 4.62e+02
c	330	14	11.8	774	20	XELINSB	X.laevis insulin I mRN 4.62e+02
c	331	14	11.8	783	23	CSZ82955	C.solida plastid rps16 4.62e+02
c	332	14	11.8	784	26	HUMPAPC	Human mRNA for PAP hom 4.62e+02
c	333	14	11.8	786	23	CIZ82946	C.intermedia plastid r 4.62e+02
c	334	14	11.8	786	23	CGZ82957	C.gotlandica plastid r 4.62e+02
c	335	14	11.8	786	23	CVZ82949	C.vesicaria plastid rp 4.62e+02
c	336	14	11.8	788	23	PLZ82953	P.lutea plastid rps16 4.62e+02
c	337	14	11.8	789	23	RNZ82954	R.numidica plastid rps 4.62e+02
c	338	14	11.8	789	23	AFZ82944	A.fungosa plastid rps1 4.62e+02
c	339	14	11.8	791	23	CEZ82948	C.edulis plastid rps16 4.62e+02
	340	14	11.8	795	15	BFU64314	Bacillus firmus putati 4.62e+02
c	341	14	11.8	797	23	CRZ82942	C.rupestris plastid rp 4.62e+02
c	342	14	11.8	797	25	S51768	PAP-H=pancreatitis-ass 4.62e+02
c	343	14	11.8	798	23	CFZ82943	C.fimbrillifera plasti 4.62e+02
c	344	14	11.8	798	21	I13456	Sequence 12 from paten 4.62e+02
c	345	14	11.8	803	23	CSZ82956	C.sempervirens plastid 4.62e+02
c	346	14	11.8	867	26	HSTBXA06	Homo sapiens thromboxa 4.62e+02
	347	14	11.8	869	18	AF017047	Dictyostelium discoide 4.62e+02
c	348	14	11.8	905	23	ATHGTPB	Arabidopsis thaliana G 4.62e+02

	349	14	11.8	925	18	AF017044	Dictyostelium discoide 4.62e+02
	350	14	11.8	938	25	HSU41644	Human germline Ig kapp 4.62e+02
	351	14	11.8	942	25	HSU41645	Human germline Ig kapp 4.62e+02
	352	14	11.8	943	25	HSU41643	Human germline Ig kapp 4.62e+02
	353	14	11.8	998	23	PSPETP	P.sativum petE gene fo 4.62e+02
c	354	14	11.8	1016	23	SCYKL167C	S.cerevisiae chromosom 4.62e+02
	355	14	11.8	1023	28	RNU33193	Rattus norvegicus gala 4.62e+02
	356	14	11.8	1033	23	TOBNDYC	Tobacco RNA-binding gl 4.62e+02
	357	14	11.8	1044	21	E12657	cDNA encoding galanin 4.62e+02
	358	14	11.8	1044	21	E12486	Mouse cDNA encoding a 4.62e+02
c	359	14	11.8	1048	20	CRAVIA2	Carassius auratus vime 4.62e+02
	360	14	11.8	1061	26	HUMT28P	Human DNA from chromos 4.62e+02
	361	14	11.8	1086	23	HVU58601	Hydrocotyle verticilla 4.62e+02
	362	14	11.8	1098	25	HSU14134	Human transcription fa 4.62e+02
	363	14	11.8	1099	15	CBFLANGE	C.botulinum (strain pr 4.62e+02
c	364	14	11.8	1101	15	TMDIHYDRE	T.maritima gene encodi 4.62e+02
	365	14	11.8	1125	23	AF009004	Pelargonium x hortorum 4.62e+02
c	366	14	11.8	1126	30	G06593	human STS WI-7563. 4.62e+02
	367	14	11.8	1133	19	SSU52141	Sus scrofa aromatase (4.62e+02
c	368	14	11.8	1133	23	ATHGF14A	Arabidopsis thaliana G 4.62e+02
	369	14	11.8	1173	25	HSU20272	Human DNA/RNA-binding 4.62e+02
c	370	14	11.8	1182	19	OCMAPKK	O.cuniculus mRNA for M 4.62e+02
c	371	14	11.8	1207	25	HUMCD1C	Human thymocyte antige 4.62e+02
c	372	14	11.8	1231	28	MUSEN1S4	Mouse Edn1 gene for pr 4.62e+02
	373	14	11.8	1241	25	HUMIGKVAE	Human germline kappa-c 4.62e+02
	374	14	11.8	1247	25	HSIKLVA18	H.sapiens DNA for IG k 4.62e+02
	375	14	11.8	1276	23	YSCMRP49A	Saccharomyces cerevisi 4.62e+02
c	376	14	11.8	1286	28	MUSUTROPH	Mouse utrophin (Dmd1) 4.62e+02
c	377	14	11.8	1308	28	USTNFR20S2	Mus musculus tumor nec 4.62e+02
c	378	14	11.8	1323	23	YSCRPC19	S.cerevisiae AC19 RNA 4.62e+02
c	379	14	11.8	1330	25	HUMETN4	Human endothelin (ETN) 4.62e+02
c	380	14	11.8	1358	28	S65225S1	PEP-19=neuron-specific 4.62e+02
	381	14	11.8	1371	16	PAU17293	Pseudomonas aeruginosa 4.62e+02
	382	14	11.8	1381	26	HUMIIIA	Human GTF3A mRNA for X 4.62e+02
	383	14	11.8	1399	10	E10962	cDNA's encoding hTFIII 4.62e+02
c	384	14	11.8	1402	26	HSA224353	Homo sapiens NIS gene, 4.62e+02
c	385	14	11.8	1404	28	MMU97327	Mus musculus calcyclin 4.62e+02
c	386	14	11.8	1424	15	BSU26732	Bacillus sphaericus D- 4.62e+02
	387	14	11.8	1450	16	AB012059	Uncultured haloarchaeo 4.62e+02
	388	14	11.8	1450	16	AB012050	Uncultured haloarchaeo 4.62e+02
c	389	14	11.8	1454	15	STRDN87577	S.pneumoniae (strain D 4.62e+02
c	390	14	11.8	1455	15	STRPBP2BA	S.sanguis penicillin-b 4.62e+02
	391	14	11.8	1484	18	AF015821	Scolopendra sp. retrot 4.62e+02
c	392	14	11.8	1491	28	SQIMHCIK95	Sciurus aberti keibab 4.62e+02
c	393	14	11.8	1493	23	D89167	Schizosaccharomyces po 4.62e+02
c	394	14	11.8	1501	28	MUSPCP4A	Mus musculus Purkinje 4.62e+02
	395	14	11.8	1505	23	PSPLAS	Pea gene for plastocya 4.62e+02
	396	14	11.8	1514	19	SSU92246	Sus scrofa aromatase c 4.62e+02
	397	14	11.8	1517	19	AF001288	Bos taurus histone H4. 4.62e+02
c	398	14	11.8	1518	23	PIAEMB21R	Picea glauca heat shoc 4.62e+02
	399	14	11.8	1520	19	SSU92245	Sus scrofa aromatase c 4.62e+02

c	400	14	11.8	1522	28	SQIMHCIA85	Scurius aberti aberti 4.62e+02
	401	14	11.8	1584	19	SSU52142	Sus scrofa aromatase t 4.62e+02
c	402	14	11.8	1585	23	SCYNL114C	S.cerevisiae chromosom 4.62e+02
	403	14	11.8	1599	25	HSKOAT	Human mRNA for kidney 4.62e+02
	404	14	11.8	1608	18	NFU37719	Naegleria fowleri acti 4.62e+02
c	405	14	11.8	1668	25	AF007140	Homo sapiens clone 237 4.62e+02
	406	14	11.8	1674	28	MMGALR3	Mus musculus galanin r 4.62e+02
	407	14	11.8	1683	28	MMU48732	Mus musculus PB-induci 4.62e+02
	408	14	11.8	1683	23	AF054455	Leavenworthia crassa c 4.62e+02
	409	14	11.8	1702	19	RABZP	Rabbit zona pellucida 4.62e+02
	410	14	11.8	1712	23	SCYLR140W	S.cerevisiae chromosom 4.62e+02
c	411	14	11.8	1740	26	AF059566	Homo sapiens sodium-io 4.62e+02
c	412	14	11.8	1770	28	MMU35233	Mus musculus preproend 4.62e+02
	413	14	11.8	1772	23	SACATALAS	Soldanella alpina mRNA 4.62e+02
	414	14	11.8	1830	33	S67522	A3R=P9L homolog. . .A1 4.62e+02
	415	14	11.8	1904	28	MMT15004	Mus musculus mRNA for 4.62e+02
c	416	14	11.8	1934	23	PHGCR1	Petunia grp-1 gene for 4.62e+02
c	417	14	11.8	1969	33	PE09100	Pea enation mosaic vir 4.62e+02
c	418	14	11.8	1970	33	PE09099	Pea enation mosaic vir 4.62e+02
	419	14	11.8	1976	24	NTA17503	Nicotiana tabacum an.2 4.62e+02
	420	14	11.8	2000	25	HUMOATM	Human ornithine aminot 4.62e+02
	421	14	11.8	2002	25	HUMOATC	Human ornithine-oxo-ac 4.62e+02
	422	14	11.8	2009	24	NTA17502	Nicotiana tabacum an.1 4.62e+02
	423	14	11.8	2013	25	HUMOAT	Human ornithine aminot 4.62e+02
c	424	14	11.8	2026	24	SPU59915	Schizosaccharomyces po 4.62e+02
c	425	14	11.8	2043	25	HSMPP4II	H.sapiens mRNA for M-p 4.62e+02
c	426	14	11.8	2048	28	MUSMTNFR1	Mouse tumor necrosis f 4.62e+02
c	427	14	11.8	2062	23	SONIRR	Spinach mRNA for nitri 4.62e+02
c	428	14	11.8	2063	28	MMP55R	Murine mRNA for p55 tu 4.62e+02
c	429	14	11.8	2088	25	HUMCYB5A01	Human cytochrome b5 (C 4.62e+02
c	430	14	11.8	2130	25	HSHSP1	H.sapiens h-Sp1 mRNA. 4.62e+02
	431	14	11.8	2136	21	A03759	H.sapiens (pXL53) gene 4.62e+02
	432	14	11.8	2136	21	A30885	Synthetic human serum 4.62e+02
c	433	14	11.8	2136	21	A03758	H.sapiens (pXL53) gene 4.62e+02
c	434	14	11.8	2154	28	MUSTNFX	Mus musculus Tumor Nec 4.62e+02
c	435	14	11.8	2179	28	MMTNFRE5	Mouse mRNA for 55-kDa 4.62e+02
c	436	14	11.8	2193	26	D88154	Homo sapiens mRNA for 4.62e+02
c	437	14	11.8	2250	23	ATU09376	Arabidopsis thaliana G 4.62e+02
	438	14	11.8	2363	19	BOVSGII	Cow secretogranin II (4.62e+02
	439	14	11.8	2386	23	TOBPARC	Tobacco DNA for PAR-C, 4.62e+02
c	440	14	11.8	2449	23	YSCYKR2A	Saccharomyces cerevisi 4.62e+02
c	441	14	11.8	2456	25	HSBMXGENE	H.sapiens Bmx mRNA for 4.62e+02
c	442	14	11.8	2458	19	D45239	Rabbit mRNA for Arnt, 4.62e+02
	443	14	11.8	2470	19	SSU37311	Sus scrofa clone ARO33 4.62e+02
c	444	14	11.8	2491	28	RNU28830	Rattus norvegicus RalB 4.62e+02
	445	14	11.8	2509	20	XLU57455	Xenopus laevis growth 4.62e+02
c	446	14	11.8	2533	16	AF081278	Bacillus sphaericus D- 4.62e+02
c	447	14	11.8	2533	26	HUMRASGAPB	Human mRNA for rasGTPa 4.62e+02
c	448	14	11.8	2562	26	D82880	Human brain mRNA for h 4.62e+02
	449	14	11.8	2588	19	SSU37312	Sus scrofa clone ARO34 4.62e+02
c	450	14	11.8	2604	26	AF045459	Homo sapiens Etk/Bmx c 4.62e+02

c	451	14	11.8	2672	25	HSU76992	Human Tat-SF1 mRNA, co 4.62e+02
	452	14	11.8	2673	28	MMWHTFACT	M.musculus mRNA for wi 4.62e+02
c	453	14	11.8	2678	27	HSU66561	Human kruppel-related 4.62e+02
c	454	14	11.8	2682	24	ACY17809	Anthoceros crispulus m 4.62e+02
	455	14	11.8	2715	22	STHK022	Coliphage HK022 int an 4.62e+02
c	456	14	11.8	2751	23	YSCHOLIG	Saccharomyces cerevisi 4.62e+02
	457	14	11.8	2755	20	XELPGFR4	Xenopus laevis mRNA fo 4.62e+02
c	458	14	11.8	2757	28	RNU25651	Rattus norvegicus phos 4.62e+02
c	459	14	11.8	2774	23	SCYNL115C	S.cerevisiae chromosom 4.62e+02
c	460	14	11.8	2961	18	DMU97685	Drosophila melanogaste 4.62e+02
c	461	14	11.8	2975	21	E07294	genomic DNA encoding d 4.62e+02
	462	14	11.8	2979	28	MMU86016	Mus musculus desmoglei 4.62e+02
	463	14	11.8	3004	20	AB007036	Xenopus laevis mRNA fo 4.62e+02
	464	14	11.8	3128	28	RATAPOLPB	Rattus norvegicus (clo 4.62e+02
	465	14	11.8	3154	19	BOVPTHG	Bovine parathyroid hor 4.62e+02
	466	14	11.8	3160	15	SPAJ5619	Streptococcus pneumoni 4.62e+02
	467	14	11.8	3160	15	SPPPR3	Streptococcus pneumoni 4.62e+02
	468	14	11.8	3161	15	AF047696	Streptococcus pneumoni 4.62e+02
	469	14	11.8	3162	15	AF047385	Streptococcus pneumoni 4.62e+02
c	470	14	11.8	3169	15	CCMCPA	C. crescentus mcpA gen 4.62e+02
	471	14	11.8	3186	25	HSOZF	H.sapiens OZF mRNA. 4.62e+02
c	472	14	11.8	3307	15	MMBLAMOR	M.morganii blaMOR gene 4.62e+02
c	473	14	11.8	3308	18	SUSETS	Strongylocentrotus pur 4.62e+02
	474	14	11.8	3320	25	S78234	nuc2 homolog [human, f 4.62e+02
	475	14	11.8	3441	28	RNU83112	Rattus norvegicus INS- 4.62e+02
c	476	14	11.8	3444	18	APGUCYC	Sea urchin (Arbacia pu 4.62e+02
c	477	14	11.8	3461	18	AF012885	Sarcocystis capracanis 4.62e+02
c	478	14	11.8	3477	21	I79236	Sequence 13 from paten 4.62e+02
	479	14	11.8	3556	26	AB00189S14	Homo sapiens PACE4 gen 4.62e+02
	480	14	11.8	3578	24	AB010388	Yarrowia lipolytica ge 4.62e+02
	481	14	11.8	3600	33	HS1UL4	Herpes simplex virus t 4.62e+02
c	482	14	11.8	3622	28	RNU82623	Rattus norvegicus cyto 4.62e+02
	483	14	11.8	3681	23	SCYOR071C	S.cerevisiae chromosom 4.62e+02
	484	14	11.8	3684	23	SCYOR070C	S.cerevisiae chromosom 4.62e+02
c	485	14	11.8	3696	15	AF065243	Agrobacterium tumefaci 4.62e+02
	486	14	11.8	3715	26	HSAC001156	Homo sapiens (subclone 4.62e+02
c	487	14	11.8	3732	28	AB009999	Rattus norvegicus mRNA 4.62e+02
c	488	14	11.8	3923	25	HSU57796	Human zinc finger prot 4.62e+02
	489	14	11.8	4084	23	SCYNR056C	S.cerevisiae chromosom 4.62e+02
	490	14	11.8	4199	15	CLOBONT	Clostridium botulinum 4.62e+02
	491	14	11.8	4209	15	CBBONTPG	C.botulinum bont/f gen 4.62e+02
c	492	14	11.8	4234	23	SONIR	S. oleracera NIR gene 4.62e+02
c	493	14	11.8	4249	16	AF053771	Staphylococcus aureus 4.62e+02
	494	14	11.8	4329	20	CHKCCPA	Gallus domesticus cone 4.62e+02
c	495	14	11.8	4422	23	TOBPSAH2	Nicotiana sylvestris p 4.62e+02
	496	14	11.8	4848	28	RNESTPROM	R.norvegicus gene enco 4.62e+02
c	497	14	11.8	4848	27	HUMLD51EP3	Human zinc finger prot 4.62e+02
	498	14	11.8	4854	28	AF038388	Rattus norvegicus acti 4.62e+02
	499	14	11.8	4863	28	AF050182	Mus musculus PERIOD 3 4.62e+02
c	500	14	11.8	4865	23	ATU20625	Arabidopsis thaliana c 4.62e+02
	501	14	11.8	5017	18	MSQHSP70A	Anopheles albimanus he 4.62e+02

c	502	14	11.8	5017	18	MSQHSP70A	Anopheles albimanus he 4.62e+02
c	503	14	11.8	5082	26	D87437	Human mRNA for KIAA025 4.62e+02
c	504	14	11.8	5085	28	MMMANNREC	M.musculus mRNA for ma 4.62e+02
	505	14	11.8	5282	30	GGALASY1	Chicken gene for mt S- 4.62e+02
	506	14	11.8	5334	23	DCY15177	Daucus carota DNA frag 4.62e+02
	507	14	11.8	5600	18	PFSC03005	Plasmodium falciparum 4.62e+02
	508	14	11.8	5653	15	PGPGAAGEN	P.gingivalis rnhB & pg 4.62e+02
c	509	14	11.8	5706	33	PEQPOLYZ	Pea enation mosaic vir 4.62e+02
	510	14	11.8	5766	27	AB007963	Homo sapiens mRNA for 4.62e+02
c	511	14	11.8	5926	18	PFSC04067	Plasmodium falciparum 4.62e+02
c	512	14	11.8	6112	22	SPP1HEADG	Bacteriophage SPP1 hea 4.62e+02
c	513	14	11.8	6241	15	PGU68468	Porphyromonas gingival 4.62e+02
c	514	14	11.8	6519	10	E11754	cDNA coding DOCK180 pr 4.62e+02
c	515	14	11.8	6519	26	HUMDOCK180	Human DOCK180 protein 4.62e+02
	516	14	11.8	6985	23	AF032443	Magnaporthe grisea ABC 4.62e+02
	517	14	11.8	7310	23	SCYNR055C	S.cerevisiae chromosom 4.62e+02
c	518	14	11.8	7486	15	ECPCO	E.coli plasmid pRJ1004 4.62e+02
c	519	14	11.8	7769	19	BTU55042	Bos taurus myosin X, c 4.62e+02
	520	14	11.8	7845	23	SCPPR1GEN	S.cerevisiae DNA for O 4.62e+02
c	521	14	11.8	8139	15	PGU54691	Porphyromonas gingival 4.62e+02
	522	14	11.8	9092	15	POOCT	P.oleovorans TF4-1L (+ 4.62e+02
c	523	14	11.8	9181	15	AFAHOXGA	Alcaligenes eutrophus 4.62e+02
	524	14	11.8	9203	33	S51392	gag. . .rev (maedi-vis 4.62e+02
	525	14	11.8	9424	23	OSU72724	Oryza sativa receptor 4.62e+02
	526	14	11.8	9470	16	PDU34346	Paracoccus denitrifica 4.62e+02
	527	14	11.8	9608	15	CSCYTOX	C.sordelii cytotoxin g 4.62e+02
c	528	14	11.8	9963	15	ECAE000186	Escherichia coli K-12 4.62e+02
	529	14	11.8	10650	15	AE000904	Methanobacterium therm 4.62e+02
	530	14	11.8	12072	15	VIBEPSCN	Vibrio cholerae genera 4.62e+02
	531	14	11.8	12095	31	BINHYGDNA	Binary vector BinHygTO 4.62e+02
	532	14	11.8	12448	16	AE001270	Treponema pallidum sec 4.62e+02
c	533	14	11.8	12461	25	HUMEDN1B	Homo sapiens endotheli 4.62e+02
	534	14	11.8	13080	15	AE001007	Archaeoglobus fulgidus 4.62e+02
	535	14	11.8	13911	31	PDGUSBTD	Promoter-trapping vect 4.62e+02
	536	14	11.8	14440	26	HSUSF2	H.sapiens USF2 gene. 4.62e+02
	537	14	11.8	14795	26	HSRCAED	H.sapiens red cell ani 4.62e+02
	538	14	11.8	14951	15	AE000756	Aquifex aeolicus secti 4.62e+02
	539	14	11.8	15202	23	ATFROHABG	A.thaliana frohA and f 4.62e+02
c	540	14	11.8	15295	15	D90723	Escherichia coli genom 4.62e+02
	541	14	11.8	15820	15	AE000737	Aquifex aeolicus secti 4.62e+02
c	542	14	11.8	15875	18	CELF49E8	Caenorhabditis elegans 4.62e+02
c	543	14	11.8	16176	15	U67571	Methanococcus jannasch 4.62e+02
	544	14	11.8	18121	18	CEF08E10	Caenorhabditis elegans 4.62e+02
	545	14	11.8	19463	18	CET19E10	Caenorhabditis elegans 4.62e+02
c	546	14	11.8	19545	18	CELH04J21	Caenorhabditis elegans 4.62e+02
c	547	14	11.8	19714	15	D90722	Escherichia coli genom 4.62e+02
	548	14	11.8	20074	27	HSA9794	Homo sapiens hoxll pro 4.62e+02
	549	14	11.8	20569	15	HPAE000636	Helicobacter pylori se 4.62e+02
c	550	14	11.8	20592	23	SCDCHRL11	S.cerevisiae (3288C) 2 4.62e+02
c	551	14	11.8	20776	18	CEK03D10	Caenorhabditis elegans 4.62e+02
c	552	14	11.8	20976	15	HPAE000631	Helicobacter pylori se 4.62e+02

	553	14	11.8	21319	25	HUMAE1ERY	Human anion exchanger 4.62e+02
c	554	14	11.8	21548	15	BACYCB20	Bacillus subtilis DNA 4.62e+02
c	555	14	11.8	22253	26	HSA1280	Human DNA sequence fro 4.62e+02
c	556	14	11.8	22607	18	CELC48A7	Caenorhabditis elegans 4.62e+02
	557	14	11.8	23379	18	CELT08E11	Caenorhabditis elegans 4.62e+02
	558	14	11.8	24418	16	PAU50396	Pseudomonas aeruginosa 4.62e+02
c	559	14	11.8	24652	18	CEW04G5	Caenorhabditis elegans 4.62e+02
	560	14	11.8	25400	18	CEF59A2	Caenorhabditis elegans 4.62e+02
c	561	14	11.8	26184	24	SPBC1A45P	S.pombe chromosome II 4.62e+02
c	562	14	11.8	26881	25	HSV434E11	Human DNA sequence fro 4.62e+02
c	563	14	11.8	27008	18	CEC34E7	Caenorhabditis elegans 4.62e+02
c	564	14	11.8	27450	18	CEK09C8	Caenorhabditis elegans 4.62e+02
c	565	14	11.8	27645	18	CELZK1055	Caenorhabditis elegans 4.62e+02
c	566	14	11.8	27878	25	HSJ3E10	Human DNA sequence fro 4.62e+02
	567	14	11.8	28678	18	CELT04D1	Caenorhabditis elegans 4.62e+02
	568	14	11.8	28701	18	CET01H8	Caenorhabditis elegans 4.62e+02
c	569	14	11.8	29096	18	CEC05C12	Caenorhabditis elegans 4.62e+02
	570	14	11.8	29154	23	YSCL9606	Saccharomyces cerevisi 4.62e+02
	571	14	11.8	29303	18	CEF57G4	Caenorhabditis elegans 4.62e+02
c	572	14	11.8	29324	15	BS16829KB	Bacillus subtilis 29kB 4.62e+02
	573	14	11.8	29363	18	CELZK185	Caenorhabditis elegans 4.62e+02
	574	14	11.8	30419	18	CELT07H8	Caenorhabditis elegans 4.62e+02
c	575	14	11.8	30563	18	CET27F6	Caenorhabditis elegans 4.62e+02
c	576	14	11.8	30827	18	CEW09H1	Caenorhabditis elegans 4.62e+02
	577	14	11.8	30875	27	HSN20A6	Human DNA sequence fro 4.62e+02
c	578	14	11.8	31661	18	CELM60	Caenorhabditis elegans 4.62e+02
c	579	14	11.8	31975	24	SPCC553	S.pombe chromosome III 4.62e+02
	580	14	11.8	31988	18	CELC50D2	Caenorhabditis elegans 4.62e+02
	581	14	11.8	32000	24	SPCC1183	S.pombe chromosome III 4.62e+02
c	582	14	11.8	32164	18	CEC09G9	Caenorhabditis elegans 4.62e+02
c	583	14	11.8	32266	18	CER07E5	Caenorhabditis elegans 4.62e+02
c	584	14	11.8	32704	18	CEF08G5	Caenorhabditis elegans 4.62e+02
c	585	14	11.8	32960	18	CEK02E11	Caenorhabditis elegans 4.62e+02
c	586	14	11.8	33164	18	CEF35C5	Caenorhabditis elegans 4.62e+02
c	587	14	11.8	33330	18	CELF37C12	Caenorhabditis elegans 4.62e+02
c	588	14	11.8	33573	18	CELP09E10	Caenorhabditis elegans 4.62e+02
	589	14	11.8	33746	18	CET09E11	Caenorhabditis elegans 4.62e+02
	590	14	11.8	33772	18	CELC35B1	Caenorhabditis elegans 4.62e+02
	591	14	11.8	34372	18	CELZC250	Caenorhabditis elegans 4.62e+02
	592	14	11.8	34372	18	CELD1005	Caenorhabditis elegans 4.62e+02
c	593	14	11.8	34388	18	CET28A8	Caenorhabditis elegans 4.62e+02
c	594	14	11.8	35024	18	CEW02B9	Caenorhabditis elegans 4.62e+02
	595	14	11.8	35056	23	SPAC3C7	S.pombe chromosome I c 4.62e+02
	596	14	11.8	35121	17	CEC43A10	Caenorhabditis elegans 4.62e+02
c	597	14	11.8	35240	18	CEF42G4	Caenorhabditis elegans 4.62e+02
c	598	14	11.8	35284	17	CEH27E01	Caenorhabditis elegans 4.62e+02
	599	14	11.8	35444	18	CELK09C4	Caenorhabditis elegans 4.62e+02
	600	14	11.8	35506	26	HS24E5	Human DNA sequence fro 4.62e+02
	601	14	11.8	35811	23	SC9718	S.cerevisiae chromosom 4.62e+02
	602	14	11.8	35843	25	HSL108F12	Human DNA sequence fro 4.62e+02
c	603	14	11.8	35863	26	AB000381	Human DNA for GPI-anch. 4.62e+02

	604	14	11.8	35938	18	CELR13A1	Caenorhabditis elegans 4.62e+02
c	605	14	11.8	36089	18	CELC28C12	Caenorhabditis elegans 4.62e+02
	606	14	11.8	36149	24	SPBC12C2	S.pombe chromosome II 4.62e+02
	607	14	11.8	36200	26	AC004792	Homo sapiens chromosom 4.62e+02
c	608	14	11.8	36337	26	HSTCRBB27	Homo sapiens T cell re 4.62e+02
	609	14	11.8	36437	25	HSU138B7	Human DNA sequence fro 4.62e+02
	610	14	11.8	36849	23	SCX11RA	S.cerevisiae DNA from 4.62e+02
c	611	14	11.8	36864	18	CELT03D3	Caenorhabditis elegans 4.62e+02
	612	14	11.8	36874	26	AC005336	Homo sapien chromosome 4.62e+02
	613	14	11.8	37082	25	HSU230B10	Human DNA sequence fro 4.62e+02
c	614	14	11.8	37394	23	SPAC2F3	S.pombe chromosome I c 4.62e+02
c	615	14	11.8	37634	18	CELK07D4	Caenorhabditis elegans 4.62e+02
c	616	14	11.8	37681	18	CEK05C4	Caenorhabditis elegans 4.62e+02
c	617	14	11.8	37730	23	YSCL8543	Saccharomyces cerevisi 4.62e+02
c	618	14	11.8	37751	16	MTCY1A6	Mycobacterium tubercul 4.62e+02
c	619	14	11.8	37752	25	HS361H4C	Human DNA sequence fro 4.62e+02
c	620	14	11.8	37800	23	SCCXIV38K	S.cerevisiae 37,800 bp 4.62e+02
c	621	14	11.8	37972	25	HSU51281	Human chromosome 11 co 4.62e+02
	622	14	11.8	37990	18	CEC17G1	Caenorhabditis elegans 4.62e+02
	623	14	11.8	38000	18	CET16G12	Caenorhabditis elegans 4.62e+02
c	624	14	11.8	38012	24	SPBC2F12	S.pombe chromosome II 4.62e+02
c	625	14	11.8	38139	23	SPAC10F6	S.pombe chromosome I c 4.62e+02
	626	14	11.8	38277	28	AC003993	Mouse Cosmid ma03m136 4.62e+02
	627	14	11.8	38400	27	AC004058	Homo sapiens chromosom 4.62e+02
c	628	14	11.8	38413	27	AC005601	Homo sapiens chromosom 4.62e+02
	629	14	11.8	38468	25	HUMN73F4	Human cosmid N73F4, co 4.62e+02
	630	14	11.8	38582	18	CELK12H4	C. elegans cosmid K12H 4.62e+02
	631	14	11.8	38610	18	CELC18H7	Caenorhabditis elegans 4.62e+02
c	632	14	11.8	38651	26	AC002306	Homo sapiens DNA from 4.62e+02
	633	14	11.8	38799	18	CEK04D7	Caenorhabditis elegans 4.62e+02
c	634	14	11.8	39030	18	CEF41D3	Caenorhabditis elegans 4.62e+02
	635	14	11.8	39118	17	AC004186	*** SEQUENCING IN PROG 4.62e+02
	636	14	11.8	39184	17	CEF12F11	*** SEQUENCING IN PROG 4.62e+02
	637	14	11.8	39190	25	HSUB2206	Human chromosome 10 cl 4.62e+02
c	638	14	11.8	39491	28	AC004101	Mouse Cosmid ma53a016 4.62e+02
c	639	14	11.8	39599	17	HSE140F8	Human DNA sequence *** 4.62e+02
	640	14	11.8	39754	26	AC000089	Homo sapiens Chromosom 4.62e+02
c	641	14	11.8	39825	27	HSU61375	Human cosmid LL12NC01- 4.62e+02
	642	14	11.8	39848	18	CEC49F5	Caenorhabditis elegans 4.62e+02
	643	14	11.8	39973	18	CEM10G6	Caenorhabditis elegans 4.62e+02
c	644	14	11.8	40176	27	HSU58675	Human olfactory recept 4.62e+02
c	645	14	11.8	40242	23	SPAC1F5	S. pombe chromosome I 4.62e+02
c	646	14	11.8	40460	27	AC005492	Homo sapiens clone UWG 4.62e+02
	647	14	11.8	40622	26	AC004199	Homo sapiens clone UWG 4.62e+02
c	648	14	11.8	40978	18	CELC09B7	Caenorhabditis elegans 4.62e+02
c	649	14	11.8	41219	26	AC005382	Homo sapiens chromosom 4.62e+02
c	650	14	11.8	41397	18	CEB02B5	Caenorhabditis elegans 4.62e+02
	651	14	11.8	41581	26	AC004711	Homo sapiens clone gl5 4.62e+02
c	652	14	11.8	41624	27	AC005680	Homo sapiens clone UWG 4.62e+02
c	653	14	11.8	41830	28	AC005240	Mus musculus clone UWG 4.62e+02
c	654	14	11.8	41883	18	CELD2096	Caenorhabditis elegans 4.62e+02

	655	14	11.8	41936	26	CH19R30879	Homo sapiens DNA from 4.62e+02
	656	14	11.8	42007	27	AC004200	Homo sapiens clone UWG 4.62e+02
	657	14	11.8	42398	17	AC003095	*** SEQUENCING IN PROG 4.62e+02
c	658	14	11.8	42734	18	CEDY3	Caenorhabditis elegans 4.62e+02
c	659	14	11.8	43843	27	AC005796	Homo sapiens chromosom 4.62e+02
c	660	14	11.8	44019	25	HSU160A4	Human DNA sequence fro 4.62e+02
	661	14	11.8	44195	18	CELC15B12	Caenorhabditis elegans 4.62e+02
c	662	14	11.8	44298	18	CELC07G1	Caenorhabditis elegans 4.62e+02
	663	14	11.8	44332	27	AC005794	Homo sapiens chromosom 4.62e+02
	664	14	11.8	45022	18	CEC34D1	Caenorhabditis elegans 4.62e+02
	665	14	11.8	45298	26	AC005404	Homo sapiens clone UWG 4.62e+02
c	666	14	11.8	45435	17	HSLUCA7	Human DNA sequence *** 4.62e+02
c	667	14	11.8	45531	18	CELC07A12	Caenorhabditis elegans 4.62e+02
	668	14	11.8	45635	18	CELT23E7	Caenorhabditis elegans 4.62e+02
	669	14	11.8	45687	26	AC004634	Homo sapiens chromosom 4.62e+02
	670	14	11.8	45692	26	HS104M094	Human DNA sequence fro 4.62e+02
c	671	14	11.8	45861	18	DMC80H7	Drosophila melanogaste 4.62e+02
c	672	14	11.8	46127	27	AC005547	Homo sapiens chromosom 4.62e+02
	673	14	11.8	47739	16	AF017113	Bacillus subtilis 300- 4.62e+02
c	674	14	11.8	47777	26	AC004189	Homo sapiens clone UWG 4.62e+02
	675	14	11.8	48000	17	AC003660	*** SEQUENCING IN PROG 4.62e+02
c	676	14	11.8	48338	18	CET02D9	Caenorhabditis elegans 4.62e+02
c	677	14	11.8	50978	18	AC002446	Drosophila melanogaste 4.62e+02
	678	14	11.8	51191	26	AC003951	Homo sapiens chromosom 4.62e+02
	679	14	11.8	54719	23	SCXV55KB	S.cerevisiae chromosom 4.62e+02
c	680	14	11.8	56128	26	AC005349	Homo sapiens chromosom 4.62e+02
c	681	14	11.8	56913	26	HS67C13	Human DNA sequence fro 4.62e+02
c	682	14	11.8	62288	27	AC005571	Homo sapiens chromosom 4.62e+02
	683	14	11.8	63726	26	HSAC002071	Human PAC clone DJ002C 4.62e+02
	684	14	11.8	64607	27	AC005021	Homo sapiens BAC clone 4.62e+02
c	685	14	11.8	64674	18	AC004247	Drosophila melanogaste 4.62e+02
	686	14	11.8	65409	17	HS36OB4	Human DNA sequence *** 4.62e+02
c	687	14	11.8	65691	18	PFMAL3P1	Plasmodium falciparum 4.62e+02
c	688	14	11.8	68620	18	AC004301	Drosophila melanogaste 4.62e+02
	689	14	11.8	69037	28	AC003949	Mus musculus chromosom 4.62e+02
c	690	14	11.8	71482	17	HSNS6E8	Human DNA sequence *** 4.62e+02
c	691	14	11.8	71664	26	AC003692	Human Chromosome 11 pa 4.62e+02
	692	14	11.8	72614	17	HS496N17	Human DNA sequence *** 4.62e+02
c	693	14	11.8	73087	26	HS147M19	Homo sapiens DNA seque 4.62e+02
	694	14	11.8	74507	26	AC004031	Homo sapiens 12q24 PAC 4.62e+02
	695	14	11.8	77137	18	AC004422	Drosophila melanogaste 4.62e+02
c	696	14	11.8	77363	23	AB006699	Arabidopsis thaliana g 4.62e+02
c	697	14	11.8	79611	26	AC002486	Human BAC clone RG367O 4.62e+02
c	698	14	11.8	79676	24	AB013396	Arabidopsis thaliana g 4.62e+02
c	699	14	11.8	79881	24	ATAC005314	Arabidopsis thaliana c 4.62e+02
	700	14	11.8	80315	23	AB005232	Arabidopsis thaliana g 4.62e+02
c	701	14	11.8	81370	24	AB017071	Arabidopsis thaliana g 4.62e+02
c	702	14	11.8	81837	26	AC004775	Homo sapiens chromosom 4.62e+02
c	703	14	11.8	82018	17	AC005430	*** SEQUENCING IN PROG 4.62e+02
c	704	14	11.8	82251	26	AC004506	Homo sapiens chromosom 4.62e+02
	705	14	11.8	82951	25	HS227P17	Human DNA sequence fro 4.62e+02

	706	14	11.8	82986	17	CEAA2	*** SEQUENCING IN PROG 4.62e+02
	707	14	11.8	83029	23	ATT8O5	Arabidopsis thaliana D 4.62e+02
c	708	14	11.8	83078	24	AF072897	Arabidopsis thaliana B 4.62e+02
	709	14	11.8	83079	26	AC004538	Homo sapiens PAC clone 4.62e+02
c	710	14	11.8	83340	18	AC004438	Drosophila melanogaste 4.62e+02
	711	14	11.8	83527	18	L49408	Drosophila melanogaste 4.62e+02
	712	14	11.8	83650	17	AC004311	*** SEQUENCING IN PROG 4.62e+02
	713	14	11.8	84055	18	AC004439	Drosophila melanogaste 4.62e+02
c	714	14	11.8	85590	26	HS453D15	Human DNA sequence fro 4.62e+02
	715	14	11.8	85616	17	AC005113	*** SEQUENCING IN PROG 4.62e+02
c	716	14	11.8	85654	26	HSP373C6	Homo sapiens DNA seque 4.62e+02
c	717	14	11.8	85855	23	AC002311	Arabidopsis thaliana c 4.62e+02
c	718	14	11.8	86064	24	AB013395	Arabidopsis thaliana g 4.62e+02
	719	14	11.8	87114	26	AC005353	Homo sapiens chromosom 4.62e+02
c	720	14	11.8	87246	26	AC002539	Homo sapiens chromosom 4.62e+02
c	721	14	11.8	87552	17	HS217O16	Human DNA sequence *** 4.62e+02
c	722	14	11.8	87585	18	AC004299	Drosophila melanogaste 4.62e+02
c	723	14	11.8	87634	17	CEK01F6	*** SEQUENCING IN PROG 4.62e+02
	724	14	11.8	88068	23	ATAC002337	Arabidopsis thaliana c 4.62e+02
	725	14	11.8	88398	17	AB018117	*** SEQUENCING IN PROG 4.62e+02
c	726	14	11.8	88871	28	MMU58105	Mus musculus Btk locus 4.62e+02
c	727	14	11.8	89015	26	HS86C11	Homo sapiens DNA seque 4.62e+02
c	728	14	11.8	89713	24	AC004809	Arabidopsis thaliana c 4.62e+02
	729	14	11.8	89990	26	AC005303	Homo sapiens chromosom 4.62e+02
	730	14	11.8	91184	17	HS633O19	Human DNA sequence *** 4.62e+02
	731	14	11.8	91539	24	ATAC004261	Arabidopsis thaliana c 4.62e+02
	732	14	11.8	91854	23	ATAC003680	Arabidopsis thaliana c 4.62e+02
c	733	14	11.8	92948	23	ATAF000657	Arabidopsis thaliana B 4.62e+02
	734	14	11.8	92975	26	AC002369	Homo sapiens Xp22 PAC 4.62e+02
	735	14	11.8	93489	24	AC005359	REVERSE-COMPLEMENT OF: 4.62e+02
c	736	14	11.8	93911	17	HSAC000002	*** SEQUENCING IN PROG 4.62e+02
	737	14	11.8	94386	26	HS452M16	Human DNA sequence *** 4.62e+02
c	738	14	11.8	94665	17	AC004979	*** SEQUENCING IN PROG 4.62e+02
c	739	14	11.8	95643	24	P6N15	Arabidopsis thaliana B 4.62e+02
	740	14	11.8	96217	26	AC002410	Human BAC clone RG264L 4.62e+02
c	741	14	11.8	96964	17	AC004906	*** SEQUENCING IN PROG 4.62e+02
	742	14	11.8	97847	26	HS874C20	Homo sapiens DNA seque 4.62e+02
c	743	14	11.8	97916	27	HS408N23	Human DNA sequence fro 4.62e+02
	744	14	11.8	98274	27	HS714B7	Human DNA sequence fro 4.62e+02
c	745	14	11.8	98471	17	AC005692	*** SEQUENCING IN PROG 4.62e+02
c	746	14	11.8	98527	23	AC004044	Arabidopsis thaliana B 4.62e+02
	747	14	11.8	98822	23	ATT6K21	Arabidopsis thaliana D 4.62e+02
	748	14	11.8	99033	27	AC005704	Homo sapiens Xp22 bins 4.62e+02
c	749	14	11.8	100000	26	AP000040	Homo sapiens genomic D 4.62e+02
c	750	14	11.8	100753	17	AC003102	*** SEQUENCING IN PROG 4.62e+02
	751	14	11.8	100783	17	HS34417	Human DNA sequence *** 4.62e+02
	752	14	11.8	100793	26	AP007717	Homo sapiens BAC clone 4.62e+02
	753	14	11.8	100849	25	HUMNEUROF	Human oligodendrocyte 4.62e+02
c	754	14	11.8	101072	27	AC005752	Homo sapiens chromosom 4.62e+02
c	755	14	11.8	101370	26	AC003962	Homo sapiens chromosom 4.62e+02
	756	14	11.8	101486	17	AC005228	*** SEQUENCING IN PROG 4.62e+02

	757	14	11.8	101817	27	HS1216H12	Homo sapiens DNA seque 4.62e+02
	758	14	11.8	102064	26	AC002558	Homo sapiens chromosom 4.62e+02
c	759	14	11.8	103319	26	AC004217	Homo sapiens 12q24.1 P 4.62e+02
c	760	14	11.8	104810	26	AC003012	Human PAC clone DJ0169 4.62e+02
	761	14	11.8	105174	17	HS171H20	Human DNA sequence *** 4.62e+02
	762	14	11.8	105232	27	HSUB5199	Homo sapiens BAC956, c 4.62e+02
c	763	14	11.8	108484	23	ATAC004238	Arabidopsis thaliana c 4.62e+02
c	764	14	11.8	109276	17	AC005281	*** SEQUENCING IN PROG 4.62e+02
c	765	14	11.8	110000	17	CEY81G3	Caenorhabditis elegans 4.62e+02
c	766	14	11.8	110000	17	CEY63D3	Caenorhabditis elegans 4.62e+02
c	767	14	11.8	110127	27	HS789O11	Human DNA sequence fro 4.62e+02
	768	14	11.8	110209	26	HUAC004533	Homo sapiens Chromosom 4.62e+02
c	769	14	11.8	110209	26	HUAC004533	Homo sapiens Chromosom 4.62e+02
c	770	14	11.8	111324	26	AC000377	Human Chromosome 11 BA 4.62e+02
	771	14	11.8	112863	17	HS482L3	Human DNA sequence *** 4.62e+02
c	772	14	11.8	114738	23	AC002130	The sequence of BAC F1 4.62e+02
c	773	14	11.8	115835	27	HS28C20	Human DNA sequence fro 4.62e+02
	774	14	11.8	116391	17	AC005588	*** SEQUENCING IN PROG 4.62e+02
c	775	14	11.8	116391	17	AC005588	*** SEQUENCING IN PROG 4.62e+02
	776	14	11.8	116552	27	AC004022	Homo sapiens BAC clone 4.62e+02
c	777	14	11.8	117101	26	AC004061	Homo sapiens chromosom 4.62e+02
c	778	14	11.8	117161	17	CEY43F11	Caenorhabditis elegans 4.62e+02
c	779	14	11.8	118308	17	AC005316	*** SEQUENCING IN PROG 4.62e+02
	780	14	11.8	118758	26	HS450C20	Human DNA sequence fro 4.62e+02
	781	14	11.8	119298	27	AC005608	Homo sapiens chromosom 4.62e+02
c	782	14	11.8	119307	26	AC004134	Homo sapiens chromosom 4.62e+02
	783	14	11.8	119704	23	OSCHLPLXX	O.sinensis complete ch 4.62e+02
c	784	14	11.8	120229	26	U90095	Human Chromosome 11p15 4.62e+02
c	785	14	11.8	120375	17	AC004442	*** SEQUENCING IN PROG 4.62e+02
c	786	14	11.8	121804	17	HS229A8	Human DNA sequence *** 4.62e+02
	787	14	11.8	122228	26	AC000003	Homo sapiens chromosom 4.62e+02
	788	14	11.8	122322	26	AC002448	Human PAC clone DJ0068 4.62e+02
c	789	14	11.8	122349	17	HS1168A5	Human DNA sequence *** 4.62e+02
	790	14	11.8	122459	17	HSAC000384	*** SEQUENCING IN PROG 4.62e+02
	791	14	11.8	123397	26	AC004808	Homo sapiens PAC clone 4.62e+02
c	792	14	11.8	123452	17	AC005079	*** SEQUENCING IN PROG 4.62e+02
c	793	14	11.8	123661	27	AC004518	Homo sapiens chromosom 4.62e+02
	794	14	11.8	124096	26	AC005247	Homo sapiens PAC clone 4.62e+02
	795	14	11.8	124321	26	HS417G15	Human DNA sequence fro 4.62e+02
	796	14	11.8	124351	27	AC005185	Homo sapiens Xp22 bins 4.62e+02
	797	14	11.8	124519	28	AC003019	Mus musculus Bac B6, c 4.62e+02
c	798	14	11.8	126263	26	HS127B14	Human DNA sequence fro 4.62e+02
c	799	14	11.8	126407	17	AC005080	*** SEQUENCING IN PROG 4.62e+02
c	800	14	11.8	126474	26	HS181N1	Human DNA sequence fro 4.62e+02
	801	14	11.8	126586	17	AC004915	*** SEQUENCING IN PROG 4.62e+02
	802	14	11.8	127176	17	AC005385	*** SEQUENCING IN PROG 4.62e+02
	803	14	11.8	127356	17	AC005063	*** SEQUENCING IN PROG 4.62e+02
	804	14	11.8	127917	26	HS257A7	Human DNA sequence fro 4.62e+02
c	805	14	11.8	128011	26	HS333E23	Human DNA sequence fro 4.62e+02
c	806	14	11.8	128169	26	HS849L7	Human DNA sequence fro 4.62e+02
c	807	14	11.8	128303	17	AC004983	*** SEQUENCING IN PROG 4.62e+02

c	808	14	11.8	128915	26	AC002416	Human Chromosome X, co 4.62e+02
c	809	14	11.8	129577	26	AC004659	Homo sapiens chromosom 4.62e+02
c	810	14	11.8	130064	17	AC002509	*** SEQUENCING IN PROG 4.62e+02
	811	14	11.8	130626	26	HS1114G22	Human DNA sequence fro 4.62e+02
	812	14	11.8	131342	17	AC004883	*** SEQUENCING IN PROG 4.62e+02
c	813	14	11.8	131458	26	AC002386	Human BAC clone RG343H 4.62e+02
c	814	14	11.8	131855	17	AC002518	*** SEQUENCING IN PROG 4.62e+02
	815	14	11.8	132493	26	AC004583	Human Chromosome 15q11 4.62e+02
	816	14	11.8	132698	17	AC005051	*** SEQUENCING IN PROG 4.62e+02
	817	14	11.8	132783	26	AC000378	Human Chromosome 11 pa 4.62e+02
c	818	14	11.8	132826	17	AC005627	*** SEQUENCING IN PROG 4.62e+02
c	819	14	11.8	132981	26	HS82J11	Human DNA sequence fro 4.62e+02
	820	14	11.8	133230	17	AC004994	*** SEQUENCING IN PROG 4.62e+02
c	821	14	11.8	133870	17	AC004861	*** SEQUENCING IN PROG 4.62e+02
c	822	14	11.8	134416	22	AF020713	Bacteriophage SPBc2 co 4.62e+02
c	823	14	11.8	134474	26	HS33F8	Human DNA sequence fro 4.62e+02
	824	14	11.8	134506	26	HS357T16	Homo sapiens DNA seque 4.62e+02
c	825	14	11.8	134898	17	HS11O7	Human DNA sequence *** 4.62e+02
c	826	14	11.8	134904	27	AC004879	Homo sapiens PAC clone 4.62e+02
c	827	14	11.8	135007	26	HS423N22	Homo sapiens DNA seque 4.62e+02
c	828	14	11.8	135289	17	AC004961	*** SEQUENCING IN PROG 4.62e+02
c	829	14	11.8	135886	26	HSXIST2	Human Xq13 3' end of P 4.62e+02
c	830	14	11.8	136222	27	AC004099	Homo sapiens chromosom 4.62e+02
	831	14	11.8	136746	17	AC004928	*** SEQUENCING IN PROG 4.62e+02
c	832	14	11.8	136746	17	AC004928	*** SEQUENCING IN PROG 4.62e+02
c	833	14	11.8	136805	17	AC005507	*** SEQUENCING IN PROG 4.62e+02
c	834	14	11.8	136975	26	HSTCRB75A	Homo sapiens T cell re 4.62e+02
c	835	14	11.8	137111	26	HS20J23	Human DNA sequence *** 4.62e+02
	836	14	11.8	137496	27	AC005288	Homo sapiens chromosom 4.62e+02
c	837	14	11.8	137728	17	AC004860	*** SEQUENCING IN PROG 4.62e+02
	838	14	11.8	138201	17	AC005016	*** SEQUENCING IN PROG 4.62e+02
c	839	14	11.8	138201	17	AC005016	*** SEQUENCING IN PROG 4.62e+02
c	840	14	11.8	139097	17	AC004591	*** SEQUENCING IN PROG 4.62e+02
	841	14	11.8	139316	24	ATP7H19	Arabidopsis thaliana D 4.62e+02
	842	14	11.8	139350	26	HS443E24	Human DNA sequence fro 4.62e+02
c	843	14	11.8	139350	26	HS443E24	Human DNA sequence fro 4.62e+02
	844	14	11.8	139389	26	HS68O2	Homo sapiens DNA seque 4.62e+02
	845	14	11.8	139488	26	AC003043	Homo sapiens chromosom 4.62e+02
	846	14	11.8	141132	17	HS192P9	Human DNA sequence *** 4.62e+02
	847	14	11.8	141916	17	AC004908	*** SEQUENCING IN PROG 4.62e+02
	848	14	11.8	142761	17	HS566G5	Human DNA sequence *** 4.62e+02
c	849	14	11.8	143092	18	CEY17G7B	Caenorhabditis elegans 4.62e+02
c	850	14	11.8	143427	26	AC002432	Human BAC clone RG317G 4.62e+02
	851	14	11.8	143747	26	HS370M22	Human DNA sequence fro 4.62e+02
	852	14	11.8	144227	26	HS119E23	Homo sapiens DNA seque 4.62e+02
c	853	14	11.8	144239	26	AC004015	Human BAC clone RG333F 4.62e+02
c	854	14	11.8	144368	26	AC004103	Homo sapiens Xp22 BAC 4.62e+02
	855	14	11.8	144474	26	AF051934	Homo sapiens chromosom 4.62e+02
c	856	14	11.8	144518	17	AC004914	*** SEQUENCING IN PROG 4.62e+02
	857	14	11.8	144545	17	AC005287	*** SEQUENCING IN PROG 4.62e+02
	858	14	11.8	145844	26	AC004931	Homo sapiens PAC clone 4.62e+02

	859	14	11.8	147102	26	AC002476	Human PAC clone DJ318C 4.62e+02
c	860	14	11.8	147394	27	HS353H6	Human DNA sequence fro 4.62e+02
	861	14	11.8	148540	26	HS212P9	Human DNA sequence fro 4.62e+02
	862	14	11.8	149298	26	AP000031	Homo sapiens genomic D 4.62e+02
c	863	14	11.8	149571	17	CEY43C5	Caenorhabditis elegans 4.62e+02
c	864	14	11.8	149687	27	AC004981	Homo sapiens PAC clone 4.62e+02
c	865	14	11.8	150158	17	HS1147O16	Human DNA sequence *** 4.62e+02
	866	14	11.8	150290	26	HS509L4	Human DNA sequence fro 4.62e+02
	867	14	11.8	151551	17	AC005481	*** SEQUENCING IN PROG 4.62e+02
c	868	14	11.8	152077	26	HS130G2	Human DNA sequence fro 4.62e+02
	869	14	11.8	152261	33	HE1CG	Herpes simplex virus (4.62e+02
c	870	14	11.8	152550	17	HSJ181N11	*** SEQUENCING IN PROG 4.62e+02
c	871	14	11.8	153019	17	AC004969	*** SEQUENCING IN PROG 4.62e+02
c	872	14	11.8	153578	26	AC004074	Homo Sapiens Chromosom 4.62e+02
c	873	14	11.8	153735	26	AC004029	Human BAC clone RG035E 4.62e+02
	874	14	11.8	154066	26	HSB24O18	Human DNA sequence *** 4.62e+02
	875	14	11.8	154252	26	AC004062	Homo sapiens chromosom 4.62e+02
c	876	14	11.8	154252	26	AC004062	Homo sapiens chromosom 4.62e+02
c	877	14	11.8	155074	27	AC003693	Human Chromosome 11p15 4.62e+02
c	878	14	11.8	155147	26	HS992D9	Human DNA sequence fro 4.62e+02
	879	14	11.8	156214	26	HSAC002064	Human BAC clone RG016J 4.62e+02
c	880	14	11.8	156432	26	HS581P7	Human DNA sequence fro 4.62e+02
c	881	14	11.8	159446	26	AC003669	Homo sapiens Xp22 BAC 4.62e+02
	882	14	11.8	159863	26	AC002349	Homo sapiens Xp22 PAC 4.62e+02
	883	14	11.8	160178	17	HS422G23	Human DNA sequence *** 4.62e+02
c	884	14	11.8	160262	26	AC005242	Homo sapiens chromosom 4.62e+02
	885	14	11.8	161742	17	AC005519	*** SEQUENCING IN PROG 4.62e+02
	886	14	11.8	161824	17	AC005061	*** SEQUENCING IN PROG 4.62e+02
	887	14	11.8	162067	17	AC005483	*** SEQUENCING IN PROG 4.62e+02
	888	14	11.8	162075	26	HS127D3	Homo sapiens DNA seque 4.62e+02
	889	14	11.8	162455	17	AC005536	*** SEQUENCING IN PROG 4.62e+02
	890	14	11.8	163023	17	AC004815	*** SEQUENCING IN PROG 4.62e+02
c	891	14	11.8	163974	27	AB011399	Homo sapiens gene for 4.62e+02
	892	14	11.8	164017	26	AC005304	Homo sapiens chromosom 4.62e+02
	893	14	11.8	164049	17	AC005232	*** SEQUENCING IN PROG 4.62e+02
c	894	14	11.8	164277	17	HS1052M9	Human DNA sequence *** 4.62e+02
	895	14	11.8	165228	26	AC005325	Homo sapiens chromosom 4.62e+02
c	896	14	11.8	165363	17	HS459L4	Human DNA sequence *** 4.62e+02
	897	14	11.8	165659	26	AF003529	Homo sapiens glypican 4.62e+02
	898	14	11.8	166687	27	AC005224	Homo sapiens chromosom 4.62e+02
c	899	14	11.8	166832	26	AF070718	Homo sapiens BAC clone 4.62e+02
c	900	14	11.8	167234	24	AC004708	CIC5B11.1 check: 4870 4.62e+02
c	901	14	11.8	167372	27	AC004945	Homo sapiens PAC clone 4.62e+02
	902	14	11.8	167575	17	HS531H16	Human DNA sequence *** 4.62e+02
	903	14	11.8	170136	26	AC003082	Human BAC clone RG011J 4.62e+02
c	904	14	11.8	170136	26	AC003082	Human BAC clone RG011J 4.62e+02
c	905	14	11.8	170392	17	HS740A11	Human DNA sequence *** 4.62e+02
c	906	14	11.8	170658	26	AC004072	Human Chromosome X clo 4.62e+02
	907	14	11.8	171101	17	AC005666	*** SEQUENCING IN PROG 4.62e+02
	908	14	11.8	171147	17	HS1059H15	Human DNA sequence *** 4.62e+02
c	909	14	11.8	171368	26	HSU95738	Human chromosome 16p13 4.62e+02

	910	14	11.8	171681	26	HUAC002299	Homo sapiens Chromosom 4.62e+02
c	911	14	11.8	171681	26	HUAC002299	Homo sapiens Chromosom 4.62e+02
	912	14	11.8	171905	26	HSAC002080	Human BAC clone RG253B 4.62e+02
	913	14	11.8	172082	17	HS998H6	Human DNA sequence *** 4.62e+02
	914	14	11.8	172984	26	HUU91318	Human chromosome 16 BA 4.62e+02
c	915	14	11.8	172995	17	AC004926	*** SEQUENCING IN PROG 4.62e+02
	916	14	11.8	173494	17	HS662N3	Human DNA sequence *** 4.62e+02
c	917	14	11.8	173717	27	AC005695	Homo sapiens chromosom 4.62e+02
c	918	14	11.8	174826	26	HUU91324	Homo sapiens Chromosom 4.62e+02
c	919	14	11.8	175162	17	AC005182	*** SEQUENCING IN PROG 4.62e+02
c	920	14	11.8	175488	17	AC004887	*** SEQUENCING IN PROG 4.62e+02
c	921	14	11.8	176011	17	AC005073	*** SEQUENCING IN PROG 4.62e+02
c	922	14	11.8	176154	17	AC005317	*** SEQUENCING IN PROG 4.62e+02
c	923	14	11.8	177163	26	HS134E15	Human DNA sequence *** 4.62e+02
c	924	14	11.8	177568	27	HS388M5	Human DNA sequence fro 4.62e+02
c	925	14	11.8	177756	17	AC005020	*** SEQUENCING IN PROG 4.62e+02
c	926	14	11.8	177821	17	AC005030	*** SEQUENCING IN PROG 4.62e+02
c	927	14	11.8	178440	17	CEY38F1	Caenorhabditis elegans 4.62e+02
c	928	14	11.8	179174	17	AC004796	*** SEQUENCING IN PROG 4.62e+02
c	929	14	11.8	179246	17	AC004854	*** SEQUENCING IN PROG 4.62e+02
	930	14	11.8	179531	26	AC003080	Human BAC clone GS368F 4.62e+02
c	931	14	11.8	179531	26	AC003080	Human BAC clone GS368F 4.62e+02
c	932	14	11.8	179544	26	AF015720	Homo sapiens chromosom 4.62e+02
	933	14	11.8	179707	17	HSAC001234	*** SEQUENCING IN PROG 4.62e+02
	934	14	11.8	179854	17	CEY19D2	Caenorhabditis elegans 4.62e+02
c	935	14	11.8	180551	26	HUAC004158	Homo sapiens Chromosom 4.62e+02
	936	14	11.8	180919	27	AF088219	Homo sapiens CC chemok 4.62e+02
	937	14	11.8	181331	17	AC005230	*** SEQUENCING IN PROG 4.62e+02
c	938	14	11.8	181331	17	AC005230	*** SEQUENCING IN PROG 4.62e+02
	939	14	11.8	184130	26	AC005209	Homo sapiens chromosom 4.62e+02
	940	14	11.8	184490	26	HSUB2828	Homo sapiens ataxia te 4.62e+02
	941	14	11.8	184706	26	AC004782	Homo sapiens chromosom 4.62e+02
c	942	14	11.8	186510	26	HS451B15	Human DNA sequence fro 4.62e+02
	943	14	11.8	186780	27	AC005740	Homo sapiens chromosom 4.62e+02
	944	14	11.8	187253	17	AC005486	*** SEQUENCING IN PROG 4.62e+02
c	945	14	11.8	188362	27	AC005019	Homo sapiens BAC clone 4.62e+02
c	946	14	11.8	188471	17	AC005540	*** SEQUENCING IN PROG 4.62e+02
	947	14	11.8	188629	17	AC004816	*** SEQUENCING IN PROG 4.62e+02
	948	14	11.8	188636	26	HUAC002287	Homo sapiens Chromosom 4.62e+02
	949	14	11.8	189215	17	AC005675	*** SEQUENCING IN PROG 4.62e+02
c	950	14	11.8	190000	17	AC005598	*** SEQUENCING IN PROG 4.62e+02
	951	14	11.8	191191	25	HSAC000119	Human BAC clone RG104I 4.62e+02
	952	14	11.8	191423	17	AC004941	*** SEQUENCING IN PROG 4.62e+02
	953	14	11.8	195016	17	HS215P16	Human DNA sequence *** 4.62e+02
	954	14	11.8	197391	17	AC005140	*** SEQUENCING IN PROG 4.62e+02
	955	14	11.8	197806	17	HS80N2	Human DNA sequence *** 4.62e+02
	956	14	11.8	198499	17	AC004870	*** SEQUENCING IN PROG 4.62e+02
	957	14	11.8	198582	26	AC005291	Homo sapiens chromosom 4.62e+02
c	958	14	11.8	199927	17	HS101A4	Human DNA sequence *** 4.62e+02
	959	14	11.8	200000	17	AC004618	*** SEQUENCING IN PROG 4.62e+02
	960	14	11.8	204082	27	AC005323	Homo sapiens chromosom 4.62e+02

	961	14	11.8	205559	23	ATFCA4	Arabidopsis thaliana D	4.62e+02
c	962	14	11.8	206214	17	AC005013	*** SEQUENCING IN PROG	4.62e+02
c	963	14	11.8	207139	17	CEY17G7	Caenorhabditis elegans	4.62e+02
	964	14	11.8	208230	15	BSUB0008	Bacillus subtilis comp	4.62e+02
c	965	14	11.8	208230	15	BSUB0008	Bacillus subtilis comp	4.62e+02
c	966	14	11.8	210000	17	AC005690	*** SEQUENCING IN PROG	4.62e+02
c	967	14	11.8	212610	15	BSUB0019	Bacillus subtilis comp	4.62e+02
c	968	14	11.8	213541	27	HUAC004381	Homo sapiens Chromosom	4.62e+02
	969	14	11.8	213680	15	BSUB0012	Bacillus subtilis comp	4.62e+02
c	970	14	11.8	215286	26	HSU91322	Human chromosome 16p13	4.62e+02
c	971	14	11.8	215807	17	AC005296	*** SEQUENCING IN PROG	4.62e+02
c	972	14	11.8	217873	26	HUAC004382	Homo sapiens Chromosom	4.62e+02
c	973	14	11.8	221160	15	BSUB0002	Bacillus subtilis comp	4.62e+02
	974	14	11.8	224064	17	AC005055	*** SEQUENCING IN PROG	4.62e+02
c	975	14	11.8	225432	26	AF027390	Homo sapiens 7q telome	4.62e+02
c	976	14	11.8	227571	17	AC005098	*** SEQUENCING IN PROG	4.62e+02
c	977	14	11.8	228952	17	HS250D10	Human DNA sequence ***	4.62e+02
c	978	14	11.8	230023	17	HS522P13	Human DNA sequence ***	4.62e+02
c	979	14	11.8	238438	17	AC005046	*** SEQUENCING IN PROG	4.62e+02
	980	14	11.8	240114	17	HS90L6	Human DNA sequence ***	4.62e+02
	981	14	11.8	244254	26	HSAC001228	244Kb Contig from Huma	4.62e+02
	982	14	11.8	244362	17	HS618F1	Human DNA sequence ***	4.62e+02
c	983	14	11.8	251124	26	HUAE000660	Homo sapiens T-cell re	4.62e+02
	984	14	11.8	251199	26	HUAE000659	Homo sapiens T-cell.re	4.62e+02
c	985	14	11.8	254200	17	CEY18D10	Caenorhabditis elegans	4.62e+02
	986	14	11.8	256404	17	AC005071	*** SEQUENCING IN PROG	4.62e+02
c	987	14	11.8	263313	17	CEY48C3	Caenorhabditis elegans	4.62e+02
c	988	14	11.8	268399	17	AC004469	*** SEQUENCING IN PROG	4.62e+02
c	989	14	11.8	269078	17	CEY76A2	Caenorhabditis elegans	4.62e+02
c	990	14	11.8	275138	17	CEY54E5	Caenorhabditis elegans	4.62e+02
	991	14	11.8	281014	17	HS58A9	Human DNA sequence ***	4.62e+02
c	992	14	11.8	292720	17	CEY47H10	Caenorhabditis elegans	4.62e+02
c	993	14	11.8	302988	17	CEY46G5	Caenorhabditis elegans	4.62e+02
	994	14	11.8	310213	17	AC004526	*** SEQUENCING IN PROG	4.62e+02
c	995	14	11.8	313573	18	CEY57G11C	Caenorhabditis elegans	4.62e+02
	996	14	11.8	315339	23	SCCHRIII	S.cerevisiae chromosom	4.62e+02
	997	14	11.8	322746	17	HS15D23	Human DNA sequence ***	4.62e+02
c	998	14	11.8	323000	26	HS229041	Homo sapiens 959 kb co	4.62e+02
c	999	14	11.8	326569	17	AC004912	*** SEQUENCING IN PROG	4.62e+02
	1000	14	11.8	348050	26	HS229042	Homo sapiens 959 kb co	4.62e+02

ALIGNMENTS

```
RESULT    1
LOCUS       S76771       6849 bp    DNA             PRI       26-JUL-1995
DEFINITION  TPO=thrombopoietin [human, Genomic, 6849 nt].
ACCESSION   S76771
NID         g914225
KEYWORDS    .
SOURCE      human.
```



```
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 6849)
  AUTHORS   Gurney,A.L., Kuang,W.J., Xie,M.H., Malloy,B.E., Eaton,D.L. and de
            Sauvage,F.J.
  TITLE     Genomic structure, chromosomal localization, and conserved
            alternative splice forms of thrombopoietin
  JOURNAL   Blood 85 (4), 981-988 (1995)
  MEDLINE   95152076
  REMARK    GenBank staff at the National Library of Medicine created this
            entry [NCBI gibbsq 163543] from the original journal article.
            This sequence comes from Fig. 5B.
FEATURES             Location/Qualifiers
     source          1. .6849
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     mRNA            join(166. .232,2046. .2058,2290. .2417,2704. .2790,4723. .4890,
                     5127. .6289)
     gene            2046. .5792
                     /note="thrombopoietin, TPO"
                     /gene="TPO"
     CDS             join(2046. .2058,2290. .2417,2704. .2790,4723. .4890,
                     5127. .5792)
                     /gene="TPO"
                     /note="This sequence comes from Fig. 5B."
                     /codon_start=1
                     /product="thrombopoietin"
                     /db_xref="PID:g914226"
                     /translation="MELTELLLVVMLLLTARLTLSSPAPPACDLRVLSKLLRDSHVLH
                     SRLSQCPEVHPLPTPVLLPAVDFSLGEWKTQMEETKAQDILGAVTLLLEGVMAARGQL
                     GPTCLSSLLGQLSGQVRLLLGALQSLLGTQLPPQGRTTAHKDPNAIFLSFQHLLRGKV
                     RFLMLVGGSTLCVRRAPPTTAVPSRTSLVLTLNELPNRTSGLLETNFTASARTTGSGL
                     LKWQQGFRAKIPGLLNQTSRSLDQIPGYLNRIHELLNGTRGLFPGPSRRTLGAPDISS
                     GTSDTGSLPPNLQPGYSPSPTHPPTGQYTLFPLPPTLPTPVVQLHPLLPDPSAPTPTP
                     TSPLLNTSYTHSQNLSQEG"
BASE COUNT     1727 a    1871 c    1616 g    1634 t       1 others
ORIGIN

  Query Match           99.2%;  Score 118;  DB 25;  Length 6849;
  Best Local Similarity  99.2%;  Pred. No. 1.80e-111;
  Matches   118;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db    6252 GATCTTTACTCTTGAGAAATGAATAAGCTTTCTCTCAGAAATGCTGTCCCTATACACTAG 6311
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy       1 GATCTTTACTCTTGAGAAATGAATAAGCTTTCTCTCAGAAATGCTGTCCCTATACACTAG 60

Db    6312 ACAAAACTGAGCCTGTATAAGGAATAAATGGGAGCGCCGAAAAGCTCCCTAAAAAGCAA 6370
           ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Qy      61 ACAAAACTGAGCCTGTATAAGGAATAAATGGGAGCGCCGAAAAGCTCCNTAAAAAGCAA 119
```



```
RESULT    2
LOCUS       HUMTA        7666 bp    DNA             PRI       05-MAR-1997
DEFINITION  Human gene for thrombopoietin, exon1-exon6, complete cds.
ACCESSION   D32046
NID         g577319
KEYWORDS    thrombopoietin.
SOURCE      Homo sapiens blood DNA, clone:lambdaHGT1.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 7666)
  AUTHORS   Sohma,Y., Akahori,H., Seki,N., Hori,T., Ogami,K., Kato,T.,
            Shimada,Y., Kawamura,K. and Miyazaki,H.
  TITLE     Molecular cloning and chromosomal localization of the human
            thrombopoietin gene
  JOURNAL   FEBS Lett. 353 (1), 57-61 (1994)
  MEDLINE   95010765
REFERENCE   2  (bases 1 to 7666)
  AUTHORS   Sohma,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JUL-1994) to the DDBJ/EMBL/GenBank databases.
            Yoshiaki Sohma, KIRIN Brewery Co.,Ltd. Pharmaceutical Research
            Laboratory; 2-2, Soujamachi 1chome, Maebashi, Gunma 371, Japan
            (Tel:0272-54-8618, Fax:0272-52-2307)
FEATURES             Location/Qualifiers
     source          1. .7666
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="lambdaHGT1"
                     /map="3q27"
                     /tissue_type="blood"
     exon            <1420. .1486
                     /number=1
     exon            3157. .3314
                     /number=2
     CDS             join(3302. .3314,3546. .3673,3960. .4046,5979. .6146,
                     6383. .7048)
                     /codon_start=1
                     /product="thrombopoietin"
                     /db_xref="PID:d1007381"
                     /db_xref="PID:g577320"
                     /translation="MELTELLLVVMLLLTARLTLSSPAPPACDLRVLSKLLRDSHVLH
                     SRLSQCPEVHPLPTPVLLPAVDFSLGEWKTQMEETKAQDILGAVTLLLEGVMAARGQL
                     GPTCLSSLLGQLSGQVRLLLGALQSLLGTQLPPQGRTTAHKDPNAIFLSFQHLLRGKV
                     RFLMLVGGSTLCVRRAPPTTAVPSRTSLVLTLNELPNRTSGLLETNFTASARTTGSGL
                     LKWQQGFRAKIPGLLNQTSRSLDQIPGYLNRIHELLNGTRGLFPGPSRRTLGAPDISS
                     GTSDTGSLPPNLQPGYSPSPTHPPTGQYTLFPLPPTLPTPVVQLHPLLPDPSAPTPTP
                     TSPLLNTSYTHSQNLSQEG"
```



```
     exon            3546. .3673
                     /number=3
     exon            3960. .4046
                     /number=4
     exon            5979. .6146
                     /number=5
     exon            6383. .7576
                     /number=6
BASE COUNT     1897 a    2154 c    1807 g    1808 t
ORIGIN

  Query Match           99.2%;  Score 118;  DB 26;  Length 7666;
  Best Local Similarity  99.2%;  Pred. No. 1.80e-111;
  Matches   118;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db    7508 GATCTTTACTCTTGAGAAATGAATAAGCTTTCTCTCAGAAATGCTGTCCCTATACACTAG 7567
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy       1 GATCTTTACTCTTGAGAAATGAATAAGCTTTCTCTCAGAAATGCTGTCCCTATACACTAG 60

Db    7568 ACAAAACTGAGCCTGTATAAGGAATAAATGGGAGCGCCGAAAAGCTCCCTAAAAAGCAA 7626
           ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Qy      61 ACAAAACTGAGCCTGTATAAGGAATAAATGGGAGCGCCGAAAAGCTCCNTAAAAAGCAA 119


RESULT    3
ID   E11961     standard; RNA; HUM; 1721 BP.
AC   E11961;
NI   d1110298
DT   08-OCT-1997 (Rel. 52, Created)
DT   08-OCT-1997 (Rel. 52, Last updated, Version 1)
DE   Human cDNA encoding thrombopoietin.
KW   JP 1996228781-A/2.
OS   Homo sapiens (human)
OC   Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates;
OC   Catarrhini; Hominidae; Homo.
RN   [1]
RP   1-1721
RA   Miyazaki H., Kato H., Ogami K., Iwamatsu A., Akahori H., Kuroki R.,
RA   Shinizu T., Muto T.;
RT   "DNA CODING FOR PROTEIN HAVING TPO ACTIVITY";
RL   Patent number JP 1996228781-A/2, 10-SEP-1996.
RL   KIRIN BREWERY CO LTD.
CC   OS   Homo sapiens (human)
CC   PN   JP 1996228781-A/2
CC   PD   10-SEP-1996
CC   PF   14-FEB-1995 JP 1995063298
CC   PR   14-FEB-1994 JP 94P    39090, 25-MAR-1994 JP 94P    79842,
CC   PR   01-JUN-1994 JP 94P   155126, 15-JUN-1994 JP 94P   167328,
CC   PR   17-AUG-1994 JP 94P   227159, 17-AUG-1994 JP 94P   193169,
CC   PR   01-NOV-1994 JP 94P   304167, 01-DEC-1994 JP 94P   298669,
CC   PR   28-DEC-1994 JP 94P   341200
```

```
CC   PI   MIYAZAKI HIROSHI, KATO HISASHI, OGAMI KINYA, IWAMATSU AKIHIKO,
CC   PI   AKAHORI HIRONORI, KUROKI RYOTA, SHIMIZU TOSHIYUKI,
CC   PI   MUTO TAKANORI
CC   PC   C12N15/09,A61K38/00,C07K14/52,C12N1/21,C12N5/10,C12P21/02,
CC   PC   (C12N1/21,
CC   PC   C12R1:19),(C12N5/10,C12R1:91),(C12P21/02,C12R1:19),(C12P21/02,
CC   PC   C12R1:91);
CC   CC   strandedness: Double;
CC   CC   topology: Linear;
CC   FH   Key             Location/Qualifiers
CC   FH
CC   FT   source          1. .1721
CC   FT                   /organism="Homo sapiens"
CC   FT                   /tissue_type="liver"
CC   FT   CDS             102. .1163
CC   FT                   /product="thrombopoietin precursor"
CC   FT   sig_peptide     102. .164
CC   FT   mat_peptide     165. .1160
CC   FT                   /product="thrombopoietin"
CC   FT   variation       replace(84,"c")
CC   FT   variation       replace(740,"t")
CC   FT   variation       replace(1198,"a")
FH   Key             Location/Qualifiers
FH
FT   source          1. .1721
FT                   /organism="Homo sapiens"
SQ   Sequence 1721 BP; 413 A; 542 C; 371 G; 395 T; 0 other;

  Query Match           58.8%;  Score 70;  DB 10;  Length 1721;
  Best Local Similarity 100.0%;  Pred. No. 4.93e-55;
  Matches   70;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1623 GATCTTTACTCTTGAGAAATGAATAAGCTTTCTCTCAGAAATGCTGTCCCTATACACTAG 1682
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      1 GATCTTTACTCTTGAGAAATGAATAAGCTTTCTCTCAGAAATGCTGTCCCTATACACTAG 60

Db   1683 ACAAAACTGA 1692
            ||||||||||
Qy     61 ACAAAACTGA 70


RESULT    4
LOCUS       D32047       1691 bp    mRNA            PRI       28-AUG-1997
DEFINITION  Homo sapiens mRNA for thrombopoietin, complete cds.
ACCESSION   D32047
NID         g2351117
KEYWORDS    thrombopoietin.
SOURCE      Homo sapiens liver cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
```



```
REFERENCE   1  (bases 1 to 1691)
  AUTHORS   Ogami,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JUL-1994) to the DDBJ/EMBL/GenBank databases. Kinya
            Ogami, KIRIN Brewery Co., Ltd., Pharmaceutical Research Laboratory;
            2-2 Soujamachi 1 chome, Maebashi, Gunma 371, Japan
            (Tel:0272-54-8618, Fax:0272-52-2307)
REFERENCE   2  (sites)
  AUTHORS   Kato,T., Ogami,K., Shimada,Y., Iwamatsu,A., Sohma,Y., Akahori,H.,
            Horie,K., Kokubo,A., Kudo,Y., Maeda,E., Kobayashi,K., Ohashi,H.,
            Ozawa,T., Inoue,H., Kawamura,K. and Miyazaki,H.
  TITLE     Purification and characterization of thrombopoietin
  JOURNAL   J. Biochem. 118 (1), 229-236 (1995)
  MEDLINE   96015174
  REMARK    Erratum:[[published erratum appears in J Biochem (Tokyo) 1996
            Jan;119(1):208]]
FEATURES             Location/Qualifiers
     source          1. .1691
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="liver"
     sig_peptide     102. .164
     CDS             102. .1163
                     /codon_start=1
                     /product="thrombopoietin"
                     /db_xref="PID:d1022789"
                     /db_xref="PID:g2351118"
                     /translation="MELTELLLVVMLLLTARLTLSSPAPPACDLRVLSKLLRDSHVLH
                     SRLSQCPEVHPLPTPVLLPAVDFSLGEWKTQMEETKAQDILGAVTLLLEGVMAARGQL
                     GPTCLSSLLGQLSGQVRLLLGALQSLLGTQLPPQGRTTAHKDPNAIPLSFQHLLRGKV
                     RFLMLVGGSTLCVRRAPPTTAVPSRTSLVLTLNELPNRTSGLLETNFTASARTTGSGL
                     LKWQQGFRAKIPGLLNQTSRSLDQIPGYLNRIHELLNGTRGLFPGPSRRTLGAPDISS
                     GTSDTGSLPPNLQPGYSPSPTHPPTGQYTLFPLPPTLPTPVVQLHPLLPDPSAPTPTP
                     TSPLLNTSYTHSQNLSQEG"
BASE COUNT      383 a    542 c    371 g    395 t
ORIGIN

  Query Match           58.0%;  Score 69;  DB 26;  Length 1691;
  Best Local Similarity 100.0%;  Pred. No. 6.87e-54;
  Matches   69;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1623 GATCTTTACTCTTGAGAAATGAATAAGCTTTCTCTCAGAAATGCTGTCCCTATACACTAG 1682
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      1 GATCTTTACTCTTGAGAAATGAATAAGCTTTCTCTCAGAAATGCTGTCCCTATACACTAG 60

Db   1683 ACAAAACTG 1691
            |||||||||
Qy     61 ACAAAACTG 69


RESULT    5
```



```
LOCUS       HUMMLCMPL    1795 bp    mRNA            PRI       07-JAN-1995
DEFINITION  Human c-mpl ligand (ML) mRNA, complete cds.
ACCESSION   L33410
NID         g506826
KEYWORDS    c-mpl ligand.
SOURCE      Homo sapiens cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1795)
  AUTHORS   de Sauvage,F.J., Hass,P.E., Spencer,S.D., Malloy,B.E., Gurney,A.L.,
            Spencer,S.A., Darbonne,W.C., Henzel,W.J., Wong,S.C., Kuang,W.-J.,
            Oles,K.J., Hultgren,B., Solberg,L.A.Jr. ., Goeddel,D.V. and
            Eaton,D.L.
  TITLE     Stimulation of megakaryocytopoiesis and thrombopoiesis by the c-Mpl
            ligand [see comments]
  JOURNAL   Nature 369 (6481), 533-538 (1994)
  MEDLINE   94261202
FEATURES             Location/Qualifiers
     source          1. .1795
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     sig_peptide     216. .278
                     /gene="ML"
     CDS             216. .1277
                     /gene="ML"
                     /note="erythropoietin homology domain  bp 66. .522"
                     /codon_start=1
                     /product="c-mpl ligand"
                     /db_xref="PID:g506827"
                     /translation="MELTELLLVVMLLLTARLTLSSPAPPACDLRVLSKLLRDSHVLH
                     SRLSQCPEVHPLPTPVLLPAVDFSLGEWKTQMEETKAQDILGAVTLLLEGVMAARGQL
                     GPTCLSSLLGQLSGQVRLLLGALQSLLGTQLPPQGRTTAHKDPNAIPLSPQHLLRGKV
                     RFLMLVGGSTLCVRRAPPTTAVPSRTSLVLTLNELPNRTSGLLETNFTASARTTGSGL
                     LKWQQGFRAKIPGLLNQTSRSLDQIPGYLNRIHELLNGTRGLFPGPSRRTLGAPDISS
                     GTSDTGSLPPNLQPGYSPSPTHPPTGQYTLFPLPPTLPTPVVQLHPLLPDPSAPTPTP
                     TSPLLNTSYTHSQNLSQEG"
     gene            216. .1277
                     /gene="ML"
     mat_peptide     279. .1274
                     /gene="ML"
                     /product="c-mpl ligand"
BASE COUNT      420 a    586 c    382 g    407 t
ORIGIN

  Query Match           34.5%;  Score 41;  DB 25;  Length 1795;
  Best Local Similarity 100.0%;  Pred. No. 6.60e-23;
  Matches   41;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1737 GATCTTTACTCTTGAGAAATGAATAAGCTTTCTCTCAGAAA 1777
            |||||||||||||||||||||||||||||||||||||||||
Qy      1 GATCTTTACTCTTGAGAAATGAATAAGCTTTCTCTCAGAAA 41
```



```


RESULT    6
LOCUS       I33525       1795 bp    DNA             PAT       27-JAN-1997
DEFINITION  Sequence 1 from patent US 5593666.
ACCESSION   I33525
NID         g1824316
KEYWORDS
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1795)
  AUTHORS   McDonald,T.P.
  TITLE     Methods and compositions for treating thrombocytopenia
  JOURNAL   Patent: US 5593666-A 1 14-JAN-1997;
FEATURES             Location/Qualifiers
     source          1. .1795
                     /organism="unknown"
BASE COUNT      420 a    586 c    382 g    407 t
ORIGIN

  Query Match           34.5%;  Score 41;  DB 21;  Length 1795;
  Best Local Similarity 100.0%;  Pred. No. 6.60e-23;
  Matches   41;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1737 GATCTTTACTCTTGAGAAATGAATAAGCTTTCTCTCAGAAA 1777
            |||||||||||||||||||||||||||||||||||||||||
Qy      1 GATCTTTACTCTTGAGAAATGAATAAGCTTTCTCTCAGAAA 41


RESULT    7
LOCUS       E12183       4506 bp    DNA             PAT       27-APR-1998
DEFINITION  Human thrombopoietin gene.
ACCESSION   E12183
NID         g1251017
KEYWORDS    JP 1996277296-A/3.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4506)
  AUTHORS   Miyazaki,H., Kato,H., Ogami,K., Iwamatsu,A., Akahori,H., Kuroki,R.,
            Shimizu,T. and Muto,T. .
  TITLE     PROTEIN HAVING TPO ACTIVITY
  JOURNAL   Patent: JP 1996277296-A 3 22-OCT-1996;
            KIRIN BREWERY CO LTD
COMMENT     OS   Homo sapiens (human)
            PN   JP 1996277296-A/3
            PD   22-OCT-1996
            PF   14-FEB-1995 JP 1995161363
```

```
            PR   14-FEB-1994 JP 94P    39090, 25-MAR-1994 JP 94P    79842, PR
            01-JUN-1994 JP 94P   155126, 15-JUN-1994 JP 94P   167328, PR
            17-AUG-1994 JP 94P   227159, 17-AUG-1994 JP 94P   193169, PR
            01-NOV-1994 JP 94P   304167, 01-DEC-1994 JP 94P   298669, PR
            28-DEC-1994 JP 94P   341200
            PI   MIYAZAKI HIROSHI, KATO HISASHI, OGAMI KINYA, IWAMATSU AKIHIKO,
            PI   AKAHORI HIRONORI, KUROKI RYOTA, SHIMIZU TOSHIYUKI, PI   MUTO
            TAKANORI
            PC   C07K14/52,C07H21/04,C12N15/09,C12P21/02//A61K38/00,(C12P21/02,
            PC   C12R1:19),
            PC   (C12P21/02,C12R1:91);
            CC   strandedness: Double;
            CC   topology: Linear;
            FH   Key              Location/Qualifiers
            FH
            FT   source           1. .4506
            FT                    /organism='Homo sapiens'
            FT   intron           284. .514
            FT   exon             545. .642
            FT   intron           643. .928
            FT   exon             929. .1015
            FT   intron           1016. .2947
            FT   exon             2948. .3115
            FT   intron           3116. .3351
            FT   CDS
            FT
            join(271. .283,515. .642,929. .1015,2948. .3115,3352. .4017) FT
                  /product='thrombopoietin'
            FT   sig_peptide      join(271. .283,515. .564).
FEATURES              Location/Qualifiers
     source           1. .4506
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
BASE COUNT     1149 a   1277 c    972 g   1108 t
ORIGIN

  Query Match           25.2%;  Score 30;  DB 21;  Length 4506;
  Best Local Similarity 100.0%;  Pred. No. 1.07e-11;
  Matches   30;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    4477 GATCTTTACTCTTGAGAAATGAATAAGCTT 4506
           ||||||||||||||||||||||||||||||
Qy       1 GATCTTTACTCTTGAGAAATGAATAAGCTT 30


RESULT    8
ID   E11962     standard; DNA; HUM; 4506 BP.
AC   E11962;
NI   d1110299
DT   08-OCT-1997 (Rel. 52, Created)
DT   08-OCT-1997 (Rel. 52, Last updated, Version 1)
```



```
DE   Human genomic DNA sequence encoding thrombopoietin.
KW   JP 1996228781-A/3.
OS   Homo sapiens (human)
OC   Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates;
OC   Catarrhini; Hominidae; Homo.
RN   [1]
RP   1-4506
RA   Miyazaki H., Kato H., Ogami K., Iwamatsu A., Akahori H., Kuroki R.,
RA   Shimizu T., Muto T.;
RT   "DNA CODING FOR PROTEIN HAVING TPO ACTIVITY";
RL   Patent number JP 1996228781-A/3, 10-SEP-1996.
RL   KIRIN BREWERY CO LTD.
CC   OS   Homo sapiens (human)
CC   PN   JP 1996228781-A/3
CC   PD   10-SEP-1996
CC   PF   14-FEB-1995 JP 1995063298
CC   PR   14-FEB-1994 JP 94P    39090, 25-MAR-1994 JP 94P    79842,
CC   PR   01-JUN-1994 JP 94P   155126, 15-JUN-1994 JP 94P   167328,
CC   PR   17-AUG-1994 JP 94P   227159, 17-AUG-1994 JP 94P   193169,
CC   PR   01-NOV-1994 JP 94P   304167, 01-DEC-1994 JP 94P   298669,
CC   PR   28-DEC-1994 JP 94P   341200
CC   PI   MIYAZAKI HIROSHI, KATO HISASHI, OGAMI KINYA, IWAMATSU AKIHIKO,
CC   PI   AKAHORI HIRONORI, KUROKI RYOTA, SHIMIZU TOSHIYUKI,
CC   PI   MUTO TAKANORI
CC   PC   C12N15/09,A61K38/00,C07K14/52,C12N1/21,C12N5/10,C12P21/02,
CC   PC   (C12N1/21,
CC   PC   C12R1:19),(C12N5/10,C12R1:91),(C12P21/02,C12R1:19),(C12P21/02,
CC   PC   C12R1:91);
CC   CC   strandedness: Double;
CC   CC   topology: Linear;
CC   FH   Key              Location/Qualifiers
CC   FH
CC   FT   source           1. .4506
CC   FT                    /organism="Homo sapiens"
CC   FT   intron           284. .514
CC   FT   exon             515. .642
CC   FT   intron           643. .928
CC   FT   exon             929. .1015
CC   FT   intron           1016. .2947
CC   FT   exon             2948. .3115
CC   FT   intron           3116. .3351
CC   FT   CDS
CC   FT                    join(271. .283,515. .642,929. .1015,2948. .3115,335
CC   FT                    2. .4017)
CC   FT                    /product="thrombopoietin precursor"
CC   FT   sig_peptide      join(271. .283,515. .564)
CC   FT   mat_peptide
CC   FT                    join(565. .642,929. .1015,2948. .3115,3352. .4014)
CC   FT                    /product="thrombopoietin"
CC   FT   variation        replace(253,"a")
CC   FT   variation        replace(3594,"a")
```



```
CC   FT   variation        replace(4052,"g")
FH   Key              Location/Qualifiers
FH
FT   source           1. .4506
FT                    /organism="Homo sapiens"
SQ   Sequence 4506 BP; 1149 A; 1277 C; 972 G; 1108 T; 0 other;

  Query Match           25.2%;  Score 30;  DB 10;  Length 4506;
  Best Local Similarity 100.0%;  Pred. No. 1.07e-11;
  Matches   30;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    4477 GATCTTTACTCTTGAGAAATGAATAAGCTT 4506
           ||||||||||||||||||||||||||||||
Qy       1 GATCTTTACTCTTGAGAAATGAATAAGCTT 30


RESULT    9
LOCUS       E12215       4506 bp    DNA             PAT       27-APR-1998
DEFINITION  Human thrombopoietin gene.
ACCESSION   E12215
NID         g3251049
KEYWORDS    JP 1996291196-A/2.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4506)
  AUTHORS   Miyazaki,H., Kato,H., Ogami,K., Iwamatsu,A., Akahori,H., Kuroki,R.,
            Shimizu,T. and Muto,T.
  TITLE     PROTEIN HAVING TPO ACTIVITY
  JOURNAL   Patent: JP 1996291196-A 2 05-NOV-1996;
            KIRIN BREWERY CO LTD
COMMENT     OS   Homo sapiens (human)
            PN   JP 1996291196-A/2
            PD   05-NOV-1996
            PF   14-FEB-1995 JP 1995355052
            PR   14-FEB-1994 JP 94P    39090, 25-MAR-1994 JP 94P    79842, PR
            01-JUN-1994 JP 94P   155126, 15-JUN-1994 JP 94P   167328, PR
            17-AUG-1994 JP 94P   227159, 17-AUG-1994 JP 94P   193169, PR
            01-NOV-1994 JP 94P   304167, 01-DEC-1994 JP 94P   298669, PR
            28-DEC-1994 JP 94P   341200
            PI   MIYAZAKI HIROSHI, KATO HISASHI, OGAMI KINYA, IWAMATSU AKIHIKO,
            PI   AKAHORI HIRONORI, KUROKI RYOTA, SHIMIZU TOSHIYUKI, PI   MUTO
            TAKANORI
            PC   C07K14/53,C07K1/22,C12N15/09//A61K38/00,C12P21/08; CC
            strandedness: Double;
            CC   topology: Linear;
            FH   Key              Location/Qualifiers
            FH
            FT   source           1. .4506
            FT                    /organism='Homo sapiens'
```



```
            FT   intron           284. .514
            FT   exon             515. .642
            FT   intron           643. .928
            FT   exon             929. .1015
            FT   intron           1016. .2947
            FT   exon             2948. .3115
            FT   intron           3116. .3351
            FT   CDS
            FT
            join(271. .283,515. .642,929. .1015,2948. .3115,3352. .4017) FT
                  /product='thrombopoietin'
            FT   sig_peptide      join(271. .283,515. .564).
FEATURES              Location/Qualifiers
     source           1. .4506
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
BASE COUNT     1149 a   1277 c    972 g   1108 t
ORIGIN

  Query Match           25.2%;  Score 30;  DB 21;  Length 4506;
  Best Local Similarity 100.0%;  Pred. No. 1.07e-11;
  Matches   30;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    4477 GATCTTTACTCTTGAGAAATGAATAAGCTT 4506
           ||||||||||||||||||||||||||||||
Qy       1 GATCTTTACTCTTGAGAAATGAATAAGCTT 30


RESULT   10
LOCUS       CEF11A1     39353 bp    DNA             INV       21-SEP-1998
DEFINITION  Caenorhabditis elegans cosmid F11A1, complete sequence.
ACCESSION   Z50857
NID         g967024
KEYWORDS    HTG; Nuclear hormone receptor; Transfer-RNA; tRNA-glu.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 39353)
  AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
  TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
```

```
            elegans
  JOURNAL   Nature 368 (6466), 32-38 (1994)
  MEDLINE   94150718
REFERENCE   2  (bases 1 to 39353)
  AUTHORS   Harris,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-NOV-1996) Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     Coding sequences below are predicted from computer analysis, using
            predictions from Genefinder (P. Green, U. Washington), and other
            available information.
            For a graphical representation of this sequence and its analysis
            see:-
            http://webace.sanger.ac.uk/cgi-
            bin/display?db=wormace&class=Sequence &object=F11A1
            Current sequence finishing criteria for the C. elegans genome
            sequencing consortium are that all bases are either sequenced
            unambiguously on both strands, or on a single strand with both a
            dye primer and dye terminator reaction, from distinct subclones.
            Exceptions are indicated by an explicit note.
            IMPORTANT:  This sequence is NOT necessarily the entire insert of
            the specified clone.  It may be shorter because we only sequence
            overlapping sections once, or longer because we arrange for a small
            overlap between neighbouring submissions.
            This sequence is the entire insert of clone F11A1. The true right
            end of clone M79 is at 5947 in this sequence. The start of this
            sequence (1. .104) overlaps with the end of sequence Z50806.
            The end of this sequence (39254. .39353) overlaps with the start of
            sequence AL009066.
FEATURES              Location/Qualifiers
     source           1. .39353
                      /organism="Caenorhabditis elegans"
                      /db_xref="taxon:6239"
                      /chromosome="X"
                      /clone="F11A1"
     gene             17129. .17456
                      /gene="F11A1.1"
     CDS              join(17129. .17302,17352. .17456)
                      /gene="F11A1.1"
                      /codon_start=1
                      /db_xref="PID:e348773"
                      /db_xref="PID:g967025"
                      /db_xref="SPTREMBL:Q19331"
                      /translation="MTSNSFRNFHKLPFYPTPYYFSAVSYETPIQTSKTPGSASVIDV
                      EKIDDDDKSSSPASRHRRIHSKNNCLTPPNSDDDVSLLLCFFNQVRLL"
     gene             complement(28112. .31091)
                      /gene="F11A1.2"
     CDS              complement(join(28112. .28424,28478. .28518,29549. .29817,
                      31055. .31091))
                      /gene="F11A1.2"
                      /codon_start=1
```



```
                      /db_xref="PID:e349167"
                      /db_xref="PID:g967026"
                      /db_xref="SPTREMBL:Q19332"
                      /translation="MYNYRQINSECAAGKKKQSSKAVRKKNEKKWGKTTDGRWVVCNI
                      LNRRVCVCLKPTLSFAARVQNEFCFSHIFRRDDAEHTHTLHSTPIKRKEVLSHNSNNT
                      PRRATGGYNGWRWCPTPSIVRVVATLKSSEVRVNEKKRLALRLAGVSSELQTEFRPVL
                      ASIVSMVGTAISHLLLSLSNLVQFTLRAAITKYTVKIIVFYARKECSVVFFNFFLPGYK
                      L"
     gene             31980. .36995
                      /gene="F11A1.3"
     CDS              join(31980. .32171,32311. .32374,32817. .32899,32947. .33069,
                      33115. .33381,33448. .33495,33544. .33615,34309. .34425,
                      34764. .34838,35174. .35271,35321. .35672,35920. .36076,
                      36160. .36436,36484. .36645,36692. .36767,36903. .36995)
                      /gene="F11A1.3"
                      /note="similar to nuclear hormone receptor"
                      /codon_start=1
                      /db_xref="PID:e348774"
                      /db_xref="PID:g967027"
                      /db_xref="SPTREMBL:Q19333"
                      /translation="MFHQQPPVVDLNNLFMYHLLVSFNEPTNYCQLLQIHNYSYPLNT
                      PQIIKCKEQEKRWRHRKIKSPQMSTPDDPVIHSPPSIGAAPGYHGSGVKLEESSGACG
                      SPDDGLLDSSEESRRRQKTCRVCGDHATGYNFNVITCESCKAFFRRNALRPKEFKCPY
                      SEDCEINSVSRRFCQKCRLRKCFTVGMKKEWILNEEQLRRRKNSRLNNTGTCNKRSQP
                      GNQQSPQGPNQQPHLSPHHPGVAIPPQPQRPLTINPMDNQMHHMQANRPNAMPQLI
                      SPPGAQPYPLTSPVGSSASDSPPNRSLTMMNGEKSPDGVSGSAGKTDPNINAHRAPP
                      PSFNNRPFKMDSGQVVLSTEEYKQLLSRIPGAQVPGLNNEEEPINKRAAYNCNGHPMPA
                      ETTPPYSAPMSDMSLSRHNSTSSGTEKNHMTHSTVSAIPGNSAQNHFDIASPGMGIVT
                      ATGGGDAAEEMYKRMNMFYENCIQSALDSPENQEPKPQEAMIPKEEYMTPTHGFQYQS
                      DPYQVPPAERNINYQLNAAELKALDAVREAFYGMDDPMEQGPDINNIMDVTMRRFVKV
                      AKGVPAFREVSQEGKFSLLKGGMIEMLTVRGVTRYDASTNSFKTPTIKGQNVSVNVDD
                      MFAKLNANAQAQKAKCLEFPGFFDEEIKKNELAVYLVMLAVLFSVRSDPPMNENDVRI
                      VTERHNHFMSLLNRYLESLFGEQARRIPERIPKALGLLNEIARNAGMLFMGTVRSGEA
                      EELPGEFFKIK"
     tRNA             complement(38536. .38607)
                      /gene="F11A1.t2"
                      /note="TTC Glu E-tRNA; predicted using tRNAscan-SE-1.11;
                      preliminary prediction; similar to tRNA-Glu"
     gene             complement(38536. .38607)
                      /gene="F11A1.t2"
     tRNA             39107. .39178
                      /gene="F11A1.t1"
                      /note="TTC Glu E-tRNA; predicted using tRNAscan-SE-1.11;
                      preliminary prediction; similar to tRNA-Glu"
     gene             39107. .39178
                      /gene="F11A1.t1"
     gene             join(complement(380. .503),complement(234. .321),
                      complement(73. .184),complement(Z50806:33620. .33673),
                      complement(Z50806:33444. .33565),
                      complement(Z50806:33155. .33242),
                      complement(Z50806:32838. .33101))
```



```
                      /gene="M79.2"
     CDS              join(complement(380. .503),complement(234. .321),
                      complement(73. .184),complement(Z50806:33620. .33673),
                      complement(Z50806:33444. .33565),
                      complement(Z50806:33155. .33242),
                      complement(Z50806:32838. .33101))
                      /gene="M79.2"
                      /codon_start=1
                      /db_xref="PID:e1322086"
                      /db_xref="PID:g3642276"
                      /translation="MCIHVVFVVYLRQGQQFSDSVWLNKLCFFPLASFVFFLFIKYML
                      LSIFTFYGSCITLTYFSTVVLIFFGLSWGVLPQWYLHVVAMIQGYFPREKLENTYENE
                      ESEEQRSEKSADQIAEERVKRIIDTDLGPCFQTAPTIRPTLLGMMPEKNWNSTEKAIF
                      FAGLIFRIFFLMPVRVCLLLTSFVFVALAGLQTAFRTLSDREKTWVAIVYCRLFCGSM
                      GLVANYENPQFRPKKPGVAVSNHLTPNDIQILWAGTPHGSSTGFVVTGQKHKGIIGIY
                      CYLRVVC"
BASE COUNT    12348 a    7137 c    6554 g   13314 t
ORIGIN

  Query Match           16.8%;  Score 20;  DB 18;  Length 39353;
  Best Local Similarity  95.2%;  Pred. No. 1.43e-02;
  Matches    20;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Db  26783 TTTGCTTTTTAGGGAGCTTTT 26803
          ||||||||||| ||||||||||
Cp    120 TTTGCTTTTTANGGAGCTTTT 100


RESULT   11
LOCUS       G37636         455 bp    DNA             STS       30-MAR-1998
DEFINITION  SHGC-58226 Human Homo sapiens STS genomic, sequence tagged site.
ACCESSION   G37636
NID         g2997287
KEYWORDS    STS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 455)
  AUTHORS   Myers,R.M.
  TITLE     Human STSs (1997)
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Richard M. Myers
            Stanford Human Genome Center (SHGC)
            Stanford University School of Medicine
            Department of Genetics, M-344, Stanford, CA 94305, USA
            Tel: 4157259687
            Fax: 4157259689
            Email: myers@shgc.stanford.edu
            Primer A: TGAAATCCCTGCCCTCAGTT
```



```
            Primer B: CCCCGAACATCAGGCATAGT
            STS site: 237
            PCR Profile:
                  Initial incubation: 95 degrees C for 10 minutes

                  Denaturation:       94 degrees C for 30 seconds
                  Annealing:          60 degrees C for 30 seconds
                  Polymerization:     72 degrees C for 23 seconds
                  PCR Cycles:         30
                  Thermal Cycler:     Perkin Elmer 9700
            Protocol:
                  Template:                  25 ng
                  Primer:                    each 1 uM
                  dNTPs:                     each 200 uM
                  AmpliTaq Gold Polymerase:  0.07 units/ul
                  Total Vol:                 5 ul

            Buffer:
                  MgCl2:          2.5 mM
                  KCl:            50 mM
                  Tris-HCl:       10 mM
                  pH:             8.3

            Prepared with primer pairs derived from W89200 -- Unigene.
FEATURES              Location/Qualifiers
     source           1. .455
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /map="1"
                      /clone_lib="Human"
     STS              120. .356
     primer_bind      120. .139
     primer_bind      complement(337. .356)
BASE COUNT      146 a      87 c      87 g     134 t       1 others
ORIGIN

  Query Match           15.1%;  Score 18;  DB 30;  Length 455;
  Best Local Similarity 100.0%;  Pred. No. 5.92e-01;
  Matches    18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Db    166 TCCCATTTATTCCTTATA 183
          ||||||||||||||||||
Cp     93 TCCCATTTATTCCTTATA 76


RESULT   12
LOCUS       CELB0416    44797 bp    DNA             INV       27-NOV-1996
DEFINITION  Caenorhabditis elegans cosmid B0416.
ACCESSION   U23516
NID         g746501
KEYWORDS    .
```

```
SOURCE      Caenorhabditis elegans strain=Bristol N2.
  ORGANISM  Caenorhabditis elegans
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Nematoda;
            Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditidae;
            Caenorhabditis.
REFERENCE   1  (bases 1 to 44797)
  AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
  TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
  JOURNAL   Nature 368 (6466), 32-38 (1994)
  MEDLINE   94150718
REFERENCE   2  (bases 1 to 44797)
  AUTHORS   Favello,A.
  TITLE     The sequence of C. elegans cosmid B0416
  JOURNAL   Unpublished (1996)
REFERENCE   3  (bases 1 to 44797)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-1995)
REFERENCE   4  (bases 1 to 44797)
  AUTHORS   Wateraton,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-NOV-1996)
COMMENT     Submitted by:
                Genome Sequencing Center
                Department of Genetics, Washington University,
                St. Louis, MO  63110, USA, and
                Sanger Centre, Hinxton Hall
                Cambridge CB10 IRQ, England
                e-mail: rw@nematode.wustl.edu and jes@sanger.ac.uk

            NOTICE:  This sequence may not be the entire insert of this clone.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we provide a small overlap between
            neighboring submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded or sequenced with an alternate
            chemistry; an attempt was made to resolve all sequencing problems,
            such as compressions and repeats; all regions were covered by
```



```
            sequence from more than one subclone

                        NEIGHBORING COSMID INFORMATION:

            The 5' cosmid is F18G5, 200 bp overlap:3' cosmid is T02A10.  Actual
            start of this cosmid is at base position 1 of CELB0416; actual end
            is at 44797 of CELB0416

            NOTES:

            Coding sequences below are predicted from computer analysis, using
            the program Genefinder(P. Green and L. Hillier, ms in preparation).
FEATURES              Location/Qualifiers
     source           1. .44797
                      /organism="Caenorhabditis elegans"
                      /strain="Bristol N2"
                      /db_xref="taxon:6239"
                      /chromosome="X"
                      /clone="B0416"
     gene             complement(3907. .5671)
                      /gene="B0416.5"
     CDS              complement(join(3907. .4095,4167. .4823,4873. .5058,
                      5105. .5345,5421. .5671))
                      /gene="B0416.5"
                      /codon_start=1
                      /evidence=not_experimental
                      /db_xref="PID:g746506"
                      /translation="MGADDGAISADKVDIRLQPLNKNGAQDETDEEAADQIQLRVYKQ
                      RWIVLLAVALLNNTNTMSWIGYAPSGNYVNSFYGESSAAWLSMVYMMCTIPVGMFAMW
                      AGREWGLRTAVLIAGWANGIGAVIRVISSLDFVPQDLRFPICMTGQGIAAIATPFIMF
                      LPTKVAGSWFFDTQRAIATSIGVMSNPLGVLMANLISPAIVKSPEHVIWLNIFTCVPS
                      LIAMLIATFGVNRSEPKIPPTFSASRPQMDFVSGMKSCFSSKYIILLIVMGGGIGMF
                      NCLYTVMLELLCPSGYSNFFSGVCAALMIVGGVPGAAASSIFVDRTKLYEETLKIALG
                      AAVFGLIPLQLTLHQGYSVILGVTCLLFGVLGLATTPIGLELASECTFPVSEATSTG
                      LIVLSGQIQSVIYVFPIMKNFARPLQFDRMHIQVCQLTPDDTINTPKDNTMSIMIFSLL
                      ATLLVLTLVVLFKPVYKRLEAERGNRATADKAELSNQNKDRITLQAESAVEPLQKK"
     gene             13797. .14678
                      /gene="B0416.4"
     CDS              join(13797. .14316,14542. .14678)
                      /gene="B0416.4"
                      /note="weakly similar to S. pneumoniae PspA (surface
                      protein A) (PIR:A41971)"
                      /codon_start=1
                      /evidence=not_experimental
                      /db_xref="PID:g746505"
                      /translation="MAPIRRSSRLAERYDAIESKKRSLKRLEEQIKAEEEQPSDKMKQ
                      LEDEIKIKEQVITMPKRKTVRREWMRNSRQATTNINIAQIESLKLQLEEGEKDIAEAE
                      KQAEPTTPQQEAELSETFKQMVRDRMKVKDVDEKLLQQYMKKENVEFEWRSCFICTME
                      YSRTDKNLHPIILNCGHNLCRSCINKLTGNGIVKCPFDRLDTRVRVTGLPRNLALINL
```



```
                      .
     gene             26547. .28201
                      /gene="B0416.3"
     CDS              join(26547. .26596,26740. .26857,27005. .27116,27275. .27351,
                      27399. .27525,27971. .28065,28148. .28201)
                      /gene="B0416.3"
                      /note="coded for by C. elegans cDNA cm5f11"
                      /codon_start=1
                      /db_xref="PID:g746504"
                      /translation="MNEDNHESSSSGCDNQIEFLVSEAEQRRRRVWHLDTEVHGLGYD
                      PFSYCGRRVHVLLLTKLISVLTIPFTVAIIIPISFFGNATSVMFSVIILGSVLISTCY
                      GAFRGAKMCLIPFVIIQLVFLIYDLILITILLLAVVFPKMPLSALLRLPLEDIPPGTD
                      QVLLGCSLLLALLLAPLVWTTHVVYIDFLFISQANQKVSQDDVSPNRMMF"
     gene             complement(28836. .31427)
                      /gene="B0416.6"
     CDS              complement(join(28836. .29043,29478. .29590,29640. .29730,
                      29775. .29924,30172. .30251,30375. .30416,30457. .30606,
                      30656. .30764,30816. .30871,30919. .30963,31139. .31297,
                      31344. .31427))
                      /gene="B0416.6"
                      /note="highly similar to alpha-1,3-mannosyl-glycoprotein
                      beta-1,2-n-acetyglucosaminyltransferase"
                      /codon_start=1
                      /evidence=not_experimental
                      /db_xref="PID:g746507"
                      /translation="MHAVTKIFIIFIFVFILWTLYVENDITNRTRNTDNIDDLLESAN
                      RLERLLKFEAKKIAAAEDVHKIRANRKGKHVIMEEMVSQDLKQWKDPIPVLVFSCNR
                      AMAVRDHVEKLIRYRPSQEKFPIIVTQDCDNENVKNEVKKFGDKVEYIKHLAGDKANI
                      TIPPSHRQTTAYYRIARHYKLALNHVFVDKGYSSVIITEGKTISTFRQTSSRTSQQEN
                      IDMTAASTLYRSDFFAGLGWMMSSKTWHELEPIWPVGFWDDWMRDPARRKDRQCIRPE
                      ISRTGMMSYGKEGASKGQFFSKHLAKIKVNDKYINFGKIDLDYLLPANFAKKTNLEVM
                      KEAVELSIDNVASFVLSSENKGKSVRVMYDGNIDYIRKADKLHIMHDFKAGVPRTAYD
                      GIVTCFINGIRIYLVPDRTKVSAYNPDWSVPPSFGE"
     gene             complement(31845. .32756)
                      /gene="B0416.7"
     CDS              complement(join(31845. .31896,31967. .32073,32383. .32479,
                      32614. .32756))
                      /gene="B0416.7"
                      /codon_start=1
                      /evidence=not_experimental
                      /db_xref="PID:g746508"
                      /translation="MNLSTAIISCLCIGVLSYENKFQPTGTVYCKSNAPWCVRIRVIE
                      VDTLVDDGVASVDFCSNEVTQTYDIGGVQENDGLLDRNFELQMVVFHNCSRNTETVFK
                      TGIWRIPLPKAPTVHAPIRQHLNLNTNNSQ"
     gene             34094. .34921
                      /gene="B0416.2"
     CDS              join(34094. .34157,34217. .34366,34411. .34525,34569. .34662,
                      34841. .34921)
                      /gene="B0416.2"
                      /codon_start=1
                      /evidence=not_experimental
```



```
                      /db_xref="PID:g746503"
                      /translation="MPFSVTKFSLIFVALLLAFALVAQSSQLESVTRKPKPFPLVKKS
                      MLVGMSETTEEVCQTVVLHNYEPVYGHLINGSLVEILQASGHKFSKTHVQCLEEDRPS
                      CHGVKDDMTISECVTVIENANTMVRLFNSFGPYRLGTIRIPILCECRLRRQYRDFERP
                      GDDDDDY"
     gene             38617. .44667
                      /gene="B0416.1"
     CDS              join(38617. .38757,39060. .39142,39202. .39342,39393. .39657,
                      39942. .40064,40117. .40262,40329. .40426,40475. .40689,
                      40734. .40849,40904. .41041,41189. .41458,41504. .41699,
                      42083. .42140,42189. .42438,42586. .42711,42759. .42837,
                      42933. .43372,43652. .43768,43844. .43900,43952. .44062,
                      44404. .44520,44571. .44667)
                      /gene="B0416.1"
                      /note="coded for by C. elegans cDNA yk16h5.5; coded for by
                      C. elegans cDNA yk4g8.5"
                      /codon_start=1
                      /db_xref="PID:g746502"
                      /translation="MPKSGAHQPLVRHDTDDGGETGQSVKSLADVSEEEIDSRMSRRS
                      SVIADLLSLFRRSSSVLVRPHTRLGPNFDDDDDEFDEEDDKEASKDRILKKIQQKKE
                      IIQKLRGQPWTMKRKRRTLEVAQKHLQQQEAKVSKARLYKAEAGRRLTQASRWLDNLK
                      IVLIPWEAIIKIESHFGSVVSSIFTFHRWVLGVNITIIFIMCMFVVIPEWLADSRTQ
                      PGDDRYNKTKAIKVMPPAVRARADELSTVWDFGGYFQYSLLFYGFYSKETFFGETIKY
                      RVPVAYFFCNIFILGFSLFIILRKMAANNRRGTLSSGKTQQYLFNWKAFTGWDYTIGN
                      PETAGNVYMANVIKFREAINDDKQKPSDKHPWIRFVARVLTNLFICAMYVFSIWAIMQ
                      CGTLKGEHFFAQNATITISLITLVFPNIFDLLGKIEKLHPRNALRFQLGRVLVLYIL
                      NIYTLIYSLMLQLEHLQKEKNASDNPISALGHPGDAIGRTIRETVLPRIPVDNNPHTY
                      YSYAPVTTTPIPATSSWTTVLPDFGPFGVYNPKASVTKDDTVFSSPVVETHMPGPNSD
                      WNETTVNAASPTGATTRASLRMSQGGLCWETIIGQEITKLVTMDLYMTVASIFLIDFL
                      RGLACRYLNLYWPWDLERTFPEYGEFKVAENVLHLVNNQGMIWLGLFPVPLLPMLNNI
                      KLIILMYIRGWAAMTCNVPASQIFRASRSSNFFALLLFLPLCTLPVGFVIASKTPS
                      KSCGPPCNQSFFYSVITDVLHENLDKTLVNGIKYSLSPGIIIPVLVLLSLVIYFLIAM
                      VTGLSQANQDLSFQLWVERTEEKKKIFELAGGKKKKSKDNTPGKQKPKQLLPPPTKGV
                      SSDDDSQHNRSTAKSVSGGRQFVPSLGSVSEVDHSTGEEQSSDSESTTSSLPPKLSLRQ
                      RFLVCIGWADPNKYGRHDDIEMEEGGGRLRELSTGSETDSDDEDSEKSNRDMSYRTAI
                      QSFDQNSQSASASSSKSTTTAPSNSEMRIEITENPLHTYITPLRIEKKSSASSSSSSH
                      QPSSSIEKQAARRLLQPISTTHNIRYGVATVENSSQDPTRPPSTDDSLGDPALHEPLW
                      ANLNPHSSYTSAMMSPIMNEVMSNDETTDDEKGRLIPDR"
BASE COUNT    14789 a   7823 c   7432 g  14753 t
ORIGIN

  Query Match           15.1%;  Score 18;  DB 18;  Length 44797;
  Best Local Similarity 100.0%;  Pred. No. 5.92e-01;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  16363 ATAAGGAATAAATGGGAG 16380
          ||||||||||||||||||
Qy     77 ATAAGGAATAAATGGGAG 94


RESULT   13
```

```
LOCUS       AC004677   128559 bp    DNA             PRI       10-MAY-1998
DEFINITION  Homo Sapiens Chromosome X clone bWXD501, complete sequence.
ACCESSION   AC004677
NID         g3126881
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 128559)
  AUTHORS   Chen,E., Brownstein,B.H., States,D.J., Schlessinger,D. and
            Mazzarella,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished (1997)
REFERENCE   2  (bases 1 to 128559)
  AUTHORS   Brownstein,B.H., States,D.J. and Mazzarella,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-1998) Center for Genetics in Medicine, Box 8232,
            Washington University School of Medicine, 4566 Scott Avenue, St.
            Louis, MO 63110, USA
COMMENT     Current status of this project is available at:
            'http://www.ibc.wustl.edu/cgm/seq_projects.html'
            Submitted by:
            Ellson Chen,
            Advanced Center for Genetic Technology,
            Applied Biosystems Division of Perlin Elmer Corp.,
            850 Lincoln Center Drive,
            Foster City, CA 94404 USA
            e-mail: ellson@genseq.apldbio.com

            and

            Buddy Brownstein,
            Center for Genetics in Medicine,
            Washington University School of Medicine, Box 8232
            4566 Scott Avenue,
            St. Louis, MO 63110, USA
            e-mail: buddy@genetics.wustl.edu

            and

            David J. States,
            Institute for Biomedical Computing
            Washington University in St. Louis
            700 South Euclid Ave.
            St. Louis, MO 63108 USA
            e-mail: states@ibc.wustl.edu.
FEATURES             Location/Qualifiers
     source          1. .128559
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
```



```
                     /clone="bWXD501"
                     /chromosome="X"
BASE COUNT    41143 a  24868 c  24792 g  37756 t
ORIGIN

  Query Match           15.1%;  Score 18;  DB 26;  Length 128559;
  Best Local Similarity 100.0%;  Pred. No. 5.92e-01;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  22169 TTGAGAAATGAATAAGCT 22186
          ||||||||||||||||||
Qy     12 TTGAGAAATGAATAAGCT 29


RESULT   14
LOCUS       MMAT01        876 bp    DNA             ROD       06-FEB-1997
DEFINITION  Mus musculus aspartate aminotransferase gene 5'-flank and exon 1
            (joined CDS).
ACCESSION   X06917
NID         g13935
KEYWORDS    aspartate aminotransferase; L-aspartate-2-oxoglutarate
            aminotransferase.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Mus.
REFERENCE   1  (bases 1 to 876)
  AUTHORS   Tsuzuki,T., Obaru,K., Setoyama,C. and Shimada,K.
  TITLE     Structural organization of the mouse mitochondrial aspartate
            aminotransferase gene
  JOURNAL   J. Mol. Biol. 198 (1), 21-31 (1987)
  MEDLINE   88118911
COMMENT     see x06917-x06926 for mAspATase gene
            Data kindly reviewed (13-JUN-1988) by Tszuki T.
FEATURES             Location/Qualifiers
     source          1. .876
                     /organism="Mus musculus"
                     /strain="C3H/HE"
                     /db_xref="taxon:10090"
                     /clone_lib="Charon4A"
     mRNA            join(<1. .806,X06918:31. .188,X06919:32. .159,X06920:41. .100,
                     X06921:41. .202,X06922:39. .143,X06923:31. .181,
                     X06924:41. .206,X06925:31. .181,X06926:37. .1152)
                     /label=ASP_mRNA
                     /product="aspartate aminotransferase"
     protein_bind    531. .540
                     /bound_moiety="SP1"
     exon            (612.637). .806
                     /label=ex1
     protein_bind    709. .714
```



```
                     /bound_moiety="SP1"
     CDS             join(718. .806,X06918:31. .188,X06919:32. .159,
                     X06920:41. .100,X06921:41. .202,X06922:39. .143,
                     X06923:31. .181,X06924:41. .206,X06925:31. .181,
                     X06926:37. .159)
                     /codon_start=1
                     /label=ASP_CDS
                     /product="aspartate aminotransferase"
                     /db_xref="PID:g872128"
                     /db_xref="MGD:9230"
                     /db_xref="SWISS-PROT:P05202"
                     /translation="MALLHSSRILSGMAAAFHPGLAAAASARASSWWTHVEMGPPDPI
                     LGVTEAFKRDTNSKKMNLGVGAYRDDNGKPYVLPSVRKAEAQIAAKNLDKEYLPIGGL
                     AEFCKASAELALGENNEVLKSGRFVTVQTISGTGALRVGASFLQRFFKFSRDVFLPKP
                     SWGNHTPIFRDAGMQLQGYRYYDPKTCGFDFSGALEDISKIPEQSVLLLHACAHNPTG
                     VDPRPEQWKEIASVVKKKNLFAFFDMAYQGFASGDGDKDAWAVRHFIEQGINVCLCQS
                     YAKNMGLYGERVGAFTVVCKDAEEAKRVESQLKILIRPLYSNPPLNGARIAATILTSP
                     DLRKQWLQEVKGMADRIISMRTQLVSNLKKEGSSHNWQHITDQIGMFCFTGLKPEQVE
                     RLTKEFSVYMTKDGRISVAGVTSGNVGYLAHAIHQVTK"
BASE COUNT      198 a    248 c    202 g    228 t
ORIGIN

  Query Match           14.3%;  Score 17;  DB 28;  Length 876;
  Best Local Similarity 100.0%;  Pred. No. 3.49e+00;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     51 AAAGCTTATTCATTTCT 67
          |||||||||||||||||
Cp     31 AAAGCTTATTCATTTCT 15


RESULT   15
LOCUS       MUSASPT01     876 bp    DNA             ROD       25-FEB-1992
DEFINITION  Mouse mitochondrial aspartate aminotransferase gene, exon 1.
ACCESSION   M37250
NID         g192039
KEYWORDS    aspartate aminotransferase; mAspATase.
SEGMENT     1 of 10
SOURCE      Mouse (strain C3H/He) male liver DNA, clones mmAspAT-[1-4].
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae;
            Murinae; Mus.
REFERENCE   1  (bases 1 to 876)
  AUTHORS   Tsuzuki,T., Obaru,K., Setoyama,C. and Shimada,K.
  TITLE     Structural organization of the mouse mitochondrial aspartate
            aminotransferase gene
  JOURNAL   J. Mol. Biol. 198, 21-31 (1987)
  MEDLINE   88118911
FEATURES             Location/Qualifiers
     source          1. .876
```



```
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
     prim_transcript 612. .>876
                     /note="mAspATase mRNA and introns (alt.)"
     prim_transcript 621. .>876
                     /note="mAspATase mRNA and introns (alt.)"
     prim_transcript 631. .>876
                     /note="mAspATase mRNA and introns (alt.)"
     prim_transcript 637. .>876
                     /note="mAspATase mRNA and introns (alt.)"
     exon            718. .806
                     /partial
                     /gene="mAspATase"
                     /number=1
                     /product="mitochondrial aspartate aminotransferase"
BASE COUNT      198 a    248 c    202 g    228 t
ORIGIN

  Query Match           14.3%;  Score 17;  DB 28;  Length 876;
  Best Local Similarity 100.0%;  Pred. No. 3.49e+00;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     51 AAAGCTTATTCATTTCT 67
          |||||||||||||||||
Cp     31 AAAGCTTATTCATTTCT 15


RESULT   16
LOCUS       CELK06H7    22073 bp    DNA             INV       22-OCT-1993
DEFINITION  C. elegans cosmid K06H7.
ACCESSION   L15314 L18807
NID         g289690
KEYWORDS    isopentenyl-diphosphate delta isomerase; protein kinase; sec7 gene.
SOURCE      Caenorhabditis elegans (strain Bristol N2) hermaphrodite mixed
            whole animal DNA.
  ORGANISM  Caenorhabditis elegans
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Nematoda;
            Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea;
            Rhabditidae; Caenorhabditis.
REFERENCE   1  (bases 1 to 22073)
  AUTHORS   Sulston,J., Du,Z., Thomas,K., Wilson,R., Hillier,L., Staden,R.,
            Halloran,N., Green,P., Thierry-Mieg,J., Qiu,L., Dear,S.,
            Coulson,A., Craxton,M., Durbin,R., Berks,M., Metzstein,M.,
            Hawkins,T., Ainscough,R. and Waterston,R.
  TITLE     The C.elegans genome sequencing project: A beginning
  JOURNAL   Nature 356, 37-41 (1992)
  MEDLINE   92168156
REFERENCE   2  (bases 1 to 22073)
  AUTHORS   Favello,A.D.
  TITLE     Sequence of the C. elegans cosmid K06H7
  JOURNAL   Unpublished (1993)
```

```
COMMENT     Submitted by:
            Nematode Sequencing Project
            Department of Genetics, Washington University,
            St. Louis, MO  63110, USA, and
            MRC Laboratory of Molecular Biology,
            Cambridge CB2 2QH, England
            e-mail: rw@nematode.wustl.edu and jes@cele.mrc-lmba.cam.ac.uk NOTE:
            Coding sequences below are predicted from computer analysis, using
            the program Genefinder (P. Green and L. Hillier, in preparation)
            Neighboring cosmid information:
            Bases 1 to 200 of this entry correspond to bases 38383 to 38582 of
            CELK12H4 (GenBank:L14331).  Bases 21074 to 22073 of this entry
            correspond to bases 1 to 1000 of CELC14B9 (L15188). The actual
            cosmid sequence of K06H7 begins at nucleotide position 32992 of
            CELK12H4 and extends through this sequence and into CELC14B9 and
            ends at position 12266 of CELC14B9.
FEATURES              Location/Qualifiers
     source           1. .22073
                      /organism="Caenorhabditis elegans"
                      /strain="Bristol N2"
                      /db_xref="taxon:6239"
                      /dev_stage="mixed"
                      /sex="hermaphrodite"
                      /tissue_type="whole animal"
                      /map="3"
     misc_feature     216
                      /note="possible trans-spliced leader site for K06H7.6"
     CDS              join(282. .334,378. .507,642. .774,820. .946,990. .1233,
                      1279. .1426,1476. .1593,1640. .1730)
                      /standard_name="K06H7.6 protein"
                      /note="putative"
                      /codon_start=1
                      /db_xref="PID:g289696"
                      /translation="MSIPDRMTTSKEAAMKYRQFCVSKKKEVWDALIYHAKYAKNIEK
                      VTDMITVLQRVQQECLMFALEKGVGQPDIRDAAELIFPLAMQKVVIKVLRPYFMSHFT
                      NYRLAMYLPHRKIDHGIDLNQELWTPEKKEICGQYVQDFNPFVNELKTGILNANRRDMN
                      AALSVYIRELAATLVMHEKAAFDNELRCSMPQEIMESLRIWSDQVTMQDAPHLISDTI
                      YKDVFQMLSKNAYEMIAVDFPVKFRSLITMKYCLLRTNNHGRVDPTNYLIDQVNTKLL
                      VASVDTKDILKAYAACVESLREMDNSCVVMHKVCGVIRETLKRRPDTVQQIISYITSN
                      KKNELEKDMVSVY"
     exon             378. .507
                      /note="putative"
                      /number=9
     exon             642. .774
                      /note="putative"
                      /number=10
     exon             820. .946
                      /note="putative"
                      /number=11
     exon             990. .1233
                      /note="putative"
```



```
                      /number=12
     exon             1279. .1426
                      /note="putative"
                      /number=13
     exon             1476. .1593
                      /note="putative"
                      /number=14
     misc_feature     1766
                      /note="possible trans-spliced leader site for K06H7.5"
     CDS              join(1787. .1895,1986. .2119,2164. .2397,2442. .2539,
                      2587. .2782,2827. .2997,3044. .3238)
                      /standard_name="K06H7.5 protein"
                      /note="putative"
                      /codon_start=1
                      /db_xref="PID:g289695"
                      /translation="MMDEEELKGVNDDFLPENMETLGWENWMPNPTDATVGDGAPGHQ
                      GVDVFNMLVSVYGSKELFVKEYRNLLAERLSSSDNKDPEFERRYLDLLKLRPQYSELQ
                      HCEVMLRDVIRSLDIDKKPEDMSERSAGNYDVPIIPISACIISSHTWPKLETEKTEAL
                      MDQPLQAAMDVYQETYLDVERDRKLEWLRSVCCVTEVSINIDGVEVDRTIPNMTALLLF
                      LFLEKETWTTAEVVEKMGKSVVVTRKRLEWLVKQGFICMNPIISSDTWFLTRNPSGIS
                      IVRPGTPDLEDDEDVEPEENSDMVDALEQYWGTTRNFIANHAPNGEVKAERKHRVYRM
                      PGSPTSAGPTLDHVSAFLQRKVALGLLTCINGSYRIIQEKKDGE"
     exon             1986. .2119
                      /note="putative"
                      /number=15
     exon             2164. .2397
                      /note="putative"
                      /number=16
     exon             2442. .2539
                      /note="putative"
                      /number=17
     exon             2587. .2782
                      /note="putative"
                      /number=18
     exon             2827. .2997
                      /note="putative"
                      /number=19
     CDS              complement(join(3416. .3541,3591. .3890,3943. .4060,
                      4124. .4202,4246. .4425,4476. .4641))
                      /standard_name="K06H7.7 protein"
                      /note="putative"
                      /codon_start=1
                      /db_xref="PID:g289697"
                      /translation="MNKLSRTDEENRQTEPSSSPLSDSVHSCLGFNRKKERKRRTSGD
                      IEQTDKRKRTEERPSDKITFASNVFDNLARVTAERDNPEKLSKYKNIPGVGTRENIDE
                      GISDINQPSSKSSGVKSKYVEVMSLFQKSIQKFSIMNLDTMYTFDWNHYSKKNIQYVV
                      FTKVCQSHAPGHFVEMLKEDCSCRDCGSADLAVKVRLTVPLDLSDIPIPCSEDIQIHV
                      SFLADDLLERHISSNVNRSSNRDDLDFSRRTCEKMYDLYQKMMRGQSSRMYTEWFDKC
                      KIFVKNFLFSKSLHLYHGNVIRRKIEESQTLNKRRCFIFAQIFDLE"
     exon             3591. .3890
                      /note="complement; putative"
```



```
                      /number=5
     exon             3943. .4060
                      /note="complement; putative"
                      /number=6
     exon             4124. .4202
                      /note="complement; putative"
                      /number=7
     exon             4246. .4425
                      /note="complement; putative"
                      /number=8
     misc_feature     5451
                      /note="possible trans-spliced leader site for K06H7.4"
     CDS              join(5480. .5548,5866. .6051,6126. .6470,6533. .6744,
                      7319. .7448,7498. .7689)
                      /standard_name="K06H7.4 protein"
                      /note="homology with sec7 gene; coded for by C. elegans
                      cDNA cm21c10 (GenBank:H89330); putative"
                      /codon_start=1
                      /db_xref="PID:g289694"
                      /translation="MTLPKVRKRKAQLVDEIEALKNEVREVDEELDQVYYTHPKSKEY
                      HKIVVNGRKKFNQDPWKALDWLASRNVVAKDPQALALWMKAGEGLSKSAIGEILGDNR
                      PFALETLDRFTKEHKLHDVPIVPALRQYLPSFRLPGESQKINRILEKFAEVYANQNPS
                      YGNADQAHTVAYSCIMVNTLLHNPNVKDKPSLEKYIEMNEQLLEKGAITIEQLTEVYE
                      SVSVTQFKIPDEVSTSGKGTVNDILLHAEREGWLFKQSSNPLFSGALSWKRRWFVLSE
                      NCLTYFDQMTDKEPKGIITLANVGIRKVEAPSRPFMFEIFSLSDGQIKACKTEQDGRL
                      VEGSHHSIYKICAVNDEDMRSWINAISRAMAPQQHLLARPKSTH"
     exon             5866. .6051
                      /note="putative"
                      /number=20
     exon             6126. .6470
                      /note="putative"
                      /number=21
     exon             6533. .6744
                      /note="putative"
                      /number=22
     exon             7319. .7448
                      /note="putative"
                      /number=23
     misc_feature     8165
                      /note="possible trans-spliced leader site for K06H7.3"
     CDS              join(8170. .8309,8364. .8408,8519. .8598,8645. .8765,
                      9158. .9271,9320. .9445,9496. .9988,10191. .10364,
                      10424. .10663,10711. .10863,11005. .11099,11338. .11516,
                      11567. .11646,12020. .12350,12427. .12551)
                      /standard_name="K06H7.3 protein"
                      /note="homology with
                      isopentenyl-diphosphate-delta-isomerase; coded for by C.
                      elegans cDNAs GenBank:HB0049 and Z14629; putative"
                      /codon_start=1
                      /db_xref="PID:g289693"
                      /translation="MGFTLKSLEFLQGVEPFRIDENHENEGASEEEEVLISAETDAGM
```



```
                      SLIMENAFVRRLGPVYNITNTLRKARNMTVYTEEDFEGIENSENDLTSSQTTIGLESD
                      DEEFDALLLPANRSFFIKNGSVFSVPRNILHVGERDVSSVTFLRPFDCAIFLWNGGHF
                      AAAMFENDKMTVQKSFHRTVARAKQGGVQSQHDSGGKGAAKSAGAQLRRTNEQKMKEE
                      IQSIMSSWKSRLQKTPLLFIRCAAYHRNIFFEADAGIETRDDRIRTIPFETKRPNIDE
                      ISDCWQRLQQVSEHGAESDFRAEMLEVREKRKKLARKVAGKKRKDGGMQMICEWSDDD
                      ENEDISKEKKTHHIKVRTIKKPEETVVQWPRLDDEWRQKTYNYVRQDSVEALKEHLAS
                      LNEDVTSEANDYLRNAKIPPNRSTPLHVSAANDARKCLKYFLEEVMCDSSTKDGAGLP
                      PYSSSANSDVKSIFIDYRVKNETAGNWARTHIPEPKKKVELTEEQERDQAERKEEKKA
                      RQKEKEKLLKKEIAKRDVEEMEERQKTVVNMSEREKRALAVDRRLAGLPPILRCHQCGVQ
                      LPPTPFQYSHYNPCSTSCVAEHRKANPQKMKPSLRGLSQIGMARRAVELSSYDAQQVE
                      YMREQCINVDENDRIIGGVSKREAHSSYQLVLHRAFSVPSFTSDNKLLMQKRSAEKIT
                      FPNLWTNTCCSHPLHTQQEMDGVPGAKRAAIRKLEHELGITGVSLNKFQMAGKYIYRA
                      EMNNAPWGEHELDYALILRGIGAENCNINENEVAEVREVVLDELKLWMRKEPNSFTPW
                      LKLFSQTETFEKWWTINKNTEESIEDTNIYKLH"
     exon             8364. .8408
                      /note="putative"
                      /number=24
     exon             8519. .8598
                      /note="putative"
                      /number=25
     exon             8645. .8765
                      /note="putative"
                      /number=26
     exon             9158. .9271
...
Note: remainder of annotations omitted.

  Query Match           14.3%;  Score 17;  DB 18;  Length 22073;
  Best Local Similarity 100.0%;  Pred. No. 3.49e+00;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   4628 GAAAGCTTATTCATTTC 4644
           |||||||||||||||||
Cp     32 GAAAGCTTATTCATTTC 16


RESULT   17
LOCUS       CEBA2       31197 bp    DNA             HTG       27-NOV-1996
DEFINITION  *** SEQUENCING IN PROGRESS *** Caenorhabditis elegans cosmid BA2;
            HTGS phase 1.
ACCESSION   Z83098
NID         g1694982
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Nematoda;
            Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditidae;
            Caenorhabditis.
REFERENCE   1  (bases 1 to 31197)
  AUTHORS   Unknown.
  TITLE     Direct Submission
```

```
JOURNAL   Submitted (26-NOV-1996) Nematode Sequencing Project, Sanger Centre,
          Hinxton, Cambridge CB10 1RQ, England and Department of Genetics,
          Washington University, St. Louis, MO 63110, USA. E-mail:
          jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT   Order of segments is not known; 800 n's separate segments.
          Cosmid=BA2; Contig ID=00084; Length=1540; Order=Unknown;
          Status=Cambridge-
          Unfinished; Author=Sulston JE; LastModification=Nov 22 10:09
          Cosmid=BA2; Contig ID=00003; Length=1938; Order=Unknown;
          Status=Cambridge-
          Unfinished; Author=Sulston JE; LastModification=Nov 22 10:09
          Cosmid=BA2; Contig ID=00004; Length=2167; Order=Unknown;
          Status=Cambridge-
          Unfinished; Author=Sulston JE; LastModification=Nov 22 10:09
          Cosmid=BA2; Contig ID=00151; Length=1092; Order=Unknown;
          Status=Cambridge-
          Unfinished; Author=Sulston JE; LastModification=Nov 22 10:09
          Cosmid=BA2; Contig ID=00032; Length=1270; Order=Unknown;
          Status=Cambridge-
          Unfinished; Author=Sulston JE; LastModification=Nov 22 10:09
          Cosmid=BA2; Contig ID=00053; Length=2559; Order=Unknown;
          Status=Cambridge-
          Unfinished; Author=Sulston JE; LastModification=Nov 22 10:09
          Cosmid=BA2; Contig ID=00078; Length=1448; Order=Unknown;
          Status=Cambridge-
          Unfinished; Author=Sulston JE; LastModification=Nov 22 10:09
          Cosmid=BA2; Contig ID=00035; Length=1077; Order=Unknown;
          Status=Cambridge-
          Unfinished; Author=Sulston JE; LastModification=Nov 22 10:09
          Cosmid=BA2; Contig ID=00124; Length=1154; Order=Unknown;
          Status=Cambridge-
          Unfinished; Author=Sulston JE; LastModification=Nov 22 10:09
          Cosmid=BA2; Contig ID=00160; Length=4080; Order=Unknown;
          Status=Cambridge-
          Unfinished; Author=Sulston JE; LastModification=Nov 22 10:09
          Cosmid=BA2; Contig ID=00043; Length=1156; Order=Unknown;
          Status=Cambridge-
          Unfinished; Author=Sulston JE; LastModification=Nov 22 10:09
          Cosmid=BA2; Contig ID=00144; Length=1134; Order=Unknown;
          Status=Cambridge-
          Unfinished; Author=Sulston JE; LastModification=Nov 22 10:09
          Cosmid=BA2; Contig ID=00080; Length=1346; Order=Unknown;
          Status=Cambridge-
          Unfinished; Author=Sulston JE; LastModification=Nov 22 10:09.
          ***                                                        ***
          *** WARNING: Phase 1 High Throughput Genome Sequence ***
          ***                                                        ***
          * This sequence is unfinished. When sequencing is complete,
          * the sequence data presented in this record will be replaced
          *by a single finished sequence with the same accession number.
FEATURES            Location/Qualifiers
```



```
     source          1. .31197
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /clone="BA2"
                     /chromosome="V"
BASE COUNT     7029 a    3762 c    3598 g    7194 t    9614 others
ORIGIN

  Query Match           14.3%;  Score 17;  DB 17;  Length 31197;
  Best Local Similarity 100.0%;  Pred. No. 3.49e+00;
  Matches    17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Db  13944 TTATTCATTTCTCAAGA 13960
          |||||||||||||||||
Cp     26 TTATTCATTTCTCAAGA 10


RESULT   18
LOCUS       CEC06C3     32043 bp    DNA             INV       21-SEP-1998
DEFINITION  Caenorhabditis elegans cosmid C06C3, complete sequence.
ACCESSION   Z36719
NID         g3217066
KEYWORDS    HTG; ATP dependent protease; LON.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 32043)
  AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
  TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
  JOURNAL   Nature 368 (6466), 32-38 (1994)
  MEDLINE   94150718
REFERENCE   2  (bases 1 to 32043)
  AUTHORS   Berks,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-NOV-1996) Nematode Sequencing Project, Sanger Centre,
            Hinxton, Cambridge CB10 1RQ, England and Department of Genetics,
            Washington University, St. Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     On Jun 13, 1998 this sequence version replaced gi:531558.
```



```
            Coding sequences below are predicted from computer analysis, using
            predictions from Genefinder (P. Green, U. Washington), and other
            available information.
            For a graphical representation of this sequence and its analysis
            see:-
            http://webace.sanger.ac.uk/cgi-
            bin/display?db=wormace&class=Sequence &object=C06C3
            Current sequence finishing criteria for the C. elegans genome
            sequencing consortium are that all bases are either sequenced
            unambiguously on both strands, or on a single strand with both a
            dye primer and dye terminator reaction, from distinct subclones.
            Exceptions are indicated by an explicit note.
            IMPORTANT:  This sequence is NOT necessarily the entire insert of
            the specified clone.  It may be shorter because we only sequence
            overlapping sections once, or longer because we arrange for a small
            overlap between neighbouring submissions.
            IMPORTANT: This sequence is not the entire insert of clone C06C3.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we arrange for a small overlap between
            neighbouring submissions.
            The true left end of clone C06C3 is at 11451 in
            sequence Z47809.
            The true right end of clone C06C3 is at 104 in
            sequence AL023845.
            The true right end of clone F42A8 is at 16760 in this sequence. The
            start of this sequence (1. .56) overlaps with the end of sequence
            Z47809.
            The end of this sequence (31940. .32043) overlaps with the start of
            sequence AL023845.
FEATURES            Location/Qualifiers
     source         1. .32043
                    /organism="Caenorhabditis elegans"
                    /db_xref="taxon:6239"
                    /chromosome="II"
                    /clone="C06C3"
     gene           complement(7334. .10527)
                    /gene="C06C3.3"
     CDS            complement(join(7334. .7542,7595. .8320,8623. .8893,
                    9335. .9559,9628. .9726,9772. .9939,10474. .10527))
                    /gene="C06C3.3"
                    /codon_start=1
                    /db_xref="PID:g531561"
                    /db_xref="SPTREMBL:Q17711"
                    /translation="MSENLHTIIHHVDFFPLRIRVFRNGDMFDRGRILVVTRKQFKHW
                    IVFLDALTDLLRTATAVRRLFTTQGIPVHHFDDLETNGETVAVETGPFVNVPYGQARF
                    TLRLKHYHPIIRAPELIPTFLHSAESMDIYLKKEGYGTITGTSNKCIRKCTCIPFPSS
                    FSGNTFSQVTLIRLMEFFREVQIGGSMEHLNELGTASWLGKNEKWWMGERVKLETQSQ
                    TLAQKSERDEQIPRLDIKEVDGGSTTSEFMRNGEKSKSTTKLPRLLNSNRSFRREPVS
                    AQKPPSEEPRSTSENLPPIDLKSIQISQNSQSTSQKKDSDTISIKTEEEVISILKPEK
                    TTTESIVSSKISRKNAPKDDDEVRFHRSYQVTETTYTTRNSDETDRRKPENVEKTVES
                    KNDRSVQDKSEKIEKKVPEKLSKPQKTETTKSKERAKWADQGNTTLKREKSIEKSTRQ
```



```
                    GKIDKSIEKLIDKSIDNTEKLSSQKPKIDMAKGELRKFDVDINEIPIRFEGPMVNKSR
                    INTSRIVTSRSTTNLFVDEIVNEDELDDVIERAQLWDLNRRRQNMRQQHTQTSGSLCR
                    SISAPRRRVVLFDPDFD"
     gene           11753. .13266
                    /gene="C06C3.4"
     CDS            join(11753. .11827,12175. .12340,12387. .12456,12519. .12696,
                    12877. .12978,13204. .13266)
                    /gene="C06C3.4"
                    /codon_start=1
                    /db_xref="PID:g531567"
                    /db_xref="SPTREMBL:Q17717"
                    /translation="MTFQNILLVFHIVFVSVVFSADIDFPLICDKGYRVSSINRAKTA
                    YSLLGSLTVECEQIALAESTNCAPQQSVPKCTGLLEGCTGNTWLGGFHVYLLENTTQA
                    AVLDPVCCSSPSVQIDAGSCINDQLNTGTRDFSHSIVADLVLKFKLNFHNPQKYRCTE
                    DGNPNVKSAHVNVELSNAQMGQVQFELFIKNFFFHVHLAIFLTSFQFLGKLIFLRCQ"
     gene           14377. .17384
                    /gene="C06C3.9"
     CDS            join(14377. .14430,15561. .15726,15773. .15842,15905. .16082,
                    16265. .16366,16590. .16774,16856. .17384)
                    /gene="C06C3.9"
                    /note="cDNA EST yk208g3.5 comes from this gene"
                    /codon_start=1
                    /db_xref="PID:g531566"
                    /db_xref="SPTREMBL:Q17716"
                    /translation="MTEVLQIDVLLADNRVTVALICDKGFRVSSINRAKTAYSLLGSL
                    TVECEQIALAESTNCAPQQSVPKCTGLLEGCTGNTWLGGFHVYLLENTTQAAVLDPVC
                    CSSPSVQIDAGSCINDQLNTGTRDFSHSIVADLVLKFKLNFHNPQKYRCTEDGNPNVK
                    NAHGNVELSNAQMEQVRFELSTKSFFFPHVHLAIFLTSFHIITFIFYLLKFYFDINLLN
                    FTISKKTDFLTVSMRALPPEIITKITNKLDPIDRLIARRVSRKIRAIIDYQVFQISTI
                    SMKIGDKKSTLKIDASTIEYLPVGDDCLAKYGSHEKKVVGKSHLDIILNDFSTLVSNL
                    KFRLEEPTFLLGSFWISEVRTNCCNYLKNILRTWRNQNTKIVSMNSISINQVIPILSQ
                    FNAEKLEELHLNILYQNWDLINEDLEIVKLDQWKN"
     gene           complement(18246. .18665)
                    /gene="C06C3.5"
     CDS            complement(join(18246. .18392,18438. .18665))
                    /gene="C06C3.5"
                    /codon_start=1
                    /db_xref="PID:g531562"
                    /db_xref="SPTREMBL:Q17712"
                    /translation="MKVGLRMRAIVQVAKKQEGIESAECSEMNKDDENIIDKDDQRSS
                    RASSTDEHPGAKPTLIHHTTPSLDCQPTKHKSIQSDFSEKCINSCIKQQGSLTSIRNV
                    RRLCTRRFSTLFRHLIKKIMFV"
     gene           20375. .23045
                    /gene="C06C3.6"
     CDS            join(20375. .20496,21390. .21776,21826. .21974,22018. .22214,
                    22261. .22567,22876. .23045)
                    /gene="C06C3.6"
                    /codon_start=1
                    /db_xref="PID:g531563"
                    /db_xref="SPTREMBL:Q17713"
                    /translation="MASGHVWSRLKIFPEFWYSTNKIENPVKMVAIQNKTNRKPKFHL
```

```
                     QYLFVFLITALATNNAQKHLPSSLLTRQFSYFIHFFFYIKLIFSRSFIMTTTTQLFEY
                     YATNYSQICQKLSYERSYIETPEFLSLCAKIIGILGVPTNFFASYIITTHTPKSMTHL
                     KLCLLNLKFWTLLIDLIYTVFLIPFVFTPIFANGYIGIPGKVPGIPSEIQFYFAMACY
                     GGIASAGILIFENRQHHNIPKGHKFRIQMCAFFVLLIIFNVLIGSSVMLMAIWLRADS
                     NELKFKFLKINPCPDPLYFTPGTFAVDSQRNEPSICIVIVLTVVFIQYTFFISHCIWY
                     IYSEDAVRYSKSTRKLQKMFLYASFSQLGIFVTVFVLPLGIFAMVLTTGYKNQGLLNI
                     CNLIIPTIGMNTSIGLVTMYKPYRDYPIGIFKEGGNGIVPPPSAPTKSVVF"
     gene            23704. .24432
                     /gene="C06C3.7"
     CDS             join(23704. .23761,23992. .24149,24241. .24432)
                     /gene="C06C3.7"
                     /codon_start=1
                     /db_xref="PID:g531564"
                     /db_xref="SPTREMBL:Q17714"
                     /translation="MSAKNVLKMANRNSANKRDILGDLLSLSSNKTVNLNEMAHGQGK
                     RKMASIDGQKENPKRIKMSQHMTAKEVKIPLLRTPDEEARKIEKLRREARIKRSKEVH
                     YDAAEQQQIMMNYEWTVQMEEQEKLRKKVEKKH"
     gene            complement(24744. .29056)
                     /gene="C06C3.8"
     CDS             complement(join(24744. .24926,24972. .25101,25205. .25532,
                     26101. .26405,26452. .26513,28161. .28319,28808. .29056))
                     /gene="C06C3.8"
                     /note="weak similarity to E. coli LON ATP dependent
                     protease (Swiss Port accession number P08177)"
                     /codon_start=1
                     /db_xref="PID:g531565"
                     /db_xref="SPTREMBL:Q17715"
                     /translation="MTQEMQQSMELSHEIVIANGHLFGIQGFTCNITLSLEQIGRTYN
                     GVSMALPFVLLIISAIKKNSLRKDYVATGDVSLAGAVLTVDYINNKIVGAINAGLKGV
                     VIPAENGEDVKPFLKTQIKVIEVTSIVDAWSTMITDQKNEMTSMGECIGEVIREDYLV
                     INSSDNCRRPLYKQPCPYPPEELPKPDPERKMSHTETYLQLSESPKLGVQKIGSKDVLL
                     KKVILNEISKSPAIWNTYKANGIYSEVAVNVFKRTGKLLNIKLIQSIPKLAKDGLRNR
                     LRDSITEKLHNEADTENHLWQWELYEFIRFYREITEGWELRMRALVAKKHEGQLRVRE
                     SAECSDMIKDDKNLININDDERNMKLDHVRTDISDTSIVQTCMDSSRLPSNDEHPGAKP
                     TLIHHTTPSLDCQPTKHESINSDFSKEMDQLVNQATRIAREHPERAETLLKALFATVS
                     TFDQEGYVCVEDKKFTEKQSK"
     gene            join(Z47809:18084. .18164,Z47809:19475. .19675,
                     Z47809:20240. .20884,Z47809:21199. .21416,200. .529,
                     1124. .1307,1375. .1511,1896. .2159,3049. .3232,3792. .3935,
                     4129. .4329,5625. .5807,5851. .5964,6014. .6181,6724. .6810)
                     /gene="C06C3.1"
     CDS             join(Z47809:18084. .18164,Z47809:19475. .19675,
                     Z47809:20240. .20884,Z47809:21199. .21416,200. .529,
                     1124. .1307,1375. .1511,1896. .2159,3049. .3232,3792. .3935,
                     4129. .4329,5625. .5807,5851. .5964,6014. .6181,6724. .6810)
                     /gene="C06C3.1"
                     /note="This predicted protein contains 2 ankyrin-like
                     repeats; cDNA EST EMBL:D34814 comes from this gene; cDNA
                     EST EMBL:D37685 comes from this gene; cDNA EST EMBL:D36298
                     comes from this gene; cDNA EST EMBL:D67967 comes from this
                     gene; cDNA EST EMBL:T02424 comes from this gene; cDNA EST
                     EMBL:D33452 comes from this gene; cDNA EST EMBL:D37524
                     comes from this gene; cDNA EST EMBL:D32820 comes from this
                     gene; cDNA EST EMBL:D73033 comes from this gene; cDNA EST
                     EMBL:D70153 comes from this gene; cDNA EST EMBL:D74714
                     comes from this gene; cDNA EST EMBL:D71957 comes from this
                     gene; cDNA EST EMBL:C11184 comes from this gene; cDNA EST
                     EMBL:C13200 comes from this gene; cDNA EST EMBL:C10946
                     comes from this gene; cDNA EST EMBL:C12936 comes from this
                     gene; cDNA EST EMBL:C11936 comes from this gene; cDNA EST
                     EMBL:C12367 comes from this gene; cDNA EST yk489d4.5 comes
                     from this gene"
                     /codon_start=1
                     /db_xref="PID:e1321732"
                     /db_xref="PID:g3641926"
...
Note: remainder of annotations omitted.

  Query Match          14.3%;  Score 17;  DB 18;  Length 32043;
  Best Local Similarity 100.0%;  Pred. No. 3.49e+00;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  14602 TCTTGAGAAATGAATAA 14618
          |||||||||||||||||
Qy     10 TCTTGAGAAATGAATAA 26


RESULT   19
LOCUS       CET07C4      33477 bp    DNA              INV       22-SEP-1998
DEFINITION  Caenorhabditis elegans cosmid T07C4, complete sequence.
ACCESSION   Z29443
NID         g1067051
KEYWORDS    HTG.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 33477)
  AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
  TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
  JOURNAL   Nature 368 (6466), 32-38 (1994)
```



```
    MEDLINE   94150718
  REFERENCE   2  (bases 1 to 33477)
    AUTHORS   Berks,M.
    TITLE     Direct Submission
    JOURNAL   Submitted (06-NOV-1996) Nematode Sequencing Project, Sanger Centre,
              Hinxton, Cambridge CB10 1RQ, England and Department of Genetics,
              Washington University, St. Louis, MO 63110, USA. E-mail:
              jes@sanger.ac.uk or rw@nematode.wustl.edu
    COMMENT   On Nov 21, 1995 this sequence version replaced gi:443827.
              Coding sequences below are predicted from computer analysis, using
              predictions from Genefinder (P. Green, U. Washington), and other
              available information.
              For a graphical representation of this sequence and its analysis
              see:-
              http://webace.sanger.ac.uk/cgi-
              bin/display?db=wormace&class=Sequence &object=T07C4
              Current sequence finishing criteria for the C. elegans genome
              sequencing consortium are that all bases are either sequenced
              unambiguously on both strands, or on a single strand with both a
              dye primer and dye terminator reaction, from distinct subclones.
              Exceptions are indicated by an explicit note.
              IMPORTANT:  This sequence is NOT necessarily the entire insert of
              the specified clone.  It may be shorter because we only sequence
              overlapping sections once, or longer because we arrange for a small
              overlap between neighbouring submissions.
              IMPORTANT: This sequence is not the entire insert of clone T07C4.
              It may be shorter because we only sequence overlapping sections
              once, or longer because we arrange for a small overlap between
              neighbouring submissions.
              The true left end of clone T07C4 is at 1 in this sequence. The true
              left end of clone C38H2 is at 33374 in this sequence. The true
              right end of clone T07A5 is at 8844 in this sequence. The start of
              this sequence (1. .101) overlaps with the end of sequence Z48055.
              The end of this sequence (33374. .33477) overlaps with the start of
              sequence Z35641.
    FEATURES            Location/Qualifiers
         source         1. .33477
                        /organism="Caenorhabditis elegans"
                        /db_xref="taxon:6239"
                        /chromosome="III"
                        /clone="T07C4"
         gene           complement(1982. .8964)
                        /gene="nex-2"
         CDS            complement(join(1982. .2497,3423. .3868,5024. .5413,
                        6234. .6288,7454. .7596,8071. .8344,8758. .8964))
                        /gene="nex-2"
                        /note="similar to Annexin; cDNA EST EMBL:C10640 comes from
                        this gene; cDNA EST EMBL:C12433 comes from this gene; cDNA
                        EST yk192f7.5 comes from this gene; cDNA EST yk318c1.5
                        comes from this gene; cDNA EST yk494a12.3 comes from this
                        gene"
```



```
                        /codon_start=1
                        /db_xref="PID:g443836"
                        /db_xref="SPTREMBL:Q27512"
                        /translation="MNSTNLHLVQEPLSRRELDVREELQDCVDLFSAEEFEEKVDLVL
                        FADARLNNSKFRKLCELVKESRSHEMLSEEQLATLYDCATTSTRLPNDFADDQEDLKN
                        IIRYGELFKACHAINSTDFIQKSEDLKDEEKVALERIVEQKLAESARKEKDIAWNVNI
                        VVANSYVAKSLRPVINIQMPTVGGDTNFEFDIDSFAQFRQQLAQAVLAPQHNVASQIR
                        HVWDSGRPGYPNQQPSYGGYGQPPQQPQPGYGNGSYDPYGQPQQQPYPGGGGQPPYPGSN
                        SNQGGGGYPGGGGAPYPGGSGGYPQAPQYGFSGGGGSAPQPNQGGYPPQQQQYPPQQG
                        YQQGGQQQGGGGFFPNQGYGQPVMIGTPSVFPVQGFNSNADAEVLRKAMKGLGCNNSK
                        VISILCQRTNWQRQEISKAFKVMYGKDLIKELKGELHGDFEDLILALMDAPAIYDAEQ
                        LHRAMEGLGTKESVLIEIMTSRTNAQIQQVRDAYKMLFKKDLERDLIGETSGHFKRLL
                        VSLCAGGRDESSQTDGLRANQDARRLLQAGEKRLGTDESTFNAILASQNFSQLRLVFE
                        EYQKASNHSIEKAIEFEPSGDIRDGLLAVIAVIRNRPAYFAKLLHDSMKGLGTRDNDL
                        IRLCVTRAEYDMGDIRNMFQSLYRTSLENMIKGDCSGAYKEGLIALVNGNRG"
         gene           10144. .10856
                        /gene="cyt-1"
         CDS            join(10144. .10230,10342. .10704,10758. .10856)
                        /gene="cyt-1"
                        /note="similar to Succinate dehydrogenase cytochrome b
                        chain; cDNA EST yk449h7.5 comes from this gene; cDNA EST
                        yk473e2.5 comes from this gene"
                        /codon_start=1
                        /db_xref="PID:g443834"
                        /db_xref="SWISS-PROT:P41956"
                        /translation="MINIPTAILCRLGARSSISRSPGTSIVTKSEAKTPIQKPGWEYL
                        LKQRSKHRPIAPHLTVYQPQLTWMLSGFHRRISGCVMAGTLLVGGIGFAVLPFDFTAPV
                        DFIRSWNLPCAVTAVFKYIIAFPIIFHTLNGIRFLGFDLAAGVNNVGQIYKSGYLVSG
                        LSAILALAIVFNSCQNKSNKTA"
         gene           11095. .12946
                        /gene="ced-9"
         CDS            join(11095. .11289,11338. .11598,12130. .12279,12767. .12946)
                        /gene="ced-9"
                        /note="similar to Apoptosis regulator proteins, Bcl-2
                        family; cDNA EST EMBL:D73048 comes from this gene; cDNA
                        EST yk503f3.3 comes from this gene; cDNA EST yk312c5.3
                        comes from this gene; cDNA EST yk312c5.5 comes from this
                        gene; cDNA EST yk201b10.3 comes from this gene; cDNA EST
                        yk201b10.5 comes from this gene"
                        /codon_start=1
                        /db_xref="PID:g443835"
                        /db_xref="SWISS-PROT:P41958"
                        /translation="MATGEMKEPLGIKGTEPTDPGINSDAQDLPSPSRQASTRRMSIG
                        ESIDGKINDWEEPRLDIEGFVVDYFTHRIRQNGMEWFGAPGLPCGVQPEHEMMRVMGT
                        IFEKKRAENFETFCEQLLAVPRISFSLYQDVVRTVGNAQTDQCPMSYGRLIGLISFGG
                        FVAAKMMESVELQGQVRNLFVYTSLFIKTRIRNNWKEHNRSWDDFMTLGKQMKEDYER
                        AEAEKVGRRKQNRRWSMIGAGVTAGAIGIVGVVVCGRMMFSLK"
         gene           13739. .15116
                        /gene="T07C4.5"
         CDS            join(13739. .13883,14869. .15116)
                        /gene="T07C4.5"
```

```
                     /note="similar to Transthyretin-like family; cDNA EST
                     CEMSG06P comes from this gene; cDNA EST CEMSG06R comes
                     from this gene; cDNA EST yk445d4.3 comes from this gene;
                     cDNA EST yk445d4.5 comes from this gene"
                     /codon_start=1
                     /db_xref="PID:g1067052"
                     /db_xref="SPTREMBL:Q22288"
                     /translation="MRALLFTSVVLLALAPVEAKKQTITVKGTTICNKKRIQAEVTLW
                     EKDTLDPDDKLASMQSNKEGEPSLTGSDDEITSISPYLIITHNCNVKKAGCKRVSEYL
                     IPKEKIGGTYDMTYVTLDILSAKDKEKC"
     gene            complement(15931. .19698)
                     /gene="tbx-9"
     CDS             complement(join(15931. .16287,16714. .16866,16924. .16999,
                     17425. .17724,18679. .18780,19553. .19698))
                     /gene="tbx-9"
                     /note="Brachyury protein (also known as tbx-9); cDNA EST
                     EMBL:T00493 comes from this gene; cDNA EST EMBL:D75158
                     comes from this gene; cDNA EST EMBL:D66339 comes from this
                     gene; cDNA EST EMBL:D72308 comes from this gene; cDNA EST
                     yk426d6.3 comes from this gene; cDNA EST yk426d6.5 comes
                     from this gene; cDNA EST yk337c3.3 comes from this gene;
                     cDNA EST yk337c3.5 comes from this gene; cDNA EST
                     yk196g6.3 comes from this gene; cDNA EST yk196g6.5 comes
                     from this gene"
                     /codon_start=1
                     /db_xref="PID:g443833"
                     /db_xref="SWISS-PROT:Q22289"
                     /translation="MNVSTLFSPPRGRHAYNEEFAPVEEDIHLINRYGIELINLKFYF
                     VREGIYQFLHLQSILSVFTQKSPLQLTGSKILQRMLFWRQKMSKVKVSIEGSQETLWK
                     IFHAEVNEMIVTKNGRKLFPKLEYIVEGLDENKLYAIMLQLQPVGESRFKFSGGKWQE
                     TGKAEKQVDAKKMWHADGVRKGSDWMWSSICFDRVKISNYSESNNASMIYLNSMHKYI
                     PVLTIYESPSESPFCVPQSSNQIVATAKFPHTEFIAVTAYQNQKITDLKIKHNSFAKG
                     FRDGNLSRKRRSPSYSDGSNSQSPSPKSRSPPEVAPLQSMPPINPFLFYPPHMLSENL
                     PVQFPPFAPPFLSPLPSTPSSSSSELSIVKEEDQEVEEDIDIVG"
     gene            215524. .25707
                     /gene="tbx-8"
     CDS             join(21524. .21754,23214. .23312,23441. .23745,24128. .24203,
                     24259. .24411,25045. .25458,25618. .25707)
                     /gene="tbx-8"
                     /note="Brachyury protein (also known as tbx-8); cDNA EST
                     EMBL:D75158 comes from this gene; cDNA EST yk325e8.3 comes
                     from this gene; cDNA EST yk325e8.5 comes from this gene"
                     /codon_start=1
                     /db_xref="PID:e218818"
                     /db_xref="PID:g1163131"
                     /db_xref="SWISS-PROT:Q22292"
                     /translation="MSVVLKRSRQTEEIKREVNQRNSGMPLLIGLHRDIGDPDKAAES
                     QPKQRCSWNKISEKAKDIRWAIKNLPALIGDRQHPAKHRKTTPTSVIATVPRGPPSGR
                     QEWGRGIQRKTMTDQVTVTLREDQDKLWNLPHYHKNEMIVTKSGRKMFPKLEYVVRGL
                     TPNKLYAMMLHIEPSDDHRYKFSSGEWVKSGRAEKHREPKKLWHADGVRSGKEWMTNP
                     VCPDRVKITNCAESTNASMIFLHSMHKYTPVMSIYESPSESPPSVPQPSTRLVTSVRL
```



```
                     TYTEFIAVTAYQNDAVIKLKIKFNPFAKGFREGSQSDRKRNSPSADDSTTDESSSQVS
                     SPQPKKSRTVSVSPPLMPRILPTMIPPPPAINPLFYSLPYPSHLAATGNFPPVPFQFP
                     FGLPCFSPMSFPTPPPSLKNVKKEEQEDIEQEINTSRQAGKNALLRAFTKNRSSIAHA
                     NPLKK"
     gene            28088. .30619
                     /gene="T07C4.3"
     CDS             join(28088. .28104,28300. .28411,28468. .28625,28669. .28817,
                     28868. .29070,29115. .29404,29452. .29941,29994. .30131,
                     30179. .30511,30560. .30619)
                     /gene="T07C4.3"
                     /note="cDNA EST EMBL:C12165 comes from this gene; cDNA EST
                     EMBL:C13043 comes from this gene; cDNA EST yk200f1.5 comes
                     from this gene; cDNA EST yk272f11.5 comes from this gene;
                     cDNA EST yk320f7.5 comes from this gene; cDNA EST
                     yk399c5.5 comes from this gene; cDNA EST yk411d2.5 comes
                     from this gene; cDNA EST yk416g3.5 comes from this gene"
                     /codon_start=1
                     /db_xref="PID:g443830"
                     /db_xref="SPTREMBL:Q22290"
                     /translation="MTDSQSTQMGRGRHAKRNKSTASVGSKKNRTQSTASGKFQPSVD
                     GFTTPIVAVSRLGKQGQEFTWHDLRKVYHKETQRQLTAEELNAIFSSKNLTKREIPEL
                     PKVKRFMTSMDGSMMCRFRLIADPQEVYDDDLPPLMSTSVETDVITEDTSSDRSQAVS
                     PSSGNPIAISTDCDDKAAGPTSDLSKSDNSPSQRSLAQQSSMTDDEGWVSGKDVRSEN
...
Note: remainder of annotations omitted.

  Query Match           14.3%;  Score 17;  DB 18;  Length 33477;
  Best Local Similarity 100.0%;  Pred. No. 3.49e+00;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   7787 TCAGTTTTGTCTAGTGT 7803
           |||||||||||||||||
Cp     70 TCAGTTTTGTCTAGTGT 54


RESULT   20
LOCUS       AC004135    73805 bp    DNA             PLN       15-MAY-1998
DEFINITION  Genomic sequence for Arabidopsis thaliana BAC T17H7, complete
            sequence.
ACCESSION   AC004135
NID         g3133272
KEYWORDS    HTG.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Charophyta/Embryophyta group;
            Embryophyta; Tracheophyta; seed plants; Magnoliophyta;
            eudicotyledons; Rosidae; Capparales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 73805)
  AUTHORS   Buehler,E., Shinn,P., Dewar,K., Feng,J., Kim,C., Li,Y., Sun,H.,
            Conway,A., Conway,A., Kurtz,D., Oji,O., Shen,Y.K., Toriumi,M.,
            Vysotskaia,V., Yu,G., Davis,R.W., Federspiel,N.A., Theologis,A. and
```



```
            Ecker,J.R.
  TITLE     Genomic sequence for Arabidopsis thaliana BAC T17H7
  JOURNAL   Unpublished (1998)
REFERENCE   2  (bases 1 to 73805)
  AUTHORS   Ecker,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-FEB-1998) Arabidopsis thaliana Genome Center,
            Department of Biology, University of Pennsylvania, 38th Street and
            Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
REFERENCE   3  (bases 1 to 73805)
  AUTHORS   Ecker,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-MAY-1998) Arabidopsis thaliana Genome Center,
            Department of Biology, University of Pennsylvania, 38th Street and
            Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
COMMENT     On May 15, 1998 this sequence version replaced gi:2887490.
FEATURES             Location/Qualifiers
     source          1. .73805
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /chromosome="1"
BASE COUNT    23939 a  12539 c  12659 g  24668 t
ORIGIN

  Query Match           14.3%;  Score 17;  DB 24;  Length 73805;
  Best Local Similarity 100.0%;  Pred. No. 3.49e+00;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  33527 TATTCATTTCTCAAGAG 33543
          |||||||||||||||||
Cp     25 TATTCATTTCTCAAGAG 9


RESULT   21
LOCUS       AC005085    85779 bp    DNA             HTG       12-JUN-1998
DEFINITION  *** SEQUENCING IN PROGRESS *** Homo sapiens clone RG302D02; HTGS
            phase 1, 9 unordered pieces.
ACCESSION   AC005085
NID         g3212906
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 85779)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 85779)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
```



```
  JOURNAL   Submitted (12-JUN-1998) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     ***                                                          ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                          ***
            * This sequence is unfinished. It consists of 9 contigs for
            * which the order is not known; their order in this record is
            * arbitrary. In some cases, the exact lengths of the gaps
            * between the contigs are also unknown; these gaps are  presented
            * as runs of N as a convenience only. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
            *        1     1560: contig of 1560 bp in length
            *     1561     1577: gap of unknown length
            *     1578     3520: contig of 1943 bp in length
            *     3521     3537: gap of unknown length
            *     3538     5533: contig of 1996 bp in length
            *     5534     5550: gap of unknown length
            *     5551     9231: contig of 3681 bp in length
            *     9232     9248: gap of unknown length
            *     9249    18919: contig of 9671 bp in length
            *    18920    18936: gap of unknown length
            *    18937    29775: contig of 10839 bp in length
            *    29776    29792: gap of unknown length
            *    29793    46124: contig of 16332 bp in length
            *    46125    46141: gap of unknown length
            *    46142    66136: contig of 19995 bp in length
            *    66137    66153: gap of unknown length
            *    66154    85779: contig of 19626 bp in length.
FEATURES             Location/Qualifiers
     source          1. .85779
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RG302D02"
BASE COUNT    27072 a  17492 c  15886 g  25193 t    136 others
ORIGIN

  Query Match           14.3%;  Score 17;  DB 17;  Length 85779;
  Best Local Similarity 100.0%;  Pred. No. 3.49e+00;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  30402 ACACTAGACAAAACTGA 30418
          |||||||||||||||||
Qy     54 ACACTAGACAAAACTGA 70


RESULT   22
LOCUS       ATAC004684  93845 bp    DNA             PLN       17-JUN-1998
DEFINITION  Arabidopsis thaliana chromosome II BAC F13M22 genomic sequence,
            complete sequence.
```

```
ACCESSION   AC004684
NID         g3236234
KEYWORDS    HTG.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Charophyta/Embryophyta group;
            Embryophyta; Tracheophyta; seed plants; Magnoliophyta;
            eudicotyledons; Rosidae; Capparales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 93845)
  AUTHORS   Rounsley,S.D., Kaul,S., Lin,X., Ketchum,K.A., Crosby,M.L.,
            Brandon,R.C., Sykes,S.M., Mason,T.M., Kerlavage,A.R., Adams,M.D.,
            Somerville,C.R. and Venter,J.C.
  TITLE     Arabidopsis thaliana chromosome II BAC F13M22 genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 93845)
  AUTHORS   Rounsley,S.D. and Lin,X.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-MAY-1998) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, rounsley@tigr.org
REFERENCE   3  (bases 1 to 93845)
  AUTHORS   Rounsley,S.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JUN-1998) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
COMMENT     On Jun 17, 1998 this sequence version replaced gi:3172168.
            Address all correspondence to:
            Steve Rounsley
            The Institute for Genomic Research
            9712 Medical Center Dr,
            Rockville, MD 20850,
            USA
            e-mail: rounsley@tigr.org
            BAC clone F13M22 is from Arabidopsis chromosome II and is near the
            molecular marker ve018.
            The orientation of the sequence is from SP6 to T7 end of the BAC
            clone.
            Genes were identified by a combination of three methods: Gene
            prediction programs including GRAIL (available by anonymous ftp
            from arthur.epm.ornl.gov), Genefinder (Phil Green, University of
            Washington), Genscan (Chris Burge,
            http://gnomic.stanford.edu/~chris/GENSCANW.html), and NetPlantGene
            (http://www.cbs.dtu.dk/netpgene/cbsnetpgene.html), searches of the
            complete sequence against a peptide database and the Arabidopsis
            EST database at TIGR (http://www.tigr.org/tdb/at/at.html).
            Annotated genes are named to indicate the level of evidence for
            their annotation. Genes with similarity to other proteins are named
            after the database hits. Genes without significant peptide
            similarity but with EST similarity are named as 'unknown' proteins.
            Genes without protein or EST similarity, that are predicted by more
            than two gene prediction programs over most of their length are
            annotated as 'hypothetical' proteins. Genes encoding tRNAs are
```



```
            predicted by tRNAscan-SE (Sean Eddy,
            http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are
            identified by repeatmasker (Arian Smit,
            http://ftp.genome.washington.edu/RM/RepeatMasker.html). Regions of
            genomic sequence that are not annotated as genes but have predicted
            exons by GRAIL are annotated as misc features.
FEATURES            Location/Qualifiers
     source         1. .93845
                    /organism="Arabidopsis thaliana"
                    /cultivar="Columbia"
                    /db_xref="taxon:3702"
                    /chromosome="II"
                    /map="ve018"
                    /clone="F13M22"
     misc_feature   complement(10. .29)
                    /note="exon predicted by xgrail, quality marginal"
     misc_feature   168. .249
                    /note="exon predicted by xgrail, quality good_shadowexon"
     mRNA           <464. .>619
                    /gene="F13M22.1"
     gene           <464. .>619
                    /note="predicted by genscan"
                    /gene="F13M22.1"
     CDS            <464. .619
                    /gene="F13M22.1"
                    /note="hypothetical protein, 5' partial"
                    /codon_start=1
                    /db_xref="PID:g3236260"
                    /translation="SRGDVRVEQRRTHKDGEDKRLLDPATRRRKLETNAIIASGTKGF
                    EIGNGNL"
     mRNA           complement(join(835. .1212,1373. .1513,1600. .1693,
                    1789. .1845,1933. .1994,2162. .2407,2492. .2548,2640. .2697,
                    2855. .3038,3239. .3405,3479. .3531,3790. .3871,3980. .4045,
                    4167. .4353,4486. .4549,4633. .4842,4938. .5006,5088. .>5654))
                    /gene="F13M22.2"
     gene           complement(835. .>5654)
                    /gene="F13M22.2"
     CDS            complement(join(1099. .1212,1373. .1513,1600. .1693,
                    1789. .1845,1933. .1994,2162. .2407,2492. .2548,2640. .2697,
                    2855. .3038,3239. .3405,3479. .3531,3790. .3871,3980. .4045,
                    4167. .4353,4486. .4549,4633. .4842,4938. .5006,5088. .5654))
                    /gene="F13M22.2"
                    /note="unknown protein"
                    /codon_start=1
                    /db_xref="PID:g3236235"
                    /translation="MPKEENGQLKRDRLDDDTDEGNKGDHFPSKKQAKEASNDDITSE
                    ISNPVASPVESTSLFRDVSSQPVKSGLVECSGSDPGSEETVSDDASVVGSSQTEQSSD
                    VLPSRPVLEIPKHLSSTGITKITFKLSKPKKEFDDLPLIKDHTWDAGVVKMPKKKIVS
                    LSYPSNVKKLLETGILEGARVKYISTPPVRQLLGIIHSGGYLCGCTTCNFSKVLSAYE
                    PEQHAGAKTRHPNNHIFLENRRAVYNIVQELKTAPRVVLEEVIRNVAGSALNEEGLRA
                    WKGDGQQVSNRVIACLTGIILRIIPLSVNALLSYHFVLLSETDSCESCSYLGPGLDES
```



```
                    QSLTPCSVENHYFSEKTYAKDTLDEPKRIAKKLTSHVSGTGCHKKVSEGSNRKRDNDL
                    HRLLFMPNGLPDGTELAYYVKTQISPSQFEAHAGMAARRQPYRHIPISSGLSLHDIAM
                    SLAMGHVITTGDSDDMCSICGDGGDLLLCAGCPQAFHTACLKFQSMPEGTWYCSSCND
                    GPISSKKATTTDPSGNARPIVIRLSRVVKAPESDIGGCVFCRSHDPSIGKFDDRTVIL
                    CDQCEKEYHVGCLRENGFCDLKEIPQEKWFCCSNCSRIHTAVQNSVVSCGPQTLPTPLL
                    DMICRKDREKGIFTDIGDTVEWRILSGKSRYPEHLPLLSRAAVIFRECFDPIVAKSGR
                    DLIPVMVYGRNISGQEFGGMYCLVLIVNSLVVSAALLRIFGQEVAELPIVATSREYQG
                    RGYFQGLYACVENLLSSLNVENLVLPAAEEAESIWTKKFGFTKMSDQQLQETQKEVQL
                    TIFEGTSMLEKKVPKATTGLSESTTLI"
     repeat_region  6010. .6135
                    /rpt_family="(GGAA)n"
     misc_feature   6023. .6202
                    /note="exon predicted by xgrail, quality
                    marginal_shadowexon"
     repeat_region  6179. .6236
                    /rpt_family="POLY_A"
     mRNA           join(<6647. .6702,7161. .7304,7322. .>7499)
                    /gene="F13M22.3"
     gene           <6647. .>7499
                    /note="predicted by genscan"
                    /gene="F13M22.3"
     CDS            join(6647. .6702,7161. .7304,7322. .7499)
                    /gene="F13M22.3"
                    /note="hypothetical protein"
                    /codon_start=1
                    /db_xref="PID:g3236236"
                    /translation="MTTWKKVRSENKSFQFLFKFAWRIPHFVYGATWTLFLTLTVAII
                    SLAPFAFVSAIFPSSSEVFSRRHIGSAGSSLGGLVFTGESFPEVEDGPRCAAGFCSN
                    CCGSFRRRREDYGPLGGRRSPLI"
     mRNA           complement(join(<7761. .7931,8019. .8065,8154. .8238,
                    8323. .8401,8494. .8647,8824. .8903,9001. .9232,9769. .>9886))
                    /gene="F13M22.4"
     gene           complement(<7761. .>9886)
                    /gene="F13M22.4"
     CDS            complement(join(7761. .7931,8019. .8065,8154. .8238,
                    8323. .8401,8494. .8647,8824. .8903,9001. .9232,9769. .9886))
                    /gene="F13M22.4"
                    /codon_start=1
                    /product="putative ribotol dehydrogenase"
                    /db_xref="PID:g3236237"
                    /translation="MGIYGVMTGKKGKSGPGSASTAEDVTQAIDASHLTAIITGGTSG
                    IGLEAARVLAMRGAHVIIAARNPKAANESKEMILQMNPNARVDYLQIDVSSIKSVRSF
                    VDQFLALNVPLNILINNAGVMFCPFKLTEDGIESQFATNHIGHFLLTNLLLDKMKSTA
                    RESGVQGRIVNLSSIAHTTTYSEGIKFQGINDPACYSERRAYGQSKLSNLLHSNALSR
                    RLQEEGVNITINSVHPGLVTTNLFRYSGFSMKVFRAMTLFPWKNIPQGAATTCYVALH
                    PDLEGVTGKYFGDCNIVAPSKFATNNSLADKLWDFSVFLIDSISK"
     repeat_region  8658. .8684
                    /rpt_family="AT_rich"
     repeat_region  complement(9842. .9915)
                    /rpt_family="(GA)n"
     misc_feature   10503. .10670
```



```
                    /note="exon predicted by xgrail, quality
                    marginal_shadowexon"
     repeat_region  10725. .10746
                    /rpt_family="AT_rich"
     mRNA           complement(join(11416. .11526,11608. .12116,12551. .12628,
                    12737. .13604))
                    /gene="F13M22.5"
     gene           complement(11416. .13604)
                    /gene="F13M22.5"
     CDS            complement(join(11610. .12116,12551. .12628,12737. .13522))
                    /gene="F13M22.5"
                    /codon_start=1
                    /product="putative ARF1 GTPase activating protein"
                    /db_xref="PID:g3236238"
                    /translation="MAAARRLRTLQSQPENKVCVDCSQKNPQWASISYGIFMCLECSG
                    KHRGLGVHISFVRSVTMDSWSEIQIKKMDAGGNERLNNFLAQYGISKETDIISKYNSN
                    AASVYRDRIQALAEGRQWRDPPIVKESVGGGLMNKKPPLSQGGGRDSGNGGWDNWDND
                    DSFRSTDMRRNQSAGDFRSSGGRGAPAKSKSSEDIYSRSQLEASAAANEESFFAKRMAE
                    NESKPEGLPPSQGGKYVGPCSSPCPAPRSNQQSGGGDVFSVMSEGPGRLSLVAASAAN
                    VVQPCTMEFTSKVKEGGLDQTVSETVNVVASKTTEIGQRTWGIMKGVMAIASQKVEEF
                    TKEEASTWNQQNKTEGNGYYQNSGIGNKTANSSPGSQSSSSGHNNSYRNSNSWDDWG
                    EENNSKKEAAPKVSTSNDDDDGGWAGWDDNDAKDDDFYYQPASDKKSVGHNGKSDTAW
                    TGGGFL"
     repeat_region  11697. .11757
                    /rpt_family="(CAT)n"
     misc_feature   13716. .13952
                    /note="exon predicted by xgrail, quality
                    excellent_shadowexon"
     misc_feature   complement(14172. .14309)
                    /note="exon predicted by xgrail, quality marginal"
     mRNA           join(14707. .14911,15566. .15765,15858. .15964,16057. .16189,
                    16311. .16602,16755. .17294)
                    /gene="F13M22.6"
     gene           14707. .17294
                    /note="identical to GB:U40269"
                    /gene="F13M22.6"
     CDS            join(15634. .15765,15858. .15964,16057. .16189,16311. .16602,
                    16755. .17182)
                    /gene="F13M22.6"
                    /codon_start=1
                    /product="origin recognition complex protein"
                    /db_xref="PID:g3236239"
                    /translation="MEDIENIEEDEYGFSRNYFLAKELGGASKRSAHKLSDIHIVDEQ
                    ELRETASTIEMKHSKEISELMSDYKTMYSKWVFELRCGFGLLMYGFGSKKALVEDFAS
...
Note: remainder of annotations omitted.

Query Match          14.3%;  Score 17;  DB 24;  Length 93845;
Best Local Similarity 100.0%;  Pred. No. 3.49e+00;
Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```

```
Db  30882 ACTCTTGAGAAATGAAT 30898
          |||||||||||||||||
Qy      8 ACTCTTGAGAAATGAAT 24
```

```
RESULT   23
LOCUS       AF007262   104793 bp    DNA             PRI       20-AUG-1997
DEFINITION  Human Xq28 cosmid 3C3 and BAC 92B2 containing fragile X mental
            retardation protein (FMR2) gene, partial cds, complete sequence.
ACCESSION   AF007262 U90127
NID         g2197194
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 104793)
  AUTHORS   Muzny,D., Ansari-Lari,M.A., Timms,K.M., Yu,W., Ding,Y., Dugan,S.,
            Haywood,M., Jain,A., Karpathy,S., Leal,B., Liu,W., Logan,O., Lu,J.,
            Nguyen,V., Perez,L., Rowland,K., Savage,L., Shen,H., Shen,Y.,
            Worley,K., Chen,E., Forcum,J., Arenson,A.D., Chiu,M.W.,
            Gorrell,J.H., Brundage,E., Di,W., Chinault,C., Nelson,D. and
            Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1997) Molecular and Human Genetics, Baylor
            College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
  REMARK    U90127 updated by submitter
COMMENT     On Jun 17, 1997 this sequence version replaced gi:1890297.
            Sequencing is completed to a minimum standard of double strand
            coverage with a minimum of 2 clones and 2 reads with no ambiguities
            or 2 chemistries with a minimum of 2 clones and 3 reads with no
            ambiguities. If the sequence quality does not meet this standard,
            it will be indicated in the annotation.
            The contig containing BAC(GSHB92-B2) was constructed by FMR2 cDNA
            probe hybridization and coding analysis of overlapping clones.
            There are at least 2 overlapping clones for this region.
            The repeat regions shown were identified using RepeatMasker by
            Adrian Smit.
            Sequence similarities were identified using Powerblast by Jinghui
            Zhang.
            Exon/Intron boundaries of identified genes were chosen if there
            were canonical splice junctions that maintained sequence continuity
            across the splice junctions.
FEATURES            Location/Qualifiers
     source         1. .104793
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="X"
                    /map="Xq28"
                    /clone="3C3; 92B2"
```



```
                    /clone_lib="LLNL X chromosome library (clone 3C3); Genome
                    Systems human BAC library (clone 92B2)"
     repeat_region  341. .394
                    /rpt_family="(GA)n"
     repeat_region  complement(727. .822)
                    /rpt_family="MIR"
     repeat_region  1080. .1118
                    /rpt_family="AT_rich"
     mRNA           join(<3843. .3883,53214. .53373,55237. .56247,57971. .58092,
                    62345. .62561,66418. .66707,67257. .67320,73094. .73230,
                    77551. .77622,77980. .78073,80356. .80408,87007. .87197,
                    90851. .94064)
                    /gene="FMR2"
                    /product="fragile X mental retardation protein"
     exon           3843. .3883
                    /gene="FMR2"
     gene           3843. .94064
                    /gene="FMR2"
     repeat_region  5869. .6022
                    /rpt_family="pTR5"
     repeat_region  6227. .6360
                    /rpt_family="LTR12"
     repeat_region  6610. .6691
                    /rpt_family="pTR5"
     repeat_region  6627. .6691
                    /rpt_family="pTR5"
     repeat_region  complement(7025. .7644)
                    /rpt_family="MER11B"
     repeat_region  complement(7258. .7967)
                    /rpt_family="MER11A"
     repeat_region  complement(7964. .8114)
                    /rpt_family="MER11A"
     repeat_region  9675. .9710
                    /rpt_family="(GA)n"
     repeat_region  complement(9933. .10061)
                    /rpt_family="MIR2"
     repeat_region  complement(10675. .10769)
                    /rpt_family="(TA)n"
     repeat_region  11341. .11787
                    /rpt_family="L1HS"
     repeat_region  complement(11793. .12002)
                    /rpt_family="L1PA7"
     repeat_region  11995. .12247
                    /rpt_family="L1PA13"
     repeat_region  complement(13555. .13608)
                    /rpt_family="MIR2"
     repeat_region  complement(15829. .16130)
                    /rpt_family="AluY"
     repeat_region  complement(16344. .16424)
                    /rpt_family="(GA)n"
     repeat_region  16438. .16561
```



```
                    /rpt_family="MIR2"
     repeat_region  17498. .17524
                    /rpt_family="(CAAA)n"
     repeat_region  19146. .19243
                    /rpt_family="AT_rich"
     repeat_region  complement(20209. .20277)
                    /rpt_family="(CA)n"
     repeat_region  complement(20874. .21204)
                    /rpt_family="L1MB8"
     repeat_region  25236. .25417
                    /rpt_family="MIR"
     repeat_region  25245. .25388
                    /rpt_family="MIR"
     repeat_region  25543. .25631
                    /rpt_family="MIR"
     repeat_region  30595. .30884
                    /rpt_family="Alu"
     repeat_region  complement(32438. .32585)
                    /rpt_family="MIR2"
     STS            complement(33659. .33780)
                    /gene="FMR2"
                    /standard_name="DXS295"
     repeat_region  45158. .45280
                    /rpt_family="MER5"
     repeat_region  complement(45369. .45498)
                    /rpt_family="MIR"
     repeat_region  complement(46068. .46624)
                    /rpt_family="MLT1d"
     exon           53214. .53373
                    /gene="FMR2"
     repeat_region  complement(54108. .54265)
                    /rpt_family="MIR"
     CDS            join(<55236. .56247,57971. .58092,62345. .62561,66418. .66707,
                    67257. .67320,73094. .73230,77551. .77622,77980. .78073,
                    80356. .80408,87007. .87197,90851. .90972)
                    /gene="FMR2"
                    /codon_start=2
                    /product="fragile X mental retardation protein"
                    /db_xref="PID:g2388546"
                    /translation="PEPPSTNKWQLDKWLNKVTSQNKSFICGQNETPMETISLPPPII
                    QPMEVQMKVKTNASQVPAEPKERPLLSLIREKARPRPTQKIPETKALKHKLSTTSETV
                    SQRTIGKKQPKKVEKNTSTDEFTWPKPNITSSTPKEKESVELHDPPRGRNKATAHKPA
                    PRKEPRPNIPLAPEKKKYRGPGKIVPKSREFIETDSSTSDSNTDQEETLQIKVLPPCI
                    ISGGNTAKSKEICGASLTLSTLMSSSGSNWNLSISNEEPTPSPIPVMQTEILSPLRDH
                    ENLKNLWVKIDLDLLSRVPGHSSLHAAPAKPDHKETATKPKRQTAVTAVEKPAPKGKR
                    KHKPIEVAEKIPEKKQRLEEATTICLLPPCISPAPPHKPPNTRENNSSRRANRREEEK
                    LPPPPLSPLPEDPPRRRNVSGNNGPPGQDKNIAMTGQITSTKPERTEGKFCATFKGIS
                    EGDTPKKASSATITVTNTAIATATVTATAIVTTTVTATATATATTTTTTTISTITST
                    ITTGLMDSSHLEMTSWAALPLLSSSSTNVRRPKLTFDDSVHNADTYMQEAKKLKHKAD
                    ALFEKPGKAVNYADAALSFTECGNAMERDPLEAKSPYTMYSETVELLRYAMRLKNFAS
                    PLASDGDKKLAVLCYRCLSLLLYLRMFKLKKDHAMKYSRSLMETFKQNASKVAQIPSPW
```



```
                    VSNGKNTPSPVSLNNVSPINAMGNCNNGPVTIPQRIHHMAASHVNITSNVLRGYEHWD
                    MADKLTRENKEFFGDLDTLMGPLTQHSSMTNLVRYVRQGLCWLRIDAHLL"
     exon           55237. .56247
                    /gene="FMR2"
     exon           57971. .58092
                    /gene="FMR2"
     misc_feature   61640
                    /gene="FMR2"
                    /note="Polymorphism"
     exon           62345. .62561
                    /gene="FMR2"
     misc_feature   63918
                    /gene="FMR2"
                    /note="Polymorphism"
     repeat_region  64063. .64350
                    /rpt_family="L1MA2"
     repeat_region  64066. .64436
                    /rpt_family="L1MB3"
     exon           66418. .66707
                    /gene="FMR2"
     misc_feature   complement(67001. .67370)
                    /gene="FMR2"
                    /note="similar to EST with GenBank Accession number
                    H18241"
     exon           67257. .67320
                    /gene="FMR2"
     repeat_region  67950. .68110
                    /rpt_family="MER5B"
     repeat_region  69032. .69366
                    /rpt_family="MER2"
     repeat_region  complement(70755. .70780)
                    /rpt_family="AT_rich"
     repeat_region  complement(71360. .71404)
                    /rpt_family="POLY_A"
     repeat_region  complement(71405. .72297)
                    /rpt_family="L1PA2"
     repeat_region  complement(72148. .72788)
                    /rpt_family="L1"
     repeat_region  complement(72821. .73034)
                    /rpt_family="MER20"
     exon           73094. .73230
                    /gene="FMR2"
     repeat_region  73883. .73913
                    /rpt_family="(GAA)n"
     repeat_region  complement(75058. .75091)
                    /rpt_family="AT_rich"
     repeat_region  75286. .75586
                    /rpt_family="AluYa5"
     repeat_region  complement(75801. .75989)
                    /rpt_family="L1MD2"
     repeat_region  77239. .77280
```

```
                     /rpt_family="(CA)n"
     exon            77551. .77622
...
Note: remainder of annotations omitted.

  Query Match           14.3%;  Score 17;  DB 26;  Length 104793;
  Best Local Similarity 100.0%;  Pred. No. 3.49e+00;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  52184 CATTTCTGAGAGAAAGC 52200
          |||||||||||||||||
Cp     43 CATTTCTGAGAGAAAGC 27


RESULT   24
LOCUS       HS232N11   123468 bp    DNA              HTG       22-SEP-1998
DEFINITION  Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone
            232N11; HTGS phase 1.
ACCESSION   AL031656
NID         g3647170
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 123468)
  AUTHORS   Hunt,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-SEP-1998) Wellcome Trust Genome Campus, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquires:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and the release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc. Order of segments is not known; 800 n's separate
            segments. Unfinished: dJ232N11  Contig_ID: 01145  acc=  Length:
            8642 bp Unfinished: dJ232N11  Contig_ID: 01307  acc=  Length: 6314
            bp Unfinished: dJ232N11  Contig_ID: 01874  acc=  Length: 9517 bp
            Unfinished: dJ232N11  Contig_ID: 01499  acc=  Length: 43435 bp
            Unfinished: dJ232N11  Contig_ID: 00259  acc=  Length: 1105 bp
            Unfinished: dJ232N11  Contig_ID: 00432  acc=  Length: 1226 bp
            Unfinished: dJ232N11  Contig_ID: 01743  acc=  Length: 1030 bp
            Unfinished: dJ232N11  Contig_ID: 01682  acc=  Length: 1046 bp
            Unfinished: dJ232N11  Contig_ID: 01841  acc=  Length: 2877 bp
            Unfinished: dJ232N11  Contig_ID: 02247  acc=  Length: 3510 bp
            Unfinished: dJ232N11  Contig_ID: 01084  acc=  Length: 36766 bp.
            ***                                                        ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                        ***
```



```
            * This sequence is unfinished. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
FEATURES             Location/Qualifiers
     source          1. .123468
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="20"
                     /clone="232N11"
BASE COUNT    34195 a   24953 c   23954 g   32365 t    8001 others
ORIGIN

  Query Match           14.3%;  Score 17;  DB 17;  Length 123468;
  Best Local Similarity 100.0%;  Pred. No. 3.49e+00;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  92959 CATTTATTCCTTATACA 92975
          |||||||||||||||||
Cp     90 CATTTATTCCTTATACA 74


RESULT   25
LOCUS       AC005051   132698 bp    DNA              HTG       12-JUN-1998
DEFINITION  *** SEQUENCING IN PROGRESS *** Homo sapiens clone RG037P03; HTGS
            phase 1, 12 unordered pieces.
ACCESSION   AC005051
NID         g3212940
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 132698)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 132698)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-1998) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     ***                                                        ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                        ***
            * This sequence is unfinished. It consists of 12 contigs for
            * which the order is not known; their order in this record is
            * arbitrary. In some cases, the exact lengths of the gaps
            * between the contigs are also unknown; these gaps are  presented
            * as runs of N as a convenience only. When sequencing is complete,
            * the sequence data presented in this record will be replaced
```



```
            *by a single finished sequence with the same accession number.
            *       1     1483: contig of 1483 bp in length
            *    1484     1500: gap of unknown length
            *    1501     3489: contig of 1989 bp in length
            *    3490     3506: gap of unknown length
            *    3507     5674: contig of 2168 bp in length
            *    5675     5691: gap of unknown length
            *    5692    10331: contig of 4640 bp in length
            *   10332    10348: gap of unknown length
            *   10349    15810: contig of 5462 bp in length
            *   15811    15827: gap of unknown length
            *   15828    20876: contig of 5049 bp in length
            *   20877    20893: gap of unknown length
            *   20894    26544: contig of 5651 bp in length
            *   26545    26561: gap of unknown length
            *   26562    35813: contig of 9252 bp in length
            *   35814    35830: gap of unknown length
            *   35831    45222: contig of 9392 bp in length
            *   45223    45239: gap of unknown length
            *   45240    60183: contig of 14944 bp in length
            *   60184    60200: gap of unknown length
            *   60201    77865: contig of 17665 bp in length
            *   77866    77882: gap of unknown length
            *   77883   132698: contig of 54816 bp in length.
FEATURES             Location/Qualifiers
     source          1. .132698
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RG037P03"
BASE COUNT    34236 a   31999 c   33721 g   32555 t     187 others
ORIGIN

  Query Match           14.3%;  Score 17;  DB 17;  Length 132698;
  Best Local Similarity 100.0%;  Pred. No. 3.49e+00;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  11846 CAGCATTTCTGAGAGAA 11862
          |||||||||||||||||
Cp     46 CAGCATTTCTGAGAGAA 30


RESULT   26
LOCUS       AC004950   133798 bp    DNA              HTG       12-JUN-1998
DEFINITION  *** SEQUENCING IN PROGRESS *** Homo sapiens clone DJ1019E05; HTGS
            phase 1, 10 unordered pieces.
ACCESSION   AC004950
NID         g3213047
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
```



```
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 133798)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 133798)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-1998) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     ***                                                        ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                        ***
            * This sequence is unfinished. It consists of 10 contigs for
            * which the order is not known; their order in this record is
            * arbitrary. In some cases, the exact lengths of the gaps
            * between the contigs are also unknown; these gaps are  presented
            * as runs of N as a convenience only. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
            *       1     1550: contig of 1550 bp in length
            *    1551     1568: gap of unknown length
            *    1569     3313: contig of 1745 bp in length
            *    3314     3331: gap of unknown length
            *    3332     6148: contig of 2817 bp in length
            *    6149     6166: gap of unknown length
            *    6167    10868: contig of 4702 bp in length
            *   10869    10886: gap of unknown length
            *   10887    18683: contig of 7797 bp in length
            *   18684    18701: gap of unknown length
            *   18702    30669: contig of 11968 bp in length
            *   30670    30687: gap of unknown length
            *   30688    43001: contig of 12314 bp in length
            *   43002    43019: gap of unknown length
            *   43020    59054: contig of 16035 bp in length
            *   59055    59072: gap of unknown length
            *   59073    91553: contig of 32481 bp in length
            *   91554    91571: gap of unknown length
            *   91572   133798: contig of 42227 bp in length.
FEATURES             Location/Qualifiers
     source          1. .133798
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="DJ1019E05"
BASE COUNT    34831 a   31738 c   30124 g   36943 t     162 others
ORIGIN

  Query Match           14.3%;  Score 17;  DB 17;  Length 133798;
  Best Local Similarity 100.0%;  Pred. No. 3.49e+00;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```

```
Db  64692 CAGCATTTCTGAGAGAA 64708
          |||||||||||||||||
Cp     46 CAGCATTTCTGAGAGAA 30


RESULT   27
LOCUS       AC004993   161453 bp    DNA             PRI       19-SEP-1998
DEFINITION  Homo sapiens PAC clone DJ1194E14 from 7p21, complete sequence.
ACCESSION   AC004993
NID         g3638948
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 161453)
  AUTHORS   Harmon,G., Bauer,C., Drone,K. and Clarke,K.
  TITLE     The sequence of Homo sapiens PAC clone DJ1194E14
  JOURNAL   Unpublished (1998)
REFERENCE   2  (bases 1 to 161453)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-1998) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   3  (bases 1 to 161453)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-SEP-1998) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Sep 19, 1998 this sequence version replaced gi:3213004.
            SUBMITTED BY:  WUGSC
            Genome Sequencing Center
            Department of Genetics
            Washington University
            St. Louis MO 63108, USA
            http://genome.wustl.edu/gsc
            mailto:sapiens@watson.wustl.edu

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded or sequenced with an alternate
            chemistry; an attempt was made to resolve all sequencing problems,
            such as compressions and repeats; all regions were covered by
            sequence from more than one subclone; and the assembly was
            confirmed by restriction digest.
```



```
            MAPPING INFORMATION:
            The sequence of this clone was established as part of a mapping and
            sequencing collaboration between the NHGRI Chromosome 7 Mapping
            Project (Eric D. Green, Director), John D. McPherson in the
            Department of Genetics (Washington University), and the Washington
            University Genome Sequencing Center.  For additional information
            about the map position of this sequence, see
            http://www.nhgri.nih.gov/DIR/GTB/CHR7 , send
            mailto:egreen@nhgri.nih.gov , or see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            This clone was derived from human PAC library RPCI-1, prepared by
            Pieter de Jong and coworkers at Roswell Park Cancer Institute,
            using the method described by Ioannou et al., Nature Genetics
            6:84-9 (1994).  The library is from one male donor.  For further
            details, see http://bacpac.med.buffalo.edu/
            The clone is available from Genome Systems, Inc.
            (http://www.genomesystems.com).
            VECTOR:  pCYPAC2
            NEIGHBORING SEQUENCE INFORMATION:
            Actual start of this clone is at base position 1 of DJ1194E14;
            actual end is at 161453 of DJ1194E14.
FEATURES             Location/Qualifiers
     source          1. .161453
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7p21"
                     /clone_lib="RPCI-1"
                     /clone="DJ1194E14"
     repeat_region   3952. .4072
                     /rpt_family="Alu"
     repeat_region   4982. .5188
                     /rpt_family="L2"
     repeat_region   5323. .5427
                     /rpt_family="L2"
     repeat_region   5500. .5931
                     /rpt_family="L2"
     repeat_region   6540. .6688
                     /rpt_family="L1"
     repeat_region   6689. .6989
                     /rpt_family="Alu"
     repeat_region   6990. .7525
                     /rpt_family="L1"
     repeat_region   8126. .8250
                     /rpt_family="MIR"
     repeat_region   8604. .9696
                     /rpt_family="L1"
     repeat_region   9688. .10614
                     /rpt_family="L1"
```



```
     repeat_region   13411. .13689
                     /rpt_family="Alu"
     repeat_region   15755. .16052
                     /rpt_family="Alu"
     repeat_region   19816. .20030
                     /rpt_family="Retroviral"
     repeat_region   20163. .20258
                     /rpt_family="L1"
     repeat_region   20277. .20360
                     /rpt_family="MER1_type"
     repeat_region   20361. .20668
                     /rpt_family="Alu"
     repeat_region   20669. .20814
                     /rpt_family="MER1_type"
     repeat_region   20957. .21798
                     /rpt_family="Other"
     repeat_region   22952. .23264
                     /rpt_family="Other"
     repeat_region   23265. .24349
                     /rpt_family="Retroviral"
     repeat_region   24350. .24538
                     /rpt_family="Other"
     repeat_region   24687. .24972
                     /rpt_family="Alu"
     repeat_region   25135. .25145
                     /rpt_family="L1"
     repeat_region   25146. .25412
                     /rpt_family="Alu"
     repeat_region   25413. .26549
                     /rpt_family="L1"
     repeat_region   26544. .26937
                     /rpt_family="L1"
     repeat_region   28051. .28164
                     /rpt_family="MIR"
     repeat_region   28442. .28666
                     /rpt_family="MIR"
     repeat_region   28980. .29222
                     /rpt_family="MIR"
     repeat_region   29249. .29832
                     /rpt_family="L1"
     repeat_region   29833. .29912
                     /rpt_family="MaLR"
     repeat_region   29913. .31040
                     /rpt_family="L1"
     repeat_region   31041. .31415
                     /rpt_family="MaLR"
     repeat_region   31416. .31909
                     /rpt_family="L1"
     repeat_region   32997. .33139
                     /rpt_family="MaLR"
     repeat_region   33140. .33450
```



```
                     /rpt_family="Alu"
     repeat_region   33451. .33806
                     /rpt_family="MaLR"
     repeat_region   35270. .35369
                     /rpt_family="MaLR"
     repeat_region   35370. .35674
                     /rpt_family="Alu"
     repeat_region   35740. .36111
                     /rpt_family="Retroviral"
     repeat_region   36318. .36448
                     /rpt_family="MER1_type?"
     repeat_region   36450. .36744
                     /rpt_family="L1"
     repeat_region   39875. .40477
                     /rpt_family="L1"
     repeat_region   42096. .42404
                     /rpt_family="Alu"
     repeat_region   45284. .46035
                     /rpt_family="L1"
     repeat_region   46139. .46230
                     /rpt_family="MER1_type"
     repeat_region   48134. .48442
                     /rpt_family="MaLR"
     repeat_region   49956. .50444
                     /rpt_family="L2"
     repeat_region   52782. .53143
                     /rpt_family="MaLR"
     STS             complement(52973. .53246)
                     /db_xref="GI:1113442"
     repeat_region   53207. .53367
                     /rpt_family="L1"
     repeat_region   53476. .53628
                     /rpt_family="L1"
     repeat_region   53629. .53939
                     /rpt_family="Alu"
     repeat_region   53940. .54164
                     /rpt_family="L1"
     repeat_region   54120. .54439
                     /rpt_family="L1"
     repeat_region   54492. .54691
                     /rpt_family="L1"
     repeat_region   54829. .55117
                     /rpt_family="Alu"
     repeat_region   55179. .56017
                     /rpt_family="MER2_type"
     repeat_region   57437. .58088
                     /rpt_family="L1"
     repeat_region   58096. .58287
                     /rpt_family="L1"
     repeat_region   58807. .59082
                     /rpt_family="L2"
```

```
     repeat_region   59216. .59328
                     /rpt_family="MIR"
     repeat_region   59328. .59466
...
Note: remainder of annotations omitted.

  Query Match           14.3%;  Score 17;  DB 27;  Length 161453;
  Best Local Similarity  94.4%;  Pred. No. 3.49e+00;
  Matches    17;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db  90104 TTTGCTTTTTATGGAGCT 90121
          |||||||||||| ||||||
Cp    120 TTTGCTTTTTANGGAGCT 103


RESULT   28
LOCUS       AC005059   170128 bp    DNA             HTG       12-JUN-1998
DEFINITION  *** SEQUENCING IN PROGRESS *** Homo sapiens clone RG074A24; HTGS
            phase 1, 25 unordered pieces.
ACCESSION   AC005059
NID         g3212932
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 170128)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 170128)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-1998) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     ***                                                        ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                        ***
            * This sequence is unfinished. It consists of 25 contigs for
            * which the order is not known; their order in this record is
            * arbitrary. In some cases, the exact lengths of the gaps
            * between the contigs are also unknown; these gaps are  presented
            * as runs of N as a convenience only. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
            *       1     1662: contig of 1662 bp in length
            *    1663     1679: gap of unknown length
            *    1680     3248: contig of 1569 bp in length
            *    3249     3265: gap of unknown length
            *    3266     4902: contig of 1637 bp in length
```



```
            *    4903     4919: gap of unknown length
            *    4920     6551: contig of 1632 bp in length
            *    6552     6568: gap of unknown length
            *    6569     8090: contig of 1522 bp in length
            *    8091     8107: gap of unknown length
            *    8108     9832: contig of 1725 bp in length
            *    9833     9849: gap of unknown length
            *    9850    11514: contig of 1665 bp in length
            *   11515    11531: gap of unknown length
            *   11532    13880: contig of 2349 bp in length
            *   13881    13897: gap of unknown length
            *   13898    16342: contig of 2445 bp in length
            *   16343    16359: gap of unknown length
            *   16360    20688: contig of 4329 bp in length
            *   20689    20705: gap of unknown length
            *   20706    24570: contig of 3865 bp in length
            *   24571    24587: gap of unknown length
            *   24588    28578: contig of 3991 bp in length
            *   28579    28595: gap of unknown length
            *   28596    35314: contig of 6719 bp in length
            *   35315    35331: gap of unknown length
            *   35332    39541: contig of 4210 bp in length
            *   39542    39558: gap of unknown length
            *   39559    46401: contig of 6843 bp in length
            *   46402    46418: gap of unknown length
            *   46419    53312: contig of 6894 bp in length
            *   53313    53329: gap of unknown length
            *   53330    60264: contig of 6935 bp in length
            *   60265    60281: gap of unknown length
            *   60282    66853: contig of 6572 bp in length
            *   66854    66870: gap of unknown length
            *   66871    75039: contig of 8169 bp in length
            *   75040    75056: gap of unknown length
            *   75057    84782: contig of 9726 bp in length
            *   84783    84799: gap of unknown length
            *   84800    97374: contig of 12575 bp in length
            *   97375    97391: gap of unknown length
            *   97392   109799: contig of 12408 bp in length
            *  109800   109816: gap of unknown length
            *  109817   124892: contig of 15076 bp in length
            *  124893   124909: gap of unknown length
            *  124910   143882: contig of 18973 bp in length
            *  143883   143899: gap of unknown length
            *  143900   170128: contig of 26229 bp in length.
FEATURES             Location/Qualifiers
     source          1. .170128
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RG074A24"
BASE COUNT    49024 a  34513 c  35801 g  50382 t    408 others
ORIGIN
```



```
  Query Match           14.3%;  Score 17;  DB 17;  Length 170128;
  Best Local Similarity 100.0%;  Pred. No. 3.49e+00;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  98749 TCAGTTTTGTCTAGTGT 98765
          |||||||||||||||||
Cp     70 TCAGTTTTGTCTAGTGT 54


RESULT   29
LOCUS       CEY51B9    191908 bp    DNA             HTG       22-SEP-1998
DEFINITION  Caenorhabditis elegans DNA *** SEQUENCING IN PROGRESS *** from
            clone Y51B9; HTGS phase 1.
ACCESSION   Z98947
NID         g3378102
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 191908)
  AUTHORS   Barlow,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-SEP-1998) Nematode Sequencing Project, Sanger Centre,
            Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics,
            Washington University, St. Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     On Aug 4, 1998 this sequence version replaced gi:2467105.
            Order of segments is not known; 800 n's separate segments.
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and the release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc.
            ***                                                        ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                        ***
            * This sequence is unfinished. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
FEATURES             Location/Qualifiers
     source          1. .191908
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /chromosome="II"
                     /clone="Y51B9"
BASE COUNT    61139 a  33459 c  33257 g  61653 t   2400 others
ORIGIN
```



```
  Query Match           14.3%;  Score 17;  DB 17;  Length 191908;
  Best Local Similarity 100.0%;  Pred. No. 3.49e+00;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  86446 TCTTGAGAAATGAATAA 86462
          |||||||||||||||||
Qy     10 TCTTGAGAAATGAATAA 26


RESULT   30
LOCUS       AC005055   224064 bp    DNA             HTG       12-JUN-1998
DEFINITION  *** SEQUENCING IN PROGRESS *** Homo sapiens clone RG050N15; HTGS
            phase 1, 26 unordered pieces.
ACCESSION   AC005055
NID         g3212936
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 224064)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 224064)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-1998) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     ***                                                        ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                        ***
            * This sequence is unfinished. It consists of 26 contigs for
            * which the order is not known; their order in this record is
            * arbitrary. In some cases, the exact lengths of the gaps
            * between the contigs are also unknown; these gaps are  presented
            * as runs of N as a convenience only. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
            *       1     1567: contig of 1567 bp in length
            *    1568     1584: gap of unknown length
            *    1585     3757: contig of 2173 bp in length
            *    3758     3774: gap of unknown length
            *    3775     5645: contig of 1871 bp in length
            *    5646     5662: gap of unknown length
            *    5663     8218: contig of 2556 bp in length
            *    8219     8235: gap of unknown length
            *    8236    10568: contig of 2333 bp in length
            *   10569    10585: gap of unknown length
            *   10586    12673: contig of 2088 bp in length
```

```
              *    12674    12690: gap of unknown length
              *    12691    14458: contig of 1768 bp in length
              *    14459    14475: gap of unknown length
              *    14476    17399: contig of 2924 bp in length
              *    17400    17416: gap of unknown length
              *    17417    20845: contig of 3429 bp in length
              *    20846    20862: gap of unknown length
              *    20863    24474: contig of 3612 bp in length
              *    24475    24491: gap of unknown length
              *    24492    30780: contig of 6289 bp in length
              *    30781    30797: gap of unknown length
              *    30798    36366: contig of 5569 bp in length
              *    36367    36383: gap of unknown length
              *    36384    41721: contig of 5338 bp in length
              *    41722    41738: gap of unknown length
              *    41739    49140: contig of 7402 bp in length
              *    49141    49157: gap of unknown length
              *    49158    54878: contig of 5721 bp in length
              *    54879    54895: gap of unknown length
              *    54896    60864: contig of 5969 bp in length
              *    60865    60881: gap of unknown length
              *    60882    72111: contig of 11230 bp in length
              *    72112    72128: gap of unknown length
              *    72129    81775: contig of 9647 bp in length
              *    81776    81792: gap of unknown length
              *    81793    92380: contig of 10588 bp in length
              *    92381    92397: gap of unknown length
              *    92398   106304: contig of 13907 bp in length
              *   106305   106321: gap of unknown length
              *   106322   119608: contig of 13287 bp in length
              *   119609   119625: gap of unknown length
              *   119626   136119: contig of 16494 bp in length
              *   136120   136136: gap of unknown length
              *   136137   157328: contig of 21192 bp in length
              *   157329   157345: gap of unknown length
              *   157346   176995: contig of 19650 bp in length
              *   176996   177012: gap of unknown length
              *   177013   198965: contig of 21953 bp in length
              *   198966   198982: gap of unknown length
              *   198983   224064: contig of 25082 bp in length.
FEATURES             Location/Qualifiers
     source          1. .224064
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RG050N15"
BASE COUNT    65068 a  45327 c  47180 g  66064 t    425 others
ORIGIN

  Query Match           14.3%;  Score 17;  DB 17;  Length 224064;
  Best Local Similarity 100.0%;  Pred. No. 3.49e+00;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```



```
Db 212516 TCAGTTTTGTCTAGTGT 212532
          |||||||||||||||||
Cp     70 TCAGTTTTGTCTAGTGT 54


RESULT   31
LOCUS       HS407F11   329751 bp    DNA             HTG       23-SEP-1998
DEFINITION  Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone
            407F11; HTGS phase 1.
ACCESSION   AL022329
NID         g3646085
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 329751)
  AUTHORS   Palmer,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-SEP-1998) Wellcome Trust Genome Campus, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquires:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 24, 1998 this sequence version replaced gi:3402800.
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and the release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc. Order of segments is not known; 800 n's separate
            segments. Unfinished: bK407F11  Contig_ID: 03127  acc=AL022329
            Length: 3718 bp Unfinished: bK407F11  Contig_ID: 02420
            acc=AL022329  Length: 2438 bp Unfinished: bK407F11  Contig_ID:
            05364  acc=AL022329  Length: 1471 bp Unfinished: bK407F11
            Contig_ID: 03968  acc=AL022329  Length: 2710 bp Unfinished:
            bK407F11  Contig_ID: 03989  acc=AL022329  Length: 1224 bp
            Unfinished: bK407F11  Contig_ID: 00319  acc=AL022329  Length: 1128
            bp Unfinished: bK407F11  Contig_ID: 02732  acc=AL022329  Length:
            1211 bp Unfinished: bK407F11  Contig_ID: 06124  acc=AL022329
            Length: 1275 bp Unfinished: bK407F11  Contig_ID: 01844
            acc=AL022329  Length: 2079 bp Unfinished: bK407F11  Contig_ID:
            01852  acc=AL022329  Length: 1128 bp Unfinished: bK407F11
            Contig_ID: 05437  acc=AL022329  Length: 1015 bp Unfinished:
            bK407F11  Contig_ID: 05460  acc=AL022329  Length: 1034 bp
            Unfinished: bK407F11  Contig_ID: 07076  acc=AL022329  Length: 36756
            bp Unfinished: bK407F11  Contig_ID: 01181  acc=AL022329  Length:
            1048 bp Unfinished: bK407F11  Contig_ID: 06202  acc=AL022329
            Length: 1910 bp Unfinished: bK407F11  Contig_ID: 02663
            acc=AL022329  Length: 1477 bp Unfinished: bK407F11  Contig_ID:
            06297  acc=AL022329  Length: 3362 bp Unfinished: bK407F11
            Contig_ID: 01268  acc=AL022329  Length: 1009 bp Unfinished:
```



```
            bK407F11  Contig_ID: 06968  acc=AL022329  Length: 168855 bp
            Unfinished: bK407F11  Contig_ID: 03486  acc=AL022329  Length: 1047
            bp Unfinished: bK407F11  Contig_ID: 03496  acc=AL022329  Length:
            4889 bp Unfinished: bK407F11  Contig_ID: 02783  acc=AL022329
            Length: 3391 bp Unfinished: bK407F11  Contig_ID: 06676
            acc=AL022329  Length: 15568 bp Unfinished: bK407F11  Contig_ID:
            02030  acc=AL022329  Length: 1100 bp Unfinished: bK407F11
            Contig_ID: 02096  acc=AL022329  Length: 1021 bp Unfinished:
            bK407F11  Contig_ID: 02833  acc=AL022329  Length: 1184 bp
            Unfinished: bK407F11  Contig_ID: 05700  acc=AL022329  Length: 1146
            bp Unfinished: bK407F11  Contig_ID: 06486  acc=AL022329  Length:
            3099 bp Unfinished: bK407F11  Contig_ID: 01415  acc=AL022329
            Length: 2179 bp Unfinished: bK407F11  Contig_ID: 01419
            acc=AL022329  Length: 3721 bp Unfinished: bK407F11  Contig_ID:
            04230  acc=AL022329  Length: 1227 bp Unfinished: bK407F11
            Contig_ID: 06512  acc=AL022329  Length: 1423 bp Unfinished:
            bK407F11  Contig_ID: 02964  acc=AL022329  Length: 1451 bp
            Unfinished: bK407F11  Contig_ID: 00041  acc=AL022329  Length: 1004
            bp Unfinished: bK407F11  Contig_ID: 05141  acc=AL022329  Length:
            1474 bp Unfinished: bK407F11  Contig_ID: 05153  acc=AL022329
            Length: 2100 bp Unfinished: bK407F11  Contig_ID: 00156
            acc=AL022329  Length: 1896 bp Unfinished: bK407F11  Contig_ID:
            02326  acc=AL022329  Length: 1150 bp Unfinished: bK407F11
            Contig_ID: 03060  acc=AL022329  Length: 1704 bp Unfinished:
            bK407F11  Contig_ID: 02370  acc=AL022329  Length: 1085 bp
            Unfinished: bK407F11  Contig_ID: 02398  acc=AL022329  Length: 2450
            bp Unfinished: bK407F11  Contig_ID: 01675  acc=AL022329  Length:
            1420 bp Unfinished: bK407F11  Contig_ID: 04527  acc=AL022329
            Length: 4574 bp.
            ***                                                         ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                         ***
            * This sequence is unfinished. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
FEATURES             Location/Qualifiers
     source          1. .329751
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="22"
                     /clone="407F11"
BASE COUNT    82181 a  60928 c  64519 g  88502 t  33621 others
ORIGIN

  Query Match           14.3%;  Score 17;  DB 17;  Length 329751;
  Best Local Similarity 100.0%;  Pred. No. 3.49e+00;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db 284255 AAATGAATAAGCTTTCT 284271
          |||||||||||||||||
Qy     17 AAATGAATAAGCTTTCT 33
```



```


RESULT   32
LOCUS       G36497       448 bp    mRNA            STS       31-DEC-1997
DEFINITION  SHGC-53095 Human Homo sapiens STS cDNA, sequence tagged site.
ACCESSION   G36497
NID         g2734164
KEYWORDS    STS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 448)
  AUTHORS   Myers,R.M.
  TITLE     Human STSs (1997)
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Richard M. Myers
            Stanford Human Genome Center (SHGC)
            Stanford University School of Medicine
            Department of Genetics, M-344, Stanford, CA 94305, USA
            Tel: 4157259687
            Fax: 4157259689
            Email: myers@shgc.stanford.edu
            Primer A: TCCACATGACACATTTATAAATTCC
            Primer B: TCCAGAATGTGCTATGGCAG
            STS size: 209
            PCR Profile:
                  Initial incubation: 95 degrees C for 10 minutes

                  Denaturation:       94 degrees C for 30 seconds
                  Annealing:          60 degrees C for 30 seconds
                  Polymerisation:     72 degrees C for 23 seconds
                  PCR Cycles:         30
                  Thermal Cycler:     Perkin Elmer 9600
            Protocol:
                  Template:                   25 ng
                  Primer:                     each 1 uM
                  dNTPs:                      each 200 uM
                  AmpliTaq Gold Polymerase: 0.07 units/ul
                  Total Vol:                  5 ul

            Buffer:
                  MgCl2:          2.5 mM
                  KCl:            50 mM
                  Tris-HCl:       10 mM
                  pH:              8.3

            Prepared with primer pairs derived from AA055645 -- Unigene.
FEATURES             Location/Qualifiers
     source          1. .448
```

```
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="6"
                     /clone_lib="Human"
     STS             150. .358
     primer_bind     150. .174
     primer_bind     complement(339. .358)
BASE COUNT      163 a      84 c      67 g     130 t       4 others
ORIGIN

  Query Match           13.4%;  Score 16;  DB 30;  Length 448;
  Best Local Similarity 100.0%;  Pred. No. 1.92e+01;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    423 ACACTAGACAAAACTG 438
          ||||||||||||||||
Qy     54 ACACTAGACAAAACTG 69


RESULT   33
LOCUS       HS3MS630      623 bp    DNA                 PRI       13-OCT-1992
DEFINITION  H.sapiens minisatellite MS630 3' flanking region.
ACCESSION   X59141
NID         g23872
KEYWORDS    Alu element; hypervariable region; minisatellite.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 623)
  AUTHORS   Armour,J.A.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-1991) J.A.L. Armour, Dept of Genetics, University
            of Leicester, Adrian Bldg, University Road, Leicester LE1 7RH, UK
REFERENCE   2  (bases 1 to 623)
  AUTHORS   Armour,J.A.L., Crosier,M. and Jeffreys,A.J.
  TITLE     Undetected 'null' alleles at human minisatellite loci
  JOURNAL   Unpublished
COMMENT     See related sequences X59140-X59144.
FEATURES             Location/Qualifiers
     source          1. .623
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="charomid; genomic"
                     /clone="cMS630"
                     /chromosome="X/Y pseudoautosomal region"
                     /map="DXYS82"
     repeat_region   1. .237
                     /note="inverted orientation Alu"
     repeat_region   590. .623
                     /note="minisatellite"
```



```
BASE COUNT      152 a    166 c    121 g    184 t
ORIGIN

  Query Match           13.4%;  Score 16;  DB 25;  Length 623;
  Best Local Similarity 100.0%;  Pred. No. 1.92e+01;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    119 TAGGGACAGCATTTCT 134
          ||||||||||||||||
Cp     52 TAGGGACAGCATTTCT 37


RESULT   34
LOCUS       HSU85248      805 bp    DNA                 PRI       07-FEB-1997
DEFINITION  Human cholesteryl ester transfer protein (CETP) gene, promoter
            region and partial exon 1.
ACCESSION   U85248
NID         g1785863
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 805)
  AUTHORS   Corbex,M., Poirier,O., Fumeron,F., Evans,A., Ruidavets,J.B.,
            Arveiler,D., Luc,G., Parra,H.J., Tiret,L. and Cambien,F.
  TITLE     Complexity of the genotype-phenotype association: four potentially
            functional polymorphisms in the CETP gene and their interaction
            with alcohol consumption on plasma HDL cholestrol and coronary
            heart disease
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 805)
  AUTHORS   Poirier,O.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-DEC-1996) INSERM U 258, Hopital Broussais, 96 rue
            Didot, Paris 75014, France
FEATURES             Location/Qualifiers
     source          1. .805
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /map="16q13.2"
     allele          147
                     /note="polymorphism in Caucasian population"
                     /frequency="0.47"
                     /replace="a"
     TATA_signal     745. .750
     exon            776. .>805
                     /number=1
     prim_transcript 776. .>805
     gene            803. .>805
```



```
                     /gene="CETP"
     CDS             803. .>805
                     /gene="CETP"
                     /note="start codon"
                     /codon_start=1
                     /product="cholesteryl ester transfer protein"
                     /db_xref="PID:g1785864"
                     /translation="M"
BASE COUNT      219 a    176 c    200 g    210 t
ORIGIN

  Query Match           13.4%;  Score 16;  DB 25;  Length 805;
  Best Local Similarity 100.0%;  Pred. No. 1.92e+01;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    470 GAGAGAAAGCTTATTC 485
          ||||||||||||||||
Cp     36 GAGAGAAAGCTTATTC 21


RESULT   35
LOCUS       CHKH1PR      1264 bp    DNA                 VRT       15-MAR-1988
DEFINITION  Chicken H1 histone gene pCH11.5E right (11R H1), complete cds with
            promoters.
ACCESSION   M17020
NID         g211838
KEYWORDS    histone; histone H1.
SOURCE      Chicken DNA, clone pCH11.5E.
  ORGANISM  Gallus gallus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Archosauria; Aves; Neognathae; Galliformes;
            Phasianidae; Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 1264)
  AUTHORS   Coles,L.S., Robins,A.J., Madley,L.K. and Wells,J.R.E.
  TITLE     Characterization of the chicken histone H1 gene complement:
            generation of a complete set of vertebrate H1 protein sequences
  JOURNAL   J. Biol. Chem. 262, 9656-9663 (1987)
  MEDLINE   87250632
COMMENT     Draft entry and clean copy of sequence [1] kindly provided by
            L.S.Coles (09-SEP-1987).
            Comparison of H1 non-coding regions in chickens and other organisms
            reveals conserved sequences in both 5' and 3' ends:  bp 218-229 (H1
            specific); bp 261-272 (g-rich); bp 1054-1078, 1086-1098, 1121-1126
            (histone-specific 3' elements).
FEATURES             Location/Qualifiers
     source          1. .1264
                     /organism="Gallus gallus"
                     /db_xref="taxon:9031"
     misc_signal     278. .287
                     /note="CAAT-like sequence"
     TATA_signal     302. .310
```



```
     mRNA            334. .1077
                     /note="11R H1 mRNA"
     CDS             355. .1014
                     /note="11R H1 protein"
                     /codon_start=1
                     /db_xref="PID:g211839"
                     /translation="MAETAPAAAPAAAAPAPAAKAAAKKPKKAAGGAKARKPAGPSVTE
                     LITKAVSASKERKGLSLAALKKALAAGGYDVEKNNSRIKLGLKSLVSKGTLVQTKGTG
                     ASGSFRLSKKPGEGLEKAPKKKASAAKPKKAAAKKPAAAAKKPKKAVAVKSKPKKAKK
                     PAASATKKSVKSPKKAAKPKKAVAAKSPAKAKAVKPKAAKPKAAKPKAAKAKKAAAKK
                     K"
BASE COUNT      335 a    386 c    369 g    174 t
ORIGIN      320 bp upstream from SacII site.

  Query Match           13.4%;  Score 16;  DB 20;  Length 1264;
  Best Local Similarity 100.0%;  Pred. No. 1.92e+01;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    304 ATAAATGGGAGCGCCG 319
          ||||||||||||||||
Qy     84 ATAAATGGGAGCGCCG 99


RESULT   36
LOCUS       HUMCETP1     1656 bp    DNA                 PRI       01-NOV-1994
DEFINITION  Human cholesteryl ester transfer protein (CETP) gene, exons 1 and
            2.
ACCESSION   M32992 J02898
NID         g180261
KEYWORDS    cholesteryl ester transfer protein.
SEGMENT     1 of 7
SOURCE      Human DNA.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1656)
  AUTHORS   Agellon,L.B., Quinet,E.M., Gillette,T.G., Drayna,D.T., Brown,M.L.
            and Tall,A.R.
  JOURNAL   Unpublished (1990)
REFERENCE   2  (sites)
  AUTHORS   Agellon,L.B., Quinet,E.M., Gillette,T.G., Drayna,D.T., Brown,M.L.
            and Tall,A.R.
  TITLE     Organization of the human cholesteryl ester transfer protein gene
  JOURNAL   Biochemistry 29 (6), 1372-1376 (1990)
  MEDLINE   90241928
COMMENT     [2]  sites for [1]; intron/exon boundaries.
            Draft entry and computer-readable sequence for [2] kindly submitted
            by L.B.Agellon, 16-MAR-1990.
FEATURES             Location/Qualifiers
     source          1. .1656
                     /organism="Homo sapiens"
```

```
                     /db_xref="taxon:9606"
     TATA_signal     330. .335
     prim_transcript 361. .>1656
                     /note="CETP mRNA and introns"
     exon            <388. .505
                     /gene="CETP"
                     /note="cholesteryl ester transferase protein precursor;
                     G00-119-773"
                     /number=1
     sig_peptide     388. .438
                     /gene="CETP"
                     /note="cholesteryl ester transferase protein signal
                     peptide"
     mat_peptide     439. .505
                     /gene="CETP"
                     /note="cholesteryl ester transferase protein"
     intron          506. .1407
                     /gene="CETP"
                     /note="CETP intron A"
     exon            1408. .1522
                     /gene="CETP"
     mat_peptide     1408. .1522
                     /gene="CETP"
                     /note="cholesteryl ester transferase protein"
     intron          1523. .>1656
                     /gene="CETP"
                     /note="CETP intron B"
BASE COUNT      370 a    411 c    500 g    375 t
ORIGIN

  Query Match           13.4%;  Score 16;  DB 25;  Length 1656;
  Best Local Similarity 100.0%;  Pred. No. 1.92e+01;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     55 GAGAGAAAGCTTATTC 70
          ||||||||||||||||
Cp     36 GAGAGAAAGCTTATTC 21


RESULT   37
LOCUS       AF052209     2610 bp    DNA             BCT       20-JUN-1998
DEFINITION  Streptococcus pneumoniae TehB homolog gene, partial cds; VacB
            homolog gene, complete cds; and competence protein (coiA) gene,
            partial cds.
ACCESSION   AF052209
NID         g3211756
KEYWORDS    .
SOURCE      Streptococcus pneumoniae.
  ORGANISM  Streptococcus pneumoniae
            Eubacteria; Firmicutes; Clostridium group firmicutes;
            Streptococcaceae; Streptococcus.
```



```
REFERENCE   1  (bases 1 to 2610)
  AUTHORS   Pestova,E.V. and Morrison,D.A.
  TITLE     Isolation and characterization of three Streptococcus pneumoniae
            transformation-specific loci by use of a lacZ reporter insertion
            vector
  JOURNAL   J. Bacteriol. 180 (10), 2701-2710 (1998)
  MEDLINE   98241533
REFERENCE   2  (bases 1 to 2610)
  AUTHORS   Pestova,E.V. and Morrison,D.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAR-1998) Biology, Univ. Illinois, 900 S. Ashland
            Ave., Chicago, IL 60607, USA
FEATURES             Location/Qualifiers
     source          1. .2610
                     /organism="Streptococcus pneumoniae"
                     /strain="Rx"
                     /db_xref="taxon:1313"
     CDS             <1. .378
                     /note="orfL2"
                     /codon_start=1
                     /transl_table=11
                     /product="TehB homolog"
                     /db_xref="PID:g3211757"
                     /translation="ELPNTVEGLIHITNLPEFYHFNERDLTLRGEKSGITFRVGQQIR
                     IRVERADRMTGEIDFSFVPSEFDVIEKGLKQSSRSGRGRGSKSSLPYKKHMTRENQDA
                     QMISVSIHKKTRRKKERNLFTRK"
     CDS             397. .1722
                     /note="orfL1"
                     /codon_start=1
                     /transl_table=11
                     /product="VacB homolog"
                     /db_xref="PID:g3211758"
                     /translation="MAKGEGKVVAQNKKARHDYTIVDTLEAGMVLTGTEIKSVRAARI
                     NLKDGPAQVKGEVWLSNVHIAPYEEGNIWNQEPERRRKLLLHKKQIQKLEQETKGTG
                     MTLVPLKVYIKDGYAKLLLGLAKGKHDYDKRESIKRREQNRDIARVMKAVNQRLKEEL
                     KMEKLVAYKRMPLWNKQTMPEAVQQKHNTKVGTWGKITVLKGALKFIELTEEGEVLAE
                     HLFEAGADNPMAQPQAWHRVEAATDDVEWYLEFTCKPEDYFAKKYNTNPVHSEVLEAM
                     QTVKQGKALDLGCGQGRNSLPLAQQDFDVTAVDQNGLALEILQSIVEQEDLDMPVGLY
                     DINSASIEQEYDFIVSTVVLMFLQADRIPAIIQNMQEKTSVGGYNLSFVPWTRRIILA
                     RLTSHSPLKKENWQTITRIGNWLSTMKIQAICTVAMRMAIVFNYALRPY"
     gene            1812. .>2610
                     /gene="coiA"
     CDS             1812. .>2610
                     /gene="coiA"
                     /codon_start=1
                     /transl_table=11
                     /product="competence protein"
                     /db_xref="PID:g3211759"
                     /translation="MFVARDARGELVNVLEDKLEKQAYTCPACGGQLHLRQGPSVRTH
                     FAHKSLKDCDFFFENESPEHLANKESLYHWLKKETKVQLEYPLSELKQIADVFVNGNL
                     ALESSVVVPCLKKVLKERSEGYRSQGYQVLWLLGQKLWLKERLTRLQAGFLYFSQNMG
```



```
                     FYVWELDKGKQVLRLKYLIYQDLRGKLHYQIKEPSYGQGSLLEILRLPYKKQKISHPT
                     VSEDKDICRYIRQQLYYQNLFWMKEQAEAYQKGENILTYGLKEWYPQI"
BASE COUNT      856 a    471 c    603 g    680 t
ORIGIN

  Query Match           13.4%;  Score 16;  DB 16;  Length 2610;
  Best Local Similarity  94.1%;  Pred. No. 1.92e+01;
  Matches   16;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db    651 GCTCCATAAAAAGCAAA 667
          ||||| ||||||||||||
Qy    104 GCTCCNTAAAAAGCAAA 120


RESULT   38
LOCUS       SPBYR1       2683 bp    DNA             PLN       12-SEP-1993
DEFINITION  Fission yeast byr1 gene.
ACCESSION   X07445
NID         g4918
KEYWORDS    byr1 gene; protein kinase; sexual differentation.
SOURCE      fission yeast.
  ORGANISM  Schizosaccharomyces pombe
            Eukaryotae; mitochondrial eukaryotes; Fungi; Ascomycota;
            Archaeascomycetes; Schizosaccharomycetales;
            Schizosaccharomycetaceae; Schizosaccharomyces.
REFERENCE   1  (bases 744 to 2016)
  AUTHORS   Nadin-Davis,S.A. and Nasim,A.
  TITLE     A gene which encodes a predicted protein kinase can restore some
            functions of the ras gene in fission yeast
  JOURNAL   EMBO J. 7 (4), 985-993 (1988)
  MEDLINE   88296442
REFERENCE   2  (bases 1 to 2683)
  AUTHORS   Nadin-Davis,S.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-FEB-1989) to the EMBL/GenBank/DDBJ databases
FEATURES             Location/Qualifiers
     source          1. .2683
                     /organism="Schizosaccharomyces pombe"
                     /db_xref="taxon:4896"
                     /clone="pFL20-2."
     CDS             866. .1888
                     /note="byr1 protein; put. protein kinase (AA 1-340)"
                     /codon_start=1
                     /db_xref="PID:g4919"
                     /db_xref="SWISS-PROT:P10506"
                     /translation="MFKRRRNPKGLVLNPNASVKSSDNDHKEELINNQKSFESNVEAF
                     MEQCAHKNRRPAWISDLDNSSLEVVRHLGEGNGGAVSLVKHRNIFMARKTVYVGSDSK
                     LQKQILRELGVLHHCRSPYIVGFYGAFQYKNNISLCMEYMDCGSLDAILREGGPIPLD
                     ILGKIINSMVKGLIYLYNVLHIIHRDLKPSNVVVNSRGEIKLCDFGVSGELVNSVAQT
                     FVGTSTYMSPERIRGGKYTVKSDIWSLGISIIELATQELPWSPSNIDDSIGILDLLHC
                     IVQEEPPRLPSSFPEDLRLFVDACLHKDPTLRASPQQLCAMPYFQQALMINVDLASWA
```



```
                     SNFRSS"
BASE COUNT      861 a    437 c    471 g    914 t
ORIGIN

  Query Match           13.4%;  Score 16;  DB 23;  Length 2683;
  Best Local Similarity  94.1%;  Pred. No. 1.92e+01;
  Matches   16;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db   1237 TTTTTATGGAGCTTTTC 1253
          ||||||| ||||||||||
Cp    115 TTTTTANGGAGCTTTTC 99


RESULT   39
LOCUS       HSCETPS1     3500 bp    DNA             PRI       15-DEC-1996
DEFINITION  Human cholesteryl ester transfer protein (CETP) gene, partial cds
            and promoter region.
ACCESSION   U71187
NID         g1732066
KEYWORDS    .
SEGMENT     1 of 2
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3500)
  AUTHORS   Oliveira,H.C.F., Chouinard,R.A., Agellon,L.B., Bruce,C., Ma,L.,
            Walsh,A., Breslow,J.L. and Tall,A.R.
  TITLE     Human cholesteryl ester transfer protein gene proximal promoter
            contains dietary cholesterol positive responsive elements and
            mediates expression in small intestine and periphery while
            predominant liver and spleen expression is controlled by 5'-distal
            sequences. Cis-acting sequences mapped in transgenic mice
  JOURNAL   J. Biol. Chem. 271 (50), 31831-31838 (1996)
  MEDLINE   97112972
REFERENCE   2  (bases 1 to 3500)
  AUTHORS   Oliveira,C.F.O., Chouinard,R.A., Agellon,L.B., Bruce,C., Ma,L.,
            Walsh,A., Breslow,J.L. and Tall,A.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-SEP-1996) Medicine, Columbia University, 630 W. 168th
            St., New York, NY 10032, USA
FEATURES             Location/Qualifiers
     source          1. .3500
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /map="16q21"
     promoter        1. .3428
                     /gene="CETP"
     TATA_signal     3399. .3403
                     /gene="CETP"
```

```
     5'UTR           3429. .3455
                     /gene="CETP"
     mRNA            3429. .>3500
                     /gene="CETP"
     CDS             3456. .>3500
                     /gene="CETP"
                     /note="lipid transfer protein"
                     /codon_start=1
                     /product="cholesteryl ester transfer protein"
                     /db_xref="PID:g1732069"
                     /translation="MLAATVLTLALLGNA"
BASE COUNT      991 a    776 c    929 g    804 t
ORIGIN

  Query Match          13.4%;  Score 16;  DB 27;  Length 3500;
  Best Local Similarity 100.0%;  Pred. No. 1.92e+01;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   3122 GAGAGAAAGCTTATTC 3137
          ||||||||||||||||
Cp     36 GAGAGAAAGCTTATTC 21


RESULT   40
LOCUS       YSCPOL2RPB    4429 bp    DNA             PLN       15-MAR-1990
DEFINITION  Yeast (S.cerevisiae) RNA polymerase II 45 kD subunit (RPB3) gene,
            complete cds.
ACCESSION   M27496
NID         g172212
KEYWORDS    RNA polymerase.
SOURCE      Yeast (S.cerevisiae, strain S288C) DNA.
  ORGANISM  Saccharomyces cerevisiae
            Eukaryotae; mitochondrial eukaryotes; Fungi; Ascomycota;
            Hemiascomycetes; Saccharomycetales; Saccharomycetaceae;
            Saccharomyces.
REFERENCE   1  (bases 1 to 4429)
  AUTHORS   Kolodziej,P. and Young,R.Y.
  JOURNAL   Unpublished (1989)
REFERENCE   2  (bases 817 to 2686)
  AUTHORS   Kolodziej,P. and Young,R.A.
  TITLE     RNA polymerase II subunit RPB3 is an essential component of the
            mRNA transcription apparatus
  JOURNAL   Mol. Cell. Biol. 9, 5387-5394 (1989)
  MEDLINE   90066488
COMMENT     Draft entry and computer-readable sequence for [2] kindly provided
            by P.A.Kolodziej, 06-SEP-1989.
FEATURES             Location/Qualifiers
     source          1. .4429
                     /organism="Saccharomyces cerevisiae"
                     /db_xref="taxon:4932"
     gene            1. .4429
```



```
                     /gene="RPB3"
                     /db_xref="SGD:L0001677"
     CDS             1306. .2262
                     /gene="RPB3"
                     /note="RNA polymerase II (EC 2.7.7.6)"
                     /codon_start=1
                     /db_xref="SGD:L0001677"
                     /db_xref="PID:g172213"
                     /translation="MSEEGPQVKIREASKDNVDFILSNVDLAMANSLRRVMIAEIPTL
                     AIDSVEVETNTTVLADEFIAHRLGLIPLQSMDIEQLEYSRDCFCEDHCDKCSVVLTLQ
                     AFGESESTTNVYSKDLVIVSNLMGRNIGHPIIQDKEGNGVLICKLRKGQELKLTCVAK
                     KGIAKEHAKWGPAAGIEPEYDPWNKLKHTDTWYEQDSAKEWPQSKNCEYEDPPNEGDP
                     FDYKAQADTFTMNVESVGSIPVDQVVVRGIDTLQKKVASILLALTQMDQDKVNFASGD
                     NNTASNMLGSNEDVMMTGAEQDPYSNASQMGNTGSGGYDNAW"
BASE COUNT     1485 a    815 c    956 g   1173 t
ORIGIN

  Query Match          13.4%;  Score 16;  DB 23;  Length 4429;
  Best Local Similarity 100.0%;  Pred. No. 1.92e+01;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   3959 CAGCATTTCTGAGAGA 3974
          ||||||||||||||||
Cp     46 CAGCATTTCTGAGAGA 31


RESULT   41
LOCUS       AF017056     4700 bp    DNA             PLN       12-SEP-1997
DEFINITION  Arabidopsis thaliana brassinosteroid insensitive 1 (BRI1) gene,
            complete cds.
ACCESSION   AF017056
NID         g2392894
KEYWORDS    .
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryotae; Viridiplantae; Charophyta/Embryophyta group;
            Embryophyta; vascular plants; seed plants; Magnoliophyta;
            Magnoliopsida; Capparales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 4700)
  AUTHORS   Li,J. and Chory,J.
  TITLE     A putative leucine-rich repeat receptor kinase involved in
            brassinosteroid signal transduction
  JOURNAL   Cell 90 (5), 929-938 (1997)
  MEDLINE   97442355
REFERENCE   2  (bases 1 to 4700)
  AUTHORS   Li,J. and Chory,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-AUG-1997) Plant Biology Laboratory, The Salk
            Institute, 10010 N. Torrey Pines Rd., La Jolla, CA 92037, USA
FEATURES             Location/Qualifiers
     source          1. .4700
```



```
                     /organism="Arabidopsis thaliana"
                     /cultivar="Col-0"
                     /db_xref="taxon:3702"
                     /chromosome="IV"
                     /map="closely linked to DHS1"
     mRNA            582. .4700
                     /product="brassinosteroid insensitive 1"
     gene            693. .4283
                     /gene="BRI1"
                     /allele="cbb2, bin1, dwarf2"
     CDS             693. .4283
                     /gene="BRI1"
                     /note="leucine-rich repeat receptor kinase; involved in
                     brassinosteroid signal transduction pathway"
                     /codon_start=1
                     /product="brassinosteroid insensitive 1"
                     /db_xref="PID:g2392895"
                     /translation="MKTPSSFFLSVTTLFPFSFFSLSFQASPSQSLYREIHQLISFKD
                     VLPDKNLLPDWSSNKNPCTFDGVTCRDDKVTSIDLSSKPLNVGFSAVSSSLLSLTGLE
                     SLFLSNSHINGSVSGFKCSASLTSLDLSRNSLSGPVTTLTSLGSCSGLKFLNVSSNTL
                     DFPGKVSGGLKLNSLEVLDLSANSISGANVVGWVLSDGCGELKHLAISGNKISGDVDV
                     SRCVNLEFLDVSSNNFSTGIPPFLGDCSALQHLDISGNKLSGDFSRAISTCTELKLLNI
                     SSNQFVGPIPPLPLKSLQYLSLAENKFTGEIPDFLSGACDTLTGLDLSGNHFYGAVPP
                     FFGSCSLLESLALSSNNFSGELPMDTLLKMRGLKVLDLSFNEFSGELPESLTNLSASL
                     LTLDLSSNNFSGPILPNLCQNPKNTLQELYLQNNGFTGKIPPTLSNCSELVSLHLSFN
                     YLSGTIPSSLGSLSKLRDLKLWLNMLEGEIPQELMYVKTLETLILDFNDLTGEIPSGL
                     SNCTNLNWISLSNNRLTGEIPKWIGRLENLAILKLSNNSFSGNIPAELGDCRSLIWLD
                     LNTNLFNGTIPAAMFKQSGKIAANFIAGKRYVYIKNDGMKKECHGAGNLLEFQGIRSE
                     QLNRLSTRNPCNITSRVYGGHTSPTFDNNGSMMFLDMSYNMLSGYIPKEIGSMPYLFI
                     LNLGHNDISGSIPDEVGDLRGLNILDLSSNKLDGRIPQAMSALTMLTEIDLSNNNLSG
                     PIPEMGQFETFPPAKFLNNPGLCGYPLPRCDPSNADGYAHHQRSHGRRPASLAGSVAM
                     GLLFSFVCIFGLILVGREMRKRRRRKEAELEMYAEGHGNSGDRTANNTNWKLTGVKEA
                     LSINLAAFEKPLRKLTFADLLQATNGFHNDSLIGSGGFGDVYKAILKDGSAVAIKKLI
                     HVSGQGDREFMAEMETIGKIKHRNLVPLLGYCKVGDERLLVYEFMKYGSLEDVLHDPK
                     KAGVKLNWSTRRKIAIGSARGLAFLHHNCSPHIIHRDMKSSNVLLDENLEARVSDFGM
                     ARLMSAMDTHLSVSTLAGTPGYVPPEYYQSFRCSTKGDVYSYGVVLLELLTGKRPTDS
                     PDFGDNNLVGWVKQHAKLRISDVFDPELMKEDPALEIELLQHLKVAVACLDDRAWRRP
                     TMVQVMAMFKEIQAGSGIDSQSTIRSIEDGGFSTIEMVDMSIKEVPEGKL"
BASE COUNT     1251 a    983 c   1009 g   1457 t
ORIGIN

  Query Match          13.4%;  Score 16;  DB 24;  Length 4700;
  Best Local Similarity 100.0%;  Pred. No. 1.92e+01;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    682 ACTCTTGAGAAATGAA 697
          ||||||||||||||||
Qy      8 ACTCTTGAGAAATGAA 23


RESULT   42
```



```
LOCUS       MTRIBC       5672 bp    DNA             BCT       08-MAY-1997
DEFINITION  M.thermoautotrophicum cbiM, cbiN, cbiQ, cbiO & ribC genes.
ACCESSION   X94292
NID         g1419074
KEYWORDS    cbiM gene; cbiN gene; cbiO gene; cbiQ gene; ribC gene; Riboflavin
            synthase.
SOURCE      Methanobacterium thermoautotrophicum.
  ORGANISM  Methanobacterium thermoautotrophicum
            Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
            Methanobacterium.
REFERENCE   1  (bases 1 to 5672)
  AUTHORS   Eberhardt,S., Korn,S., Lottspeich,F. and Bacher,A.
  TITLE     Biosynthesis of riboflavin: an unusual riboflavin synthase of
            Methanobacterium thermoautotrophicum
  JOURNAL   J. Bacteriol. 179 (9), 2938-2943 (1997)
  MEDLINE   97284497
REFERENCE   2  (bases 1 to 5672)
  AUTHORS   Eberhardt,S.M.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-DEC-1995) S.M.R. Eberhardt, TU Muenchen, Lehrstuhl
            III f. Org.Chemie u.Biochemie, Lichtenbergstr 4, D-85747 Garching,
            FRG
  REMARK    Revised by [3]
REFERENCE   3  (bases 1 to 5672)
  AUTHORS   Eberhardt,S.M.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-APR-1996) S.M.R. Eberhardt, TU Muenchen, Lehrstuhl
            III f. Org.Chemie u.Biochemie, Lichtenbergstr 4, D-85747 Garching,
            FRG
FEATURES             Location/Qualifiers
     source          1. .5672
                     /organism="Methanobacterium thermoautotrophicum"
                     /strain="Marburg"
                     /db_xref="taxon:2166"
     gene            793. .1452
                     /gene="cbiM"
     CDS             793. .1452
                     /gene="cbiM"
                     /function="active transport of cobalt ions"
                     /codon_start=1
                     /transl_table=11
                     /db_xref="PID:e237182"
                     /db_xref="PID:g1419075"
                     /db_xref="SPTREMBL:Q50798"
                     /translation="MEGFLPPEWRLFWYLLSAPVIVYGLFRIIRISDERPEAKPLLAV
                     SGAFMFVLSSLKLPSVTGSCSHPTGNGLGAVLFGPAVTSVLALIVLIFQALLLAHGGI
                     TTLGANVFSMGIIGPVCAWGIYRITRSAIPSSAAVFLAAFLRDLMTFVTTSIQLALAF
                     PKPGFMEALVKFMVIFAYTQVPLAVAEGLLTVVIFENILKLKPDIMEKLQVIGKPAVE
                     G"
     gene            1454. .1741
                     /gene="cbiN"
```

```
     CDS             1454. .1741
                     /gene="cbiN"
                     /function="active transport of cobalt ions"
                     /codon_start=1
                     /transl_table=11
                     /db_xref="PID:e237183"
                     /db_xref="PID:g1419076"
                     /db_xref="SPTREMBL:Q50799"
                     /translation="MDKRHTIMLIAVAVIAIAPLVIYSGLGEDQGYPGGADDSASKAI
                     SETGYKPWFQPIWEPPSGEIESLLFALQAAIGALIIGYVPGYIRGRGESPE"
     gene            1764. .2528
                     /gene="cbiQ"
     CDS             1764. .2528
                     /gene="cbiQ"
                     /function="active transport of cobalt ions"
                     /codon_start=1
                     /transl_table=11
                     /db_xref="PID:e237184"
                     /db_xref="PID:g1419077"
                     /db_xref="SPTREMBL:Q50800"
                     /translation="MNISVDHYAHTNGLRNTSPGFKMALAIITMLISVLSPNPVIPLI
                     IAFLMTLLILFAADIPASYYLRFAAAIPVGFGVLTLVLMALPFGVNPLVTFMGLRIYSD
                     GLHTGILVFSRILGGFSCLAFLALTTPLNEVFHELGRLGLPAAFIEIALLMYRSIFVF
                     LEEASVMYNAQDSRLGYSGIKNSYRSLGLLAGNLFIRSWLRGETVYRSMEGRCYQGTM
                     PSIRRESIGAREAVMIALFDGLLLVGLLFTSKGVIF"
     gene            2615. .3448
                     /gene="cbiO"
     CDS             2615. .3448
                     /gene="cbiO"
                     /function="active transport of cobalt ions"
                     /codon_start=1
                     /transl_table=11
                     /db_xref="PID:e237185"
                     /db_xref="PID:g1419078"
                     /db_xref="SPTREMBL:Q50801"
                     /translation="MKIIEAVNIRYTYPDGTEALKGIDFHAKKGEVVALLGPNGAGKS
                     TLFLHFNGILQPTSGKVLVDGKEIDTSKSGLIKVRQKVGIVFQNPDDQLFAPRVDEDV
                     AFGPLNMGLDEDEVEERVKDALRKVGMSGTRTDPHHLSGGEKKRVAIAGILAMKPDIM
                     ILDEPTSGLDPRGASQILRLLHELNEEGMTIIISTHDVDLAPLYSDRIYIISGGEIIS
                     EGTPGDVFSDIDTIRGADLRLPRIAHLVEILQKEDDLPFGEPTPLTIGEARRKLRDQL
                     K"
     gene            3458. .3919
                     /gene="ribC"
     CDS             3458. .3919
                     /gene="ribC"
                     /EC_number="2.5.1.9"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="riboflavin synthase"
                     /db_xref="PID:e218505"
```



```
                     /db_xref="PID:g1419079"
                     /db_xref="SPTREMBL:Q59587"
                     /translation="MIKVGICDTTPARYDMGGAAIDEIKKHATGIKIIRRTVPGIKDL
                     PVACKKLIEEEGCEMVMALGMPGPEEKDKVCAHEASTGLIQAQLMTNTHILEVFVHED
                     EEDDPEDLKVLADNRAREHAQNLIMMLFRPERLTRDAGMGMREGKPDVGPL"
BASE COUNT     1455 a   1301 c   1485 g   1431 t
ORIGIN

  Query Match           13.4%;  Score 16;  DB 15;  Length 5672;
  Best Local Similarity 100.0%;  Pred. No. 1.92e+01;
  Matches    16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   3703 CCTTATACAGGCTCAG 3718
           ||||||||||||||||
Cp     82 CCTTATACAGGCTCAG 67


RESULT   43
LOCUS       AF027656     5887 bp    DNA             PRI       20-NOV-1997
DEFINITION  Homo sapiens cholesteryl ester transfer protein gene, promoter
            region.
ACCESSION   AF027656
NID         g2625008
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5887)
  AUTHORS   Williams,S., Hayes,L., Elsenboss,L., Williams,A., Andre,C.,
            Abramson,R., Thompson,J.F. and Milos,P.M.
  TITLE     Sequencing of the cholesteryl ester transfer protein 5' regulatory
            region using artificial transposons
  JOURNAL   Gene 197 (1-2), 101-107 (1997)
  MEDLINE   97473500
REFERENCE   2  (bases 1 to 5887)
  AUTHORS   Williams,S., Hayes,L., Elsenboss,L., Williams,A., Andre,C.,
            Abramson,R., Thompson,J.F. and Milos,P.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-SEP-1997) Molecular Sciences, Pfizer Central
            Research, Eastern Point Road, Groton, CT 06340, USA
FEATURES             Location/Qualifiers
     source          1. .5887
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /map="16q21"
     promoter        1. .5780
                     /note="cholesteryl ester transfer protein gene"
     repeat_region   <1. .142
                     /note="Mermaid repeat"
```



```
     repeat_region   complement(283. .437)
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   939. .1196
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   1565. .1851
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   complement(2021. .2313)
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   2468. .2761
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   3360. .3646
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   3640. .3837
                     /note="polymorphic GA repeat"
                     /rpt_type=tandem
     mRNA            5781. .>5887
                     /product="cholesteryl ester transfer protein"
     CDS             5808. .>5887
                     /codon_start=1
                     /product="cholesteryl ester transfer protein"
                     /db_xref="PID:g2625009"
                     /translation="MLAATVLTLALLGNAHACSKGTSHEAG"
BASE COUNT     1693 a   1308 c   1490 g   1396 t
ORIGIN

  Query Match           13.4%;  Score 16;  DB 26;  Length 5887;
  Best Local Similarity 100.0%;  Pred. No. 1.92e+01;
  Matches    16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   5474 GAGAGAAAGCTTATTC 5489
           ||||||||||||||||
Cp     36 GAGAGAAAGCTTATTC 21


RESULT   44
LOCUS       CHKY         8372 bp    DNA             VRT       22-JUN-1995
DEFINITION  Gallus gallus ovalbumin (Y) gene, complete cds.
ACCESSION   J00922 V00439
NID         g212899
KEYWORDS    ovalbumin.
SOURCE      Gallus gallus DNA.
  ORGANISM  Gallus gallus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Archosauria; Aves; Neognathae; Galliformes;
            Phasianidae; Phasianinae; Gallus.
```



```
REFERENCE   1  (bases 1 to 2237)
  AUTHORS   Heilig,R., Muraskowsky,R. and Mandel,J.L.
  TITLE     The ovalbumin gene family. The 5' end region of the X and Y genes
  JOURNAL   J. Mol. Biol. 156 (1), 1-19 (1982)
  MEDLINE   82242056
REFERENCE   2  (bases 1614 to 8372)
  AUTHORS   Heilig,R., Muraskowsky,R., Kloepfer,C. and Mandel,J.L.
  TITLE     The ovalbumin gene family: complete sequence and structure of the Y
            gene
  JOURNAL   Nucleic Acids Res. 10 (14), 4363-4382 (1982)
  MEDLINE   83014329
COMMENT     This sequence shows approximately 73% homology with the ovalbumin
            and ovalbumin-x genes. See <chkoval>,<chkx>. Homology is strongest
            and exonic DNA suggesting strong selection pressure for
            conservation. A leader exon and seven exons were determined by
            electron microscopy and comparison with the ovalbumin region.
            Dinucleotide frequencies and purine and pyrimidine stretches are
            reported and discussed.
FEATURES             Location/Qualifiers
     source          1. .8372
                     /organism="Gallus gallus"
                     /db_xref="taxon:9031"
     exon            1614. .1663
                     /gene="oval-Y"
                     /number=1
     mRNA            join(1614. .1663,3421. .3602,4148. .4198,4554. .4682,
                     5530. .5647,5869. .6011,6093. .6248,7125. .8267)
                     /gene="oval-Y"
     gene            join(1614. .1663,3421. .3602,4148. .4198,4554. .4682,
                     5530. .5647,5869. .6011,6093. .6248,7125. .8267)
                     /gene="oval-Y"
     intron          1664. .3420
                     /gene="oval-Y"
                     /number=1
     exon            3421. .3602
                     /partial
                     /gene="oval-Y"
                     /number=2
     CDS             join(3435. .3602,4148. .4198,4554. .4682,5530. .5647,
                     5869. .6011,6093. .6248,7125. .7526)
                     /gene="oval-Y"
                     /note="y protein"
                     /codon_start=1
                     /product="ovalbumin"
                     /db_xref="PID:g212900"
                     /translation="MDSISVTNAKPCFDVFNEMKVHHVNENILYCPLSILTALAMVYL
                     GARGNTESQMKKVLHFDSITGAGSTTDSQCGSSEYVHNLFKELLSEITRPNATYSLEI
                     ADKLYVDKTFSVLPEYLSCARKFYTGGVEEVNFKTAAEEARQLINSWVEKETNGQIKD
                     LLVSSSIDPGTTMVFINTIYFKGIWKIAFNTEDTREMPFSMTKEESKPVQMMCMNNSF
                     NVATLPAEKMKILELPYASGDLSMLVLLPDEVSGLERIEKTINFDKLREWTSTNAMAR
                     KSMKVYLPRMKIEEKYNLTSILMALGMTDLFSRSANLTGISSVDNLMISDAVHGVFME
```

```
                     VNEEGTEATGSTGAIGNIKHSLELEEFRADHPFLPFIRYNPTNAILFPGRYWSP"
     intron          3603. .4147
                     /gene="oval-Y"
                     /number=2
     exon            4148. .4198
                     /gene="oval-Y"
                     /number=3
     intron          4199. .4553
                     /gene="oval-Y"
                     /number=3
     exon            4554. .4682
                     /gene="oval-Y"
                     /number=4
     intron          4683. .5529
                     /gene="oval-Y"
                     /number=4
     exon            5530. .5647
                     /gene="oval-Y"
                     /number=5
     intron          5648. .5868
                     /gene="oval-Y"
                     /number=5
     exon            5869. .6011
                     /gene="oval-Y"
                     /number=6
     intron          6012. .6092
                     /gene="oval-Y"
                     /note="oval-y intron F"
                     /number=6
     exon            6093. .6248
                     /gene="oval-Y"
                     /number=7
     intron          6249. .7124
                     /gene="oval-Y"
                     /number=7
     exon            7125. .8267
                     /gene="oval-Y"
                     /number=8
     polyA_signal    8249. .8254
                     /gene="oval-Y"
                     /note="putative"
BASE COUNT     2475 a    1571 c    1556 g    2770 t
ORIGIN      About 2.1 kb upstream from HindIII site.

  Query Match           13.4%;  Score 16;  DB 20;  Length 8372;
  Best Local Similarity 100.0%;  Pred. No. 1.92e+01;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   6799 ACAGCATTTCTGAGAG 6814
         ||||||||||||||||
Cp     47 ACAGCATTTCTGAGAG 32
```



```
RESULT   45
LOCUS       PFSC03016   19574 bp    DNA             INV       20-MAR-1998
DEFINITION  Plasmodium falciparum DNA *** SEQUENCING IN PROGRESS *** from
            contig 3-16, complete sequence.
ACCESSION   AL008970
NID         g2982534
KEYWORDS    HTG.
SOURCE      malaria parasite.
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 19574)
  AUTHORS   Lawson,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAR-1998) P.falciparum Genome Sequencing Consortium,
            The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA, UK
COMMENT     On Mar 25, 1998 this sequence version replaced gi:2894447.
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and the release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc.
            For more information about this sequence or the Malaria Project,
            see http://www.sanger.ac.uk/Projects/P_falciparum.
FEATURES             Location/Qualifiers
     source          1. .19574
                     /organism="Plasmodium falciparum"
                     /db_xref="taxon:5833"
                     /chromosome="3"
                     /clone="3-16"
BASE COUNT     7658 a    1961 c    1856 g    8099 t
ORIGIN

  Query Match           13.4%;  Score 16;  DB 18;  Length 19574;
  Best Local Similarity 100.0%;  Pred. No. 1.92e+01;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   7204 AAGCTTATTCATTTCT 7219
         ||||||||||||||||
Cp     30 AAGCTTATTCATTTCT 15


RESULT   46
LOCUS       CEF43D2     27777 bp    DNA             INV       21-SEP-1998
DEFINITION  Caenorhabditis elegans cosmid F43D2, complete sequence.
ACCESSION   Z83226
NID         g1729565
KEYWORDS    HTG.
```



```
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 27777)
  AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
  TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
  JOURNAL   Nature 368 (6466), 32-38 (1994)
  MEDLINE   94150718
REFERENCE   2  (bases 1 to 27777)
  AUTHORS   Mortimore,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-NOV-1996) Nematode Sequencing Project, Sanger Centre,
            Hinxton, Cambridge CB10 1RQ, England and Department of Genetics,
            Washington University, St. Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     Coding sequences below are predicted from computer analysis, using
            predictions from Genefinder (P. Green, U. Washington), and other
            available information.
            For a graphical representation of this sequence and its analysis
            see:-
            http://webace.sanger.ac.uk/cgi-
            bin/display?db=wormace&class=Sequence &object=F43D2
            Current sequence finishing criteria for the C. elegans genome
            sequencing consortium are that all bases are either sequenced
            unambiguously on both strands, or on a single strand with both a
            dye primer and dye terminator reaction, from distinct subclones.
            Exceptions are indicated by an explicit note.
            IMPORTANT: This sequence is NOT necessarily the entire insert of
            the specified clone.  It may be shorter because we only sequence
            overlapping sections once, or longer because we arrange for a small
            overlap between neighbouring submissions.
            IMPORTANT: This sequence is not the entire insert of clone F43D2.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we arrange for a small overlap between
            neighbouring submissions.
            The true left end of clone F43D2 is at 19281 in
            sequence Z93386.
            The start of this sequence (1. .103) overlaps with the end of
            sequence Z93386.
```



```
            The end of this sequence (27674. .27777) overlaps with the start of
            sequence Z75532.
FEATURES             Location/Qualifiers
     source          1. .27777
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /chromosome="V"
                     /clone="F43D2"
     gene            11. .1043
                     /gene="F43D2.1"
     CDS             join(11. .64,99. .199,249. .333,378. .503,547. .669,756. .1043)
                     /gene="F43D2.1"
                     /note="Weak similarity to Human G1/S-specific cyclin C
                     (SW:P24863)"
                     /codon_start=1
                     /db_xref="PID:e329937"
                     /db_xref="PID:g2266971"
                     /db_xref="SPTREMBL:O17869"
                     /translation="MSDWIWPLEALKARTPNSTTPSIRAGLTKEQELLWRREGIKLLS
                     EVGNALNCKPRPTIGVAAVYFHEFYMEHSFQSFSREVTALSCLFLAGKVEDFPKKCKD
                     VCQAAVTHYPEIYSKYQNLVDDVMGLERVLLHSLKFDLHVALPYDALLDYKMMFPDMN
                     REKITDAVQIAWTFINDSIYTTLCITTEPQMIAIALLHLAFTVKGYQPVQKNMDPCWW
                     SADVSNWPQESVDKACHLVLDFYAATKEKPVLEKKKLTTP"
     gene            19345. .20560
                     /gene="F43D2.3"
     CDS             join(19345. .19476,19726. .19971,20247. .20366,20423. .20560)
                     /gene="F43D2.3"
                     /note="predicted using Genefinder"
                     /codon_start=1
                     /db_xref="PID:e284771"
                     /db_xref="PID:g1729571"
                     /translation="MNKLEAKLRKLQGSLRGCAKMIAMNTGRKRKLTEDLPDTPDDNQ
                     TTSTNMSPRRSVRIRRMPVRYGSTMQPPATSNGQRNTRGQGPKPKGRAATRNKSPDNS
                     PRMYSSGQVDRASGSVLPVAHDNRTLDRDSAQNQIDAPQVMTAPVAPTANKIVVPGTT
                     PNGKTMIQMTIDMKVVVGQLLKGSDQPVELVPKKIVVAGQECTVDSKSKND"
     gene            complement(23036. .24616)
                     /gene="F43D2.2"
     CDS             complement(join(23036. .23195,23443. .23550,23599. .23689,
                     23770. .23866,24024. .24131,24178. .24315,24542. .24616))
                     /gene="F43D2.2"
                     /codon_start=1
                     /db_xref="PID:e329938"
                     /db_xref="PID:g2266972"
                     /db_xref="SPTREMBL:O17870"
                     /translation="MPIGFAASEMYNFEKGTPLDKKTCEKNLEKLVKYDREVDYHEWL
                     NLAQQCRTENQHLMEKMAPGMAFTSAMSNKLHKQVGEKTDDRKPTGSNDVTFLIESAG
                     PWDREIDDSDPHMMCYTTSKNASTYIPIDKDFFNGISDCLIAGPGQEHSKDRKLQVRK
                     LITEVMSYKLSRVRAHYNSINGNTSSEKAAKWMREEIAKEDGELKECQLTLSHASKTS
                     HWLEDLPQKDKKLIDSSIKCLQLQWENLNGKVPFIVHCDD"
BASE COUNT     8683 a    5136 c    4774 g    9184 t
ORIGIN
```

```
  Query Match          13.4%;  Score 16;  DB 18;  Length 27777;
  Best Local Similarity 100.0%;  Pred. No. 1.92e+01;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  19341 AGAAATGAATAAGCTT 19356
          ||||||||||||||||
Qy     15 AGAAATGAATAAGCTT 30


RESULT   47
LOCUS       AC001230    29117 bp    DNA             ROD       29-MAY-1997
DEFINITION  Genomic sequence from Mouse 11, complete sequence.
ACCESSION   AC001230
NID         g2133905
KEYWORDS    HTG.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae;
            Murinae; Mus.
REFERENCE   1  (bases 1 to 29117)
  AUTHORS   Hawkins,T.L., Reeve,M.P., Christoffersen,A., Birren,B.W.,
            Fasman,K.H. and Lander,E.S.
  TITLE     Genomic sequence from Mouse 11
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 29117)
  AUTHORS   Hawkins,T.L., Reeve,M.P., Christoffersen,A., Birren,B.W.,
            Fasman,K.H., Lander,E.S., McKernan,K., Munro,C., Richardson,P.,
            Barna,N., Chang,A., Cooke,P., Daly,M.J., Forrest,C., Fripp,W.J.,
            Gage,D., Geraigery,K., Hagos,B., Jacotot,L., Lane,M., MacKenzie,J.,
            Marquis,N., McDermott,J., Moloney,N., Morrow,J., Nachman,A.,
            Naylor,J., O'Connor,T., Peterson,K., Rollins,G., Spencer,J.,
            Stilwell,J., Stone,C., Strickland,C., Sydney,K., Wilmer,F. and
            Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-APR-1997) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 29117)
  AUTHORS   Hawkins,T.L., Reeve,M.P., Christoffersen,A., Birren,B.W.,
            Fasman,K.H., Lander,E.S., McKernan,K., Munro,C., Richardson,P.,
            Barna,N., Chang,A., Cooke,P., Daly,M.J., Forrest,C., Fripp,W.J.,
            Gage,D., Geraigery,K., Hagos,B., Jacotot,L., Lane,M., MacKenzie,J.,
            Marquis,N., McDermott,J., Moloney,N., Morrow,J., Nachman,A.,
            Naylor,J., O'Connor,T., Peterson,K., Rollins,G., Spencer,J.,
            Stilwell,J., Stone,C., Strickland,C., Sydney,K., Wilmer,F. and
            Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAY-1997) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 29, 1997 this sequence version replaced gi:1943576.
```



```
            The Staden databases, finishing information, and all
            chromatographic files used in the assembly of this clone are
            available from our anonymous ftp site.

            All repeats were identified using RepeatMasker: Smit, A.F.A. &
            Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html.
FEATURES            Location/Qualifiers
     source         1. .29117
                    /organism="Mus musculus"
                    /db_xref="taxon:10090"
                    /clone="JS1"
                    /clone_lib="unknown"
                    /chromosome="11"
     repeat_region  complement(144. .276)
                    /rpt_family="MMB1"
     repeat_region  complement(277. .308)
                    /rpt_family="AT_rich"
     repeat_region  complement(480. .550)
                    /rpt_family="MMB1F"
     repeat_region  1474. .1620
                    /rpt_family="MMB1"
     repeat_region  complement(2072. .2160)
                    /rpt_family="MMFLAM"
     repeat_region  complement(2092. .2160)
                    /rpt_family="MMB1F"
     repeat_region  complement(2581. .3123)
                    /rpt_family="MML"
     repeat_region  complement(3652. .3786)
                    /rpt_family="MMB2"
     repeat_region  complement(3830. .3940)
                    /rpt_family="(TAAAA)n"
     repeat_region  complement(3882. .4027)
                    /rpt_family="(CAAA)n"
     repeat_region  complement(4030. .4127)
                    /rpt_family="MMFLAM"
     repeat_region  complement(5450. .5570)
                    /rpt_family="MML"
     repeat_region  6592. .6733
                    /rpt_family="MMB1"
     repeat_region  complement(6932. .6991)
                    /rpt_family="AT_rich"
     repeat_region  7017. .7135
                    /rpt_family="MMB1"
     repeat_region  complement(7306. .7451)
                    /rpt_family="MMB2"
     repeat_region  complement(7605. .7947)
                    /rpt_family="MTB"
     repeat_region  complement(7951. .8140)
                    /rpt_family="L1ME3A"
     repeat_region  8251. .8307
```



```
                    /rpt_family="(CA)n"
     repeat_region  complement(9006. .9085)
                    /rpt_family="MMFLAM"
     repeat_region  complement(9098. .9197)
                    /rpt_family="MMFLAM"
     repeat_region  complement(9367. .9436)
                    /rpt_family="ID4"
     repeat_region  9647. .9830
                    /rpt_family="MMB2"
     repeat_region  complement(9972. .10146)
                    /rpt_family="MMB2"
     repeat_region  12188. .12252
                    /rpt_family="ID2"
     repeat_region  13245. .13374
                    /rpt_family="MER42c"
     repeat_region  13256. .13413
                    /rpt_family="L1MC3"
     repeat_region  13649. .13719
                    /rpt_family="ID3"
     repeat_region  13743. .13838
                    /rpt_family="MMFLAM"
     repeat_region  14827. .14863
                    /rpt_family="(GA)n"
     repeat_region  complement(14892. .14958)
                    /rpt_family="MMFLAM"
     repeat_region  complement(15779. .15832)
                    /rpt_family="MMB1F"
     repeat_region  complement(15780. .15869)
                    /rpt_family="MMFLAM"
     repeat_region  complement(16685. .16777)
                    /rpt_family="MMFLAM"
     repeat_region  16836. .16900
                    /rpt_family="ID1"
     repeat_region  17837. .17938
                    /rpt_family="MMB1F"
     repeat_region  complement(19154. .19264)
                    /rpt_family="MMB1"
     repeat_region  complement(19154. .19299)
                    /rpt_family="ORR1B-INT"
     repeat_region  complement(19289. .19336)
                    /rpt_family="MMFLAM"
     repeat_region  complement(20342. .20368)
                    /rpt_family="(CAAA)n"
     repeat_region  complement(20369. .20495)
                    /rpt_family="MMB1"
     repeat_region  complement(20508. .20646)
                    /rpt_family="MMB1"
     repeat_region  20685. .20854
                    /rpt_family="MMB2"
     repeat_region  21073. .21148
                    /rpt_family="MMB1F"
```



```
     repeat_region  21073. .21176
                    /rpt_family="MMFLAM"
     repeat_region  21397. .21583
                    /rpt_family="MMB2"
     repeat_region  21586. .21701
                    /rpt_family="MMFLAM"
     repeat_region  21706. .21756
                    /rpt_family="(GAAA)n"
     repeat_region  21998. .22144
                    /rpt_family="MMB1"
     repeat_region  complement(22287. .22316)
                    /rpt_family="AT_rich"
     repeat_region  22500. .22599
                    /rpt_family="MMB1F"
     repeat_region  complement(22605. .22635)
                    /rpt_family="(CA)n"
     repeat_region  complement(23513. .23575)
                    /rpt_family="POLY_A"
     repeat_region  complement(24118. .24253)
                    /rpt_family="MIR"
     repeat_region  25070. .25165
                    /rpt_family="MMB1F"
     repeat_region  25173. .25275
                    /rpt_family="(GGGA)n"
     repeat_region  25632. .25754
                    /rpt_family="MER21B"
     repeat_region  complement(27060. .27116)
                    /rpt_family="(GA)n"
     repeat_region  27284. .27384
                    /rpt_family="MMB2"
     repeat_region  27388. .27439
                    /rpt_family="MMFLAM"
     repeat_region  complement(27879. .27947)
                    /rpt_family="MMSQR2"
     repeat_region  complement(28576. .28721)
                    /rpt_family="MMB1"
     repeat_region  28796. .28865
                    /rpt_family="MIR2"
     repeat_region  28822. .28864
                    /rpt_family="MIR"
     repeat_region  28983. .29065
                    /rpt_family="MMB1"
BASE COUNT     8464 a   5626 c   5931 g   9096 t
ORIGIN

  Query Match          13.4%;  Score 16;  DB 28;  Length 29117;
  Best Local Similarity  94.1%;  Pred. No. 1.92e+01;
  Matches   16;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db  1315 TTTTTAAGGAGCTTTTC 1331
         |||||| ||||||||||
```

```
Cp    115 TTTTTANGGAGCTTTTC 99


RESULT   48
LOCUS       CELT19A5    30722 bp    DNA             INV       04-APR-1996
DEFINITION  Caenorhabditis elegans cosmid T19A5.
ACCESSION   U53153
NID         g1255408
KEYWORDS    .
SOURCE      Caenorhabditis elegans strain=Bristol N2.
  ORGANISM  Caenorhabditis elegans
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Nematoda;
            Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea;
            Rhabditidae; Caenorhabditis.
REFERENCE   1  (bases 1 to 30722)
  AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
  TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
  JOURNAL   Nature 368 (6466), 32-38 (1994)
  MEDLINE   94150718
REFERENCE   2  (bases 1 to 30722)
  AUTHORS   Bradshaw,H.
  TITLE     The sequence of C. elegans cosmid T19A5
  JOURNAL   Unpublished (1996)
REFERENCE   3  (bases 1 to 30722)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-1996) Robert Waterston
COMMENT               Submitted by:
                      Genome Sequencing Center
                      Department of Genetics, Washington University,
                      St. Louis, MO  63110, USA, and
                      Sanger Centre, Hinxton Hall
                      Cambridge CB10 IRQ, England
                      e-mail: rw@nematode.wustl.edu and jes@sanger.ac.uk
                      NEIGHBORING COSMID INFORMATION:
            The 5' end lies in a gap followed by the cosmid ZC411;3' cosmid is
            F59E11.  Actual start of this cosmid is at base position 1 of
            CELT19A5

            NOTES:
```



```
            Coding sequences below are predicted from computer analysis, using
            the program Genefinder(P. Green and L. Hillier, ms in preparation).
FEATURES            Location/Qualifiers
     source         1. .30722
                    /organism="Caenorhabditis elegans"
                    /strain="Bristol N2"
                    /db_xref="taxon:6239"
                    /chromosome="V"
                    /clone="T19A5"
     gene           complement(582. .3782)
                    /gene="T19A5.3"
     CDS            complement(join(582. .767,942. .1558,1688. .1800,1846. .2401,
                    2497. .2652,2744. .2885,3253. .3489,3531. .3782))
                    /gene="T19A5.3"
                    /note="coded for by C. elegans cDNA yk14g12.3; coded for
                    by C. elegans cDNA yk92c9.5; coded for by C. elegans cDNA
                    yk87e4.5; coded for by C. elegans cDNA yk92c9.3;  similar
                    to mouse bullous pemphigoid antigen, BPAG2 (PIR:A46053)"
                    /codon_start=1
                    /db_xref="PID:g1255409"
                    /translation="MGAPYVVAHSSTQVLSHLRISKSKPQNLVCKPACPLSRSRKREA
                    SELEEERQSRFSHSRFHPETKLSSPAAMTSLISSFLFQSKDQVFVPRAVAPPRASRPI
                    VTSSTTESAESSAETVPTTYVVQTLTKYGKSLDEARRKRIEARWKRLGINFKTINHPL
                    LNTESSESSHSGHHYTIRQLNAPKKVEKLVEVERPAIDQPILHVVNPLDQGFVPSEPE
                    PQPILPAELPISGNYHSRTKPILKGVKFVQEKLSVRPRLRIRRPGEKAELPMPEDPNL
                    TPFMNNARAKFGYGTTTLPPETTTIYVPPTTASRRPTTTTTTEAPVEASPGISGGDS
                    GIIDTGDSGIVGNGESGLVGTGDSGIVGNGDSGVLGPPGDSGAGPGPGPSELPPEPQN
                    IGPGVGGGPQGFNFNVPADKMEPKEETTTTTAAPTTTTPATTTMKPVPKKPKHVQRKQ
                    PNAAIDKDEPVVIPQNSGLRPVSPPKEFAGAGGPGSPGGGGGGGGPGGGGGGFGGGGG
                    GFGSGSPFGGAAPSPPSGPDPDFGGDTAAAGPTEEEFFTGESALSRSKGPTGDGYGPP
                    VFPGGAPPPPVPAVGLGGAAAGKSPYSGQEVDQVENTGTTVKPSALLSVLSKADTGFN
                    QVIDHFENGTPPEAAFIDILEVALGSQKLDSQAKLLGHVDRTIGLDNLQRLQRWANTA
                    GAMDVFKDQENNFFSSPTLNSPNLFFSFYFTKFNSPESHYLTHSLLIVLFPTFFFLLL
                    HFHCISIDLKTG"
     gene           8632. .16235
                    /gene="T19A5.2"
     CDS            join(8632. .8709,12679. .12903,13283. .13448,13839. .13996,
                    14365. .14553,14608. .14705,14759. .14963,15345. .16130,
                    16179. .16235)
                    /gene="T19A5.2"
                    /note="coded for by C. elegans cDNA yk34b11.5; coded for
                    by C. elegans cDNA yk19h11.5; coded for by C. elegans cDNA
                    yk21f8.5; coded for by C. elegans cDNA yk8f12.5; coded for
                    by C. elegans cDNA yk29b9.5;  Similar to serine.threonine
                    kinase."
                    /codon_start=1
                    /db_xref="PID:g1255410"
                    /translation="MTTTSSDELPRQADDDSMKWDRIYIQKLDPEVIFTKQERIGRGS
                    FGEVYKGIDNRTGRVVAIKIIDLEQAEDEIEDIQQEIQVLSQCDSQYVTKYFGSFLKG
                    SKLWIIMEYLGGGSALDLTKSGKLDESHIAVILREILKGLEYLHSERKIHRDIKAANV
                    LVSEHGDVKVADFGVAGQLTETVKKRITFVGSPFWMAPELIKQSSYDYKADIWSLGIT
```



```
                    AIELANGEPPHSDLHPMRVLFLIPKNPPPVLQGSQWSKPFKEFVEMCLNKDPENRPSA
                    STLLKHQFIKRAKKNSILVDLIERAAETYLRTGVSSDSDLDEDSDGGGGTSKWDYPTV
                    RGPRVSADDDGTVRQRTDRPRAQVDRRSPSGSPGGTIVRGSPQVAAVAEQLRNSSVGS
                    SGYGSGGNSASSQYATSSLPQSHTASSGGATTITLGSPNGSPTSSLARTQSMVSPSGQ
                    RSGSAQSWELERGNRPMSERVSSQVSPSKYNQHRTSSSNGVQGGSGGRREYINGSGSG
                    LNGNSSNQNHSEYSDAVRQRGPGGSGGLDYRESHVPTSSQENLNHGRMYGTGAPPPS
                    REANVVPMPRVKGALDCSLLPAIEHLSRTRHATAALDQLRHVFRDVEDSCPGICNELI
                    EELMQRIAVPQVSQSDLDAAIRRLTTPPS"
     gene           17892. .22571
                    /gene="T19A5.1"
     CDS            join(17892. .18005,18054. .18241,18370. .18571,18616. .18777,
                    18827. .18933,18983. .19112,19161. .19367,19417. .19702,
                    19753. .19835,19887. .20098,20175. .20318,20366. .20483,
                    20531. .20683,20734. .21186,21237. .22128,22180. .22571)
                    /gene="T19A5.1"
                    /note="coded for by C. elegans cDNA yk12c12.3; coded for
                    by C. elegans cDNA yk88b1.3; coded for by C. elegans cDNA
                    yk88a5.5; coded for by C. elegans cDNA yk133h6.3; coded
                    for by C. elegans cDNA yk12c12.5; coded for by C. elegans
                    cDNA yk133h6.5; coded for by C. elegans cDNA yk88b1.5;
                    coded for by C. elegans cDNA CEESJ05F;  one short region
                    of weak similarity to S. cerevisiae protease A inhibitor 3
                    (SP:P01094) and another short region of weak similarity to
                    S. cerevisiae glucose repression mediator protein
                    (SP:P14922)"
                    /codon_start=1
                    /db_xref="PID:g1255411"
                    /translation="MNRKFPAWLPTLLILNNDVVLSAYHWKISEDGKKIQAVPDTRYH
                    LKLPGSLTDFITQVENVKVVDRQANELHGSMGKIKNHENFDNPKPEEDFRRQEEACVK
                    SRGIRLQSDSFHTSFTAELCPEFHESYQQYTAFVKEYFGSPPDSDKRPPRCNYWPEAV
                    KMKGQNNVDLGKSLQDKYDADDDMYLKVLDLYFPGFAKFTRGSEGMSRPGNMLSSMLS
                    HADERVSRNAYLHMAAAAYWRAQGKISDAFTCYATGIHFVNDENRAPHIGAFHPSHAE
                    HTIRLSIATLLNRAGYSVEAHIILKQLTSTNTVACYSGMMHAATGDASVLTGPESAID
                    YDTEQISIQLKLTNKERKSMFSNTFYNYAKAASDFKAAKKEHNQKDGELFDQSFDSVE
                    SKASVISCHLRLYSALEKQKKNLEELVVEKTNPQESYNKRVEMNSNILSLMPSAEARA
                    NSLLFYDQIKYGAHPHRKCQRAQINRFPATNHKLNWATTMICSLGCTPEEYEESMAPL
                    REKTLRNRPAINETSWKVTEVYPPPNAFEKTLFDISRNSQQDGLPFEMIEPRFTNKEID
                    ESVDRFWRRADWPNSVDCQAVVSSSNPFRINTFPQVFISPENRGWLVSELLTTHIGLQ
                    SADSLPLPWFEPRCDFVKMDDLRVFSGFEAIERLLNPRFDRKSEGYAEKTLKTTFVRL
                    ADRVIEEHEMAARIQLLMKHTIGPRWVSLNLAGLYWRIQGNPHQATKCLLGAFLDKPE
                    ESFVSIVQLAQVMLKSTRMVNDAHGFLNQQLNLLGYSEPLFHYVQGRMKLLLHDVDKA
                    IQHLKDAMDKDPNNLVIQEDLLKIACSGKSTKSAISSQFPTVCCSPVVQNAVCIRPRK
                    DSEEQCYIVESSNIPDASPQLVYHRCNGFYDGFSKKTADYASIVSPFLVFNTITPRN
                    DVTNWVNQVDGIQTVESTEMPLDYGGWNSFFTDRPAEWWDSATNEMKYILPTKKEKEV
                    DSWEESWESDLAEIPGKPLSYLWIREKSMMIQYDAKLPSELPKPSIHQIRRGVSLFPP
                    PRVATMSCNGVAKPEAMFENAPSTWVSLTAKGEDIEKYVDLRGPMPAMASLQPICPRM
                    DKYNSPLLGLGHIPAFALSDQFLFYQPEKALAEALKSLGNERDTIEHVAARLHAAML
                    HSNAKDGHVSWVLCVLSSLYWRVTGNAENAIGCLRCALHTAPPHMRDVALVSLANICH
                    QAGLFNSALISAGAALSASPRLVAIHFTLGNIYASIGDYQRALNFYYSTLSLQANFQP
                    AKDKIRTIYCHAGKEFEF"
     gene           complement(22781. .29227)
```



```
                    /gene="T19A5.4"
     CDS            complement(join(22781. .22882,22939. .23209,23256. .23350,
                    23543. .23638,23687. .23905,24180. .24311,24402. .24490,
                    24543. .24601,24648. .24725,24775. .24920,25687. .26052,
                    26111. .26206,26258. .26476,26519. .26673,26718. .27000,
                    27064. .27147,27760. .27848,28954. .29227))
                    /gene="T19A5.4"
                    /note="similar to steroid/thyroid/retinoic nuclear hormone
                    receptors"
                    /codon_start=1
                    /evidence=not_experimental
                    /db_xref="PID:g1255412"
                    /translation="MIASTSTDSLETNSPERSTKCLVYFQSAHGNHPGIDSCRACAAF
                    FRRVFVTHKQHFKCQTCQNQCVPDPRGRWNCKKCRTDRCPQLGMRPDMWWGVYGPIYW
                    DLLPDNHSLLPIKIACHEQVALTKIPCGHIVNIKVYSFLRLALPEESMGSLSSPSTSI
                    STTSSPKLISEKCLVCDQPSHGNHFGVDSCRACAAFFRRVFVTHKQQYKCINELEQCV
                    PDARGRWKCRKCRTDRCPELGMRPDNIQYNRDMFSCTGEFPKQVIYKSTRIIPSKPEC
                    ITYFDFSSLLESVENILSETTPKCVPPQCQTPLQRLAFGLKEIRKTQLWDNIPVINCI
                    GKEETFENWKVQIERTTTWLSYMDQYRDLNFDDKMTIVQCMWFPFARLERCGMTAKLR
                    KANKCQKNDYAFKSDSLVNTQTVKFDFDWLSHLSNAQMKAFLGDPSEMLCEGLTECMM
                    EIQPTDEELCYILCELCFHHLGRKLGGPMQETMERFLKILSDNLHVYIFEQDRMSRYS
                    HRLLKILKINEKYKMDSISSPSTSSSASSTPKLISEKCLVCPQPSHGNHFGVDSCRAC
                    AAFFRRVFVTHKQQFPCREGDNKCTPDEWGRWSCKRCRSDKCFALGMKPDNIQRDRDR
                    FIFSDNFRDDRKEKTSESIVPLSVERFVGKQLVTTYTSTGSGKNIEYITFFDLTPIIK
                    DADYILKRIPELKSVKVKSSLSQLAPGLQEVRKEQLPESVPELRKIGKMETWDNWVK
                    GMRRAGEWIMHFEEFRQLEQEEKMVILKCMWHLFIRLERISMTAEMRRMKLCDDKEFI
                    YGTEQRINYDTLEIDRDWFSEATDREVRSFIGPLPRWFCETIIDALMELRPSDVELSF
                    MLCNLCFHLTGQKLGGRIQEITDRLQDVLANDLHKYYLAKDKYSRYSYRLTKLLSLNR
                    QYKNDLEIRRQGIFLANTLNIFRVKLSHPEMPQFSC"
     gene           complement(29558. .30563)
                    /gene="T19A5.5"
     CDS            complement(join(29558. .29665,29715. .30080,30127. .30222,
                    30274. .30492,30543. .30563))
                    /gene="T19A5.5"
                    /codon_start=1
                    /evidence=not_experimental
                    /db_xref="PID:g1255413"
                    /translation="METILSDNRRKEVPIDNQSPLKRLSYGLKEIRKTQLGEKFQTAD
                    CIGKNESFNIWKISMERASTWLTYLDQFQSLEFDEKITITRCMWLLFRRIERCAMTAE
                    LRRTKMCKKNEYLIKTDTVVNTDSVKWDFNWLSHFSPDQLRTFVGNTSAILCESVTEC
                    MMEVQPTDEELCYILCVLCFYHIGNKIGGSMQKTTERLLNILADNLHEYYVEQKDATR
                    YSYRLTKILKISQQYKTLMEEKRRTEVLGDVFGAFRVKWSHSEIFQHSHFQ"
BASE COUNT     9558 a    5703 c    5652 g    9809 t
ORIGIN

  Query Match          13.4%;  Score 16;  DB 18;  Length 30722;
  Best Local Similarity 100.0%;  Pred. No. 1.92e+01;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  10799 CTTATTCATTTCTCAA 10814
          ||||||||||||||||
```

```
Cp     27 CTTATTCATTTCTCAA 12


RESULT   49
LOCUS       SPAC1D4     31201 bp    DNA             PLN       30-NOV-1997
DEFINITION  S.pombe chromosome I cosmid c1D4.
ACCESSION   Z69239
NID         g1177333
KEYWORDS    byr1; cct2; CDP-alcohol phosphatidyltransferase; chaperonin; csk1;
            cyclin suppressing protein kinase; DNA repair helicase; protein
            kinase; rad15; rhp3; serine threonine protein kinase; TCP1 beta
            subunit homologue; tf2 LTR.
SOURCE      fission yeast.
  ORGANISM  Schizosaccharomyces pombe
            Eukaryotae; Fungi; Ascomycota; Archiascomycetes;
            Schizosaccharomycetales; Schizosaccharomycetaceae;
            Schizosaccharomyces.
REFERENCE   1  (bases 1 to 31201)
  AUTHORS   Lye,G. and Churcher,C.M.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 31201)
  AUTHORS   Barrell,B.G., Rajandream,M.A. and Walsh,S.V.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-FEB-1995) Schizosaccharomyces pombe chromosome I
            sequencing project, Sanger Centre, Hinxton Hall, Hinxton, Cambridge
            CB10 1RQ E-mail: barrell@sanger.ac.uk
COMMENT     Notes:
            Details of yeast sequencing at the Sanger Centre are available on
            the World Wide Web.
            (URL, http://www.sanger.ac.uk/yeast/home.html)
            Protein coding regions (CDS) have been predicted with the help of
            computer analysis using the Genefinder program in PomBase (an ACEDB
            database) with additional predictions for the branch-acceptor sites
            supplied by the program Sp3splice. CAUTION: It is possible that for
            any individual CDS we may have underestimated or overestimated the
            number of introns/exons or we may not have chosen the correct
            splice donor/acceptor sites. CDS are numbered using the following
            system eg SPAC5H10.01c. SP (S. pombe), A (chromosome 1), c5H10
            (cosmid name), .01 (first CDS), c (complementary strand).
            The more significant matches with motifs in the PROSITE database
            are also included but some of these may be fortuitous. The length
            in codons is given for each CDS.
            IMPORTANT: This sequence MAY NOT be the entire insert of the
            sequenced clone.  It may be shorter because we only sequence
            overlapping sections once, or longer, because we arrange for a
            small overlap between neighbouring submissions. Cosmid c1D4
            overlapped by c1F3 5' and c22F3 3'.
FEATURES                Location/Qualifiers
     source             1. .31201
                        /organism="Schizosaccharomyces pombe"
                        /strain="972h-"
```



```
                        /db_xref="taxon:4896"
                        /chromosome="I"
                        /clone="cosmid c1D4"
                        /map="IL"
     misc_feature       1. .133
                        /gene="SPAC1D4.01"
                        /note="the sequence of this region is derived from cosmid
                        c1F3, the true end of c1D4 is at position 134, while the
                        true end of c1F3 is at position 4137 in this sequence"
     CDS                <1. .864
                        /gene="SPAC1D4.01"
                        /note="SPAC1D4.01, unknown, partial orf, len:>287"
                        /codon_start=1
                        /product="unknown"
                        /db_xref="PID:e1188804"
                        /db_xref="PID:g1177334"
                        /translation="RNMRQIFGTCTGYMNGIYFQEGLLEIPTTKSIRMNSIHIKKKSN
                        RSFRRRKVFGNEKEFDLEELDDNDIRLRQALEATKRRRKIRNSIIGINAEKLLNQETKK
                        EKQLNTANEPHEANDQTSAQSSKLIEAQLPTVEDRPAKQTNEVDINTHLLNFVEKKLK
                        QERLAQNYSENGETNALNTKNESTVQNIKNSLEHPNEHSFIRDAAALGAIREVDLGIIS
                        TDVDNLKNGRKRQKKRARMKEKLDSKALRTSEDAARDEPIEKMLKPISQDEESKGIYR
                        RFRVYKDGTQD"
     gene               1. .864
                        /gene="SPAC1D4.01"
     gene               complement(1088. .2150)
                        /gene="SPAC1D4.02"
     CDS                complement(join(1088. .1636,1686. .1873,1970. .2150))
                        /gene="SPAC1D4.02"
                        /note="SPAC1D4.02c, unknown, len: 305"
                        /codon_start=1
                        /product="unknown"
                        /db_xref="PID:e1188805"
                        /db_xref="PID:g1177335"
                        /translation="MGSYWYELIIVLSLLKCFSQNGDPSMFWALLRDSSPEVTLEVPS
                        LKGQITRWYNIKINSDEKIGMVLQWASIAPAVDAIWHILNVIDDSPVARASLVPTEDY
                        IVGTPEGMMTGEKALSDLIESHLNRPLRLYIYNHYRDSTRQVTIVPNRHWGGNGAIGC
                        GVGHGVLHRLPAPLSGPPPQPGDIVPSNPMLGGPDHKVSQPSETENFLPTPEPPKIAS
                        ANAGSSNEISIPHTQRHKKSHKGAIQDSSIQSYLDEEEKLSRELDHKTKDASSTNDSQ
                        TTPLPPPPPVAVNSTNDESAPQNEELVKN"
     misc_feature       complement(1637. .1651)
                        /gene="SPAC1D4.02"
                        /note="splice branch and acceptor sequence,
                        ctaacaatttaacag"
     misc_feature       complement(1680. .1685)
                        /gene="SPAC1D4.02"
                        /note="splice donor sequence, gtatgt"
     misc_feature       complement(1874. .1886)
                        /gene="SPAC1D4.02"
                        /note="splice branch and acceptor sequence, ctaacgactacag"
     misc_feature       complement(1964. .1969)
                        /gene="SPAC1D4.02"
```



```
                        /note="splice donor sequence, gtaagt"
     gene               complement(2574. .4201)
                        /gene="SPAC1D4.03c"
     CDS                complement(join(2574. .2820,2907. .4201))
                        /gene="SPAC1D4.03c"
                        /note="SPAC1D4.03c, unknown, len: 513"
                        /codon_start=1
                        /product="unknown"
                        /db_xref="PID:e1188806"
                        /db_xref="PID:g1177336"
                        /translation="MEPTSEHSSIKEEVENDNVHRSHESECGSLLLNPGNVLMAAPSV
                        SEDDQEVSRSTPELRSHVENVEQLLSDILHDNSSPLNVSTSVSSSSNNTAVDEIIKLL
                        SLISFDLAKQKRTYLIFSSSGKPVFSNIVDDSIEPSTVGALQAIISSFEVSKEELTSF
                        STFSNVIVVLSKNPLYLVGVSPSTTLSAAYLLSELNLLYCQILTGVTAKAMQLTLNSR
                        PNFDLRRLIGSNEQFLKELCDQLWDYELVPTLNAISPLPLRSSFRDQLSQLLLRETPK
                        SLLFTFIAIEGRLVCMVKAKKLLLHANDLYLFLSLFRTCGFHDSNEHMVPVCFPTTLN
                        PDAYIYIYSYFLCKDTVLIMGSSESGVPFEMQSVRCKVAQEIQDHGWLKKLIYCEEMD
                        RTTPRNPGSPCISHYLFYSKKYSQFTTPGYSFSTPNFNTRTLYAIYASLHDQAFHKKN
                        SFSINMTVHESLLLFTWSTASFDFHCIANATTSSQLLIANVNKILRWIRREENRLFIQ
                        TNLSF"
     misc_feature       complement(2821. .2836)
                        /gene="SPAC1D4.03c"
                        /note="splice branch and acceptor sequence,
                        ctaactatgtttttag"
     misc_feature       complement(2901. .2906)
                        /gene="SPAC1D4.03c"
                        /note="splice donor sequence, gtacgt"
     gene               5400. .7066
                        /gene="SPAC1D4.04"
     CDS                join(5400. .5402,5439. .5570,5618. .7066)
                        /gene="SPAC1D4.04"
                        /note="SPAC1D4.04, len: 527, putative cct2 gene, TCP1 beta
                        subunit homologue, similar to SW:TCPB_MOUSE P80314
                        T-complex protein 1, beta subunit (66.9% identity in 525
                        aa overlap), contains PS00995 Chaperonins TCP-1 signature
                        3"
                        /codon_start=1
                        /product="unknown"
                        /db_xref="PID:e1188807"
                        /db_xref="PID:g1177337"
                        /translation="MSLNPHQIFNESGIQERGENARLSSFVGAIAVGDLVKSTLGPKG
                        MDKILQSNSSGDIVVTNDGATILKSIALDNAAAKVLVHISKVQDDEVGDGTTSVCVFA
                        AELLRQAEIMVNAKIHPQVIIDGYRIATKTAIDALRASSIDNSSDPAKFRSDLENIAR
                        TTLSSKILSQNKNHFAQLAVDAVLRLKGSTNLDNIQIIKILGGKLDDSFLDEGFILNK
                        TIGVNCPKVMENANILIANTAMDTDKVKVFGARVRVDTTGKLAELERAEREKMKAKVE
                        KIKSHNINCFINRQLIYNWPEQLFADAGIMSIEHADFDGIERLSLVTGGEIASTFDHP
                        ELVKLGHCKKIEEIIGEDKMIKFSGVEACTCTVLRGATHQLLDESERAIHDALAV
                        LSQTVAESRVTLGGGCAEMLMAKAVEEAATHEPGKKAVAVSAFAKALSQLPTILADNA
                        GFDSSELVAQLKAAHYDGNDTNGLDMDEGEIADMRAKGILEALKLKQAVVSSGSEGAQ
                        LLLRVDTILKAAPRPRERM"
     misc_feature       5425. .5438
```



```
                        /gene="SPAC1D4.04"
                        /note="splice branch and acceptor sequence,
                        tactaacgatatag"
     misc_feature       5571. .5576
                        /gene="SPAC1D4.04"
                        /note="splice donor sequence, gtaagg"
     misc_feature       5604. .5617
                        /gene="SPAC1D4.04"
                        /note="splice branch and acceptor sequence,
                        ttaacgatttatag"
     misc_feature       5741. .5767
                        /gene="SPAC1D4.04"
                        /note="PS00995 Chaperonins TCP-1 signature 3"
     gene               complement(7282. .8497)
                        /gene="SPAC1D4.05c"
     CDS                complement(join(7282. .8366,8419. .8497))
                        /gene="SPAC1D4.05c"
                        /note="SPAC1D4.05c, unknown, len: 387, some similarity to
                        ERD1, SW:ERD1_YEAST P16151 ERD1 protein PIR:S22841 (27.2%
                        identity in 191 aa overlap)"
                        /codon_start=1
                        /product="unknown"
                        /db_xref="PID:e1188808"
                        /db_xref="PID:g1177338"
                        /translation="MDLEVVEFPLHHKLALPFRIGLLVIVGTWLWSVCYHLIYILNRY
                        QPISPNPRGSLNSKWYHLLQPLSNRHTDLEENTEFKANLVSPVDFHAGYCFAAILSI
                        SWATGFILFLKKTQGDIGGLYSHPIYPLLWVITAFILIVPFFPWRYRSSQRGLRKSII
                        RVFLFFQADFRSPYKDFIVSEIFTSYAKALGDFYIFGCVLQGHISKFTLRPDLKCDGT
                        FFVPLAMAYPFIVAILQCLHYGLSRRKHTFKINLLSALKHATALPVIYLSAIIHAKQT
                        KFTLTSGHGYLFWLILSALLSSAYTFLWDVFIDWRIRFPPHKSINHKRFPMFIYAIG
                        CFINFILRVTWSWKLHPRLHQFHEYEMGIFSFEMLEILRRFLWLFFHLDAISS"
     misc_feature       complement(8367. .8389)
                        /gene="SPAC1D4.05c"
                        /note="splice branch and acceptor sequence,
                        tactaacaatctctattctttag"
     misc_feature       complement(8413. .8418)
                        /gene="SPAC1D4.05c"
                        /note="splice donor sequence, gtatga"
     gene               complement(8990. .10082)
                        /gene="csk1"
     CDS                complement(join(8990. .9190,9303. .9518,9579. .10082))
                        /gene="csk1"
                        /note="SPAC1D4.06c, csk1 gene, len: 306, SW:CSK1_SCHPO
                        P36615 cyclin suppressing protein kinase"
                        /codon_start=1
                        /product="Csk1"
                        /db_xref="PID:e1188809"
                        /db_xref="PID:g1177339"
                        /translation="MKSVGHFVPWLTDIRHLTDGTISEVFVGERKNSKKLYVIKVQGL
                        VFKRPPHDAMRGKLILESIGHPHIERIVDSFIDNEAGSVYLITSFKSFVLSDVMDEIS
                        IDTKCKIVLQISSALEYLEKHGILHRDIHPNNILLDSHNGPAYLSDFSIAWSKQHPGE
```

...
Note: remainder of annotations omitted.

Query Match 13.4%; Score 16; DB 23; Length 31201;
Best Local Similarity 94.1%; Pred. No. 1.92e+01;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 29213 TTTTTATGGAGCTTTTC 29229
|||||| ||||||||||
Cp 115 TTTTTANGGAGCTTTTC 99

RESULT 50
LOCUS HSN120B6B 32359 bp DNA PRI 19-AUG-1998
DEFINITION Human DNA sequence from cosmid N120B6 on chromosome 22q12-qter.
ACCESSION Z73988
NID g2193880
KEYWORDS 22q12-qter.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 32359)
AUTHORS Hunt,A.
TITLE Direct Submission
JOURNAL Submitted (28-MAY-1997) E-mail enquires: humquery@sanger.ac.uk
Clone requests: clonerequest@sanger.ac.uk
COMMENT On Jun 12, 1997 this sequence version replaced gi:1403048.
IMPORTANT: This sequence is not the entire insert of clone N120B6.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.
This sequence was generated from part of bacterial clone contigs of
human chromosome 22, constructed by the Sanger Centre chromosome 22
mapping group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr22/
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The sequence from clone N120B6 has been finished in more than one
contig. This sequence (N120B6B) is separated from the preceding one
(N120B6A) by a gap of 150bp sized by restriction analysis.
The true left end of clone N4F11 is at 2151.
The true right end of clone N37D7 is at 19947.
The true right end of clone N120B6 is at 32359.
N120B6 is from the human chromosome 22-specific cosmid library
LL22NCO3, constructed at the Biomedical Sciences Division, Lawrence
Livermore National Laboratory, Livermore, CA 94550 under the

auspices of the National Laboratory Gene Library Project sponsored
by the US Department of Energy. The source of the flow sorted
chromosomes was a human/hamster hybrid containing chromosomes Y, 22
and 9.
VECTOR: Lawrist 16.
N120B6 is part of a cosmid contig isolated using YACs and markers
from the Sanger Centre chromosome 22 YAC contig described in
Collins, J.E. et al Nature 377 Suppl., 367-379.
FEATURES Location/Qualifiers
source 1. .32359
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22"
/map="22q11.2-qter"
/clone="N120B6"
/clone_lib="LL22NCO3"
repeat_region 1. .110
/note="AluSx repeat: matches 104. .1 of consensus;
incomplete repeat"
unsure 1. .130
/note="single clone"
repeat_region 528. .685
/note="L1MD2 repeat: matches 1088. .925 of consensus"
repeat_region 680. .800
/note="L1MD2 repeat: matches 884. .761 of consensus"
repeat_region 813. .1109
/note="AluSg repeat: matches 2. .299 of consensus"
repeat_region 1552. .1591
/note="20 copies of 2 mer 83 % conserved"
repeat_region 2644. .3015
/note="L1MB8 repeat: matches 299. .677 of consensus"
repeat_region 3776. .4066
/note="AluSp repeat: matches 1. .292 of consensus"
repeat_region 4726. .4886
/note="MIR repeat: matches 81. .235 of consensus"
repeat_region 5528. .5667
/note="MER4B repeat: matches 147. .9 of consensus"
repeat_region 5674. .5812
/note="MIR repeat: matches 260. .122 of consensus"
repeat_region 5682. .5728
/note="MIR2 repeat: matches 130. .84 of consensus"
repeat_region 7956. .8258
/note="AluSq repeat: matches 303. .1 of consensus"
repeat_region 8861. .8910
/note="MIR2 repeat: matches 97. .146 of consensus"
repeat_region 9247. .9388
/note="MIR repeat: matches 250. .106 of consensus"
repeat_region 10925. .11229
/note="AluY repeat: matches 2. .301 of consensus"
repeat_region 12063. .12359
/note="AluSq repeat: matches 1. .303 of consensus"

repeat_region 13536. .13585
/note="MIR repeat: matches 148. .100 of consensus"
repeat_region 13612. .13909
/note="AluSp repeat: matches 301. .3 of consensus"
repeat_region 13930. .14364
/note="MLT1D repeat: matches 505. .69 of consensus"
repeat_region 14368. .14671
/note="AluSg repeat: matches 299. .1 of consensus"
repeat_region 14672. .14753
/note="MLT1D repeat: matches 79. .1 of consensus"
repeat_region 15118. .15261
/note="MIR2 repeat: matches 4. .142 of consensus"
repeat_region 16393. .16680
/note="AluJo repeat: matches 298. .1 of consensus"
repeat_region 17432. .17608
/note="MER5B repeat: matches 1. .178 of consensus"
repeat_region 17656. .17855
/note="AluJo repeat: matches 4. .199 of consensus;
incomplete repeat"
repeat_region 17874. .18007
/note="AluJb repeat: matches 176. .302 of consensus;
incomplete repeat"
repeat_region 18682. .18983
/note="AluSq repeat: matches 1. .303 of consensus"
repeat_region 20002. .20027
/note="13 copies of 2 mer 92 % conserved"
repeat_region 21208. .21493
/note="AluSx repeat: matches 3. .292 of consensus"
repeat_region 22066. .22221
/note="AluJo repeat: matches 302. .134 of consensus;
incomplete repeat"
repeat_region 22222. .22251
/note="15 copies of 2 mer 100 % conserved"
repeat_region 22253. .22539
/note="AluSx repeat: matches 290. .1 of consensus"
repeat_region 22541. .22672
/note="AluJb repeat: matches 145. .13 of consensus;
incomplete repeat"
repeat_region 23381. .23684
/note="AluSx repeat: matches 302. .1 of consensus"
repeat_region 23783. .23935
/note="MIR repeat: matches 51. .204 of consensus"
repeat_region 23960. .24288
/note="MER1B repeat: matches 337. .1 of consensus"
repeat_region 24327. .24614
/note="AluSg repeat: matches 13. .300 of consensus"
repeat_region 24641. .24744
/note="L1MC3 repeat: matches 2382. .2486 of consensus"
repeat_region 24842. .25218
/note="MSTA repeat: matches 3. .396 of consensus"
repeat_region 25220. .25514

/note="AluSp repeat: matches 296. .2 of consensus"
repeat_region 25515. .25555
/note="MSTA repeat: matches 386. .426 of consensus"
repeat_region 26567. .26796
/note="L1PA6 repeat: matches 723. .506 of consensus"
repeat_region 26799. .26944
/note="L1PA5 repeat: matches 741. .886 of consensus"
repeat_region 26945. .26978
/note="17 copies of 2 mer 100 % conserved"
repeat_region 27138. .27303
/note="MER20 repeat: matches 2. .177 of consensus"
repeat_region 27302. .27998
/note="L1PB3 repeat: matches 184. .892 of consensus"
repeat_region 28012. .28064
/note="MER20 repeat: matches 166. .218 of consensus"
repeat_region 28615. .28814
/note="AluY repeat: matches 104. .301 of consensus;
incomplete repeat"
repeat_region 30026. .30336
/note="AluJo repeat: matches 2. .302 of consensus"
repeat_region 30367. .30539
/note="MIR repeat: matches 256. .70 of consensus"
repeat_region 31656. .31769
/note="MIR2 repeat: matches 146. .30 of consensus"
repeat_region 31773. .32072
/note="MER2 repeat: matches 344. .37 of consensus"
BASE COUNT 9270 a 7425 c 6932 g 8732 t
ORIGIN

Query Match 13.4%; Score 16; DB 27; Length 32359;
Best Local Similarity 100.0%; Pred. No. 1.92e+01;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3526 CTTTACTCTTGAGAAA 3541
||||||||||||||||
Qy 4 CTTTACTCTTGAGAAA 19

Search completed: Tue Dec 22 21:29:40 1998
Job time : 472 secs.

MPSRCH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Tue Dec 22 21:47:52 1998; MasPar time 258.44 Seconds
1147.429 Million cell updates/sec

Tabular output not generated.

Title: >US-08-530-112A-2965
Description: (1-172) from US08530112A.seq
Perfect Score: 171
N.A. Sequence: 1 GATCTGAACGATGCCGTGTC.....AAAATAAATAATAAGTGAAA 172
Comp: CTAGACTTGCTACGGCACAG.....TTTTATTTATTATTCACTTT

Scoring table: TABLE jmetric
Gap 60

Nmatch STD : Dbase 0; Query 0

Searched: 2275026 seqs, 895388244 bases x 2

Post-processing: Minimum Match 0%
Listing first 1000 summaries

Database: embl-est55
1:em_est1 2:em_gss1 3:em_gss2 4:em_gss3
Database: genbank-est107
5:gb_est1 6:gb_est10 7:gb_est11 8:gb_est12 9:gb_est13
10:gb_est14 11:gb_est15 12:gb_est16 13:gb_est17
14:gb_est18 15:gb_est19 16:gb_est2 17:gb_est20
18:gb_est21 19:gb_est3 20:gb_est4 21:gb_est5 22:gb_est6
23:gb_est7 24:gb_est8 25:gb_est9 26:gb_gss1 27:gb_gss2
28:gb_gss3 29:gb_gss4

Statistics: Mean 7.991; Variance 1.272; scale 6.285

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
c	1	171	100.0	281	17	AI004135	ot43g02.s1 Soares_test	0.00e+00
c	2	171	100.0	314	12	AA724792	ai05e02.s1 Soares para	0.00e+00
c	3	171	100.0	341	10	AA626315	zu93e06.s1 Soares test	0.00e+00
c	4	171	100.0	394	7	AA448861	zx11d01.s1 Soares tota	0.00e+00
c	5	171	100.0	420	18	AI077370	oy87f05.x1 Soares_feta	0.00e+00
c	6	171	100.0	423	20	N78952	zb18c05.s1 Homo sapien	0.00e+00
c	7	171	100.0	428	10	AA435896	zt80e12.s1 Soares test	0.00e+00
c	8	171	100.0	437	21	W68342	zd34c01.s1 Soares feta	0.00e+00
c	9	171	100.0	447	18	AI147031	qa82c01.x1 Soares_feta	0.00e+00
c	10	171	100.0	449	18	AI092949	qa81c06.x1 Soares_feta	0.00e+00
c	11	171	100.0	468	20	N78829	zb17a05.s1 Homo sapien	0.00e+00
c	12	171	100.0	549	13	AA824370	aj29a06.s1 Soares test	0.00e+00
c	13	168	98.2	379	13	AA862620	oh07c11.s1 NCI_CGAP_Ki	0.00e+00
c	14	168	98.2	407	5	T98518	ye60h05.s1 Homo sapien	0.00e+00
c	15	168	98.2	450	17	AI015595	ov09h08.x1 NCI_CGAP_Ki	0.00e+00
	16	151	88.3	470	7	AA447925	zx11d01.r1 Soares tota	0.00e+00
c	17	150	87.7	381	22	W93038	zd93b04.s1 Soares feta	0.00e+00
c	18	147	86.0	405	15	AA981266	oe31c03.s1 NCI_CGAP_Pr	0.00e+00
c	19	146	85.4	253	10	AA652731	ns65b03.s1 NCI_CGAP_Pr	0.00e+00
	20	126	73.7	481	20	W17346	zb18c05.r1 Soares feta	2.00e-258
	21	113	66.1	424	8	AA401790	zt57g10.r1 Soares test	2.87e-225
c	22	109	63.7	520	22	AA027214	ze97b09.s1 Soares feta	4.04e-215
	23	81	47.4	477	20	W17259	zb17a05.r1 Soares feta	4.94e-145
c	24	80	46.8	262	14	AA878662	oj22c11.s1 NCI_CGAP_Ki	1.44e-142
	25	63	36.8	334	5	R07662	ye98b02.r1 Homo sapien	2.91e-101
	26	53	31.0	442	21	W68383	zd34c01.r1 Soares feta	1.19e-77
c	27	47	27.5	388	18	AI143531	qc03a03.x1 Soares_feta	7.81e-64
	28	38	22.2	326	20	W25674	zb72d07.r1 Soares feta	8.59e-44
	29	36	21.1	504	11	AA070633	zm53a01.r1 Stratagene	1.75e-39
c	30	22	12.9	352	5	R07618	ye98b02.s1 Homo sapien	1.02e-11
c	31	19	11.1	441	18	AI137425	UI-R-C2p-ok-d-10-0-UI.	1.26e-06
	32	19	11.1	551	14	AA980297	ua33a08.r1 Soares mous	1.26e-06
c	33	18	10.5	291	28	AQ101466	HS_3028_B1_H04_MR CIT	4.92e-05
c	34	18	10.5	393	19	N68090	za10e03.s1 Homo sapien	4.92e-05
c	35	18	10.5	440	29	AQ246068	HS_2057_B2_A10_T7 CIT	4.92e-05
	36	18	10.5	475	6	C06551	similar to hypothetica	4.92e-05
	37	18	10.5	572	7	AA532018	TgESTzz46g01.r1 TgME49	4.92e-05
	38	18	10.5	572	7	AA531880	TgESTzz32e10.r1 TgME49	4.92e-05
c	39	18	10.5	604	26	B72809	RPCI11-10F20.TP RPCI11	4.92e-05
	40	18	10.5	680	26	AQ017441	CIT-HSP-2306F1.TF CIT-	4.92e-05
	41	18	10.5	1148	26	B13244	T18C16-T7 TAMU Arabido	4.92e-05
	42	17	9.9	225	23	AA259656	va36a09.r1 Soares mous	1.65e-03

	43	17	9.9	273	23	AA098029	mn82b06.r1 Stratagene	1.65e-03
c	44	17	9.9	301	5	T36476	EST101417 Saccharomyce	1.65e-03
c	45	17	9.9	316	8	AA476276	zw29a11.s1 Soares ovar	1.65e-03
c	46	17	9.9	318	14	AA686340	EST109943 Rat PC-12 ce	1.65e-03
	47	17	9.9	329	24	AA270724	va67e04.r1 Soares mous	1.65e-03
c	48	17	9.9	360	9	C46755	c.elegans cDNA clone y	1.65e-03
c	49	17	9.9	373	5	R11080	yf39b03.r1 Homo sapien	1.65e-03
	50	17	9.9	393	29	AQ238631	RPCI11-71A5.TK RPCI11	1.65e-03
	51	17	9.9	394	11	AA069104	zm10f06.r1 Stratagene	1.65e-03
c	52	17	9.9	405	7	AA456808	zw28e11.r1 Soares ovar	1.65e-03
c	53	17	9.9	406	7	AA454913	zx79d08.r1 Soares ovar	1.65e-03
	54	17	9.9	412	24	AA324452	EST27277 Cerebellum II	1.65e-03
c	55	17	9.9	427	16	R79391	yi83f11.r1 Homo sapien	1.65e-03
c	56	17	9.9	432	20	N91229	za13h02.r1 Soares feta	1.65e-03
c	57	17	9.9	441	19	N68672	za21f01.s1 Homo sapien	1.65e-03
c	58	17	9.9	446	28	AQ154601	HS_3017_B2_B12_T7 CIT	1.65e-03
	59	17	9.9	461	9	C72464	Rice cDNA, partial seq	1.65e-03
c	60	17	9.9	462	6	AA410942	st32h02.r1 Soares ovar	1.65e-03
	61	17	9.9	480	19	H65691	yr62c09.r1 Homo sapien	1.65e-03
	62	17	9.9	577	26	B72264	RPCI11-7O24.TP RPCI11	1.65e-03
c	63	17	9.9	588	17	AU018183	Mus musculus 2-cell em	1.65e-03
	64	17	9.9	668	11	C83894	Dictyostelium discoide	1.65e-03
c	65	17	9.9	695	17	AI110130	GH09604.5prime GH Dros	1.65e-03
c	66	17	9.9	739	28	AQ200209	RPCI11-4503.TJ RPCI11	1.65e-03
	67	17	9.9	813	26	B18861	F20C18-Sp6 IGF Arabido	1.65e-03
	68	16	9.4	112	14	AA895312	vy33a06.r1 Stratagene	4.72e-02
c	69	16	9.4	118	26	AQ015520	CIT-HSP-2304L12.TR CIT	4.72e-02
c	70	16	9.4	189	29	AQ242069	CITBI-E1-2501I18.TR.1	4.72e-02
	71	16	9.4	213	22	AA208683	T3785 MVAT4 bloodstrea	4.72e-02
	72	16	9.4	217	10	AA646437	vs37e02.r1 Stratagene	4.72e-02
c	73	16	9.4	218	7	AA466690	vd95e03.r1 Soares mous	4.72e-02
c	74	16	9.4	219	23	AA186065	mt97h01.r1 Soares mous	4.72e-02
c	75	16	9.4	222	10	AA583683	nn58d10.s1 NCI_CGAP_Ki	4.72e-02
	76	16	9.4	223	27	AQ059254	CIT-HSP-2349D16.TF CIT	4.72e-02
	77	16	9.4	226	26	AQ010962	F23I1TFB IGF Arabidops	4.72e-02
	78	16	9.4	245	27	AQ063427	CIT-HSP-2346H14.TF CIT	4.72e-02
	79	16	9.4	247	23	AA087830	mn94c02.r1 Stratagene	4.72e-02
	80	16	9.4	249	15	AI026540	TENU0752 T. cruzi epim	4.72e-02
	81	16	9.4	263	20	N79012	zb42f10.s1 Homo sapien	4.72e-02
	82	16	9.4	267	5	T02473	0003C7 Plasmodium falc	4.72e-02
c	83	16	9.4	271	14	AA857557	of64d09.s1 NCI_CGAP_Co	4.72e-02
	84	16	9.4	272	5	T69428	yc38e01.s1 Homo sapien	4.72e-02
	85	16	9.4	283	7	AA452118	zx15c04.s1 Soares tota	4.72e-02
c	86	16	9.4	283	24	AA300160	EST13146 Thymus tumor	4.72e-02
	87	16	9.4	287	6	AA406333	zv11h12.r1 Soares NhHM	4.72e-02
	88	16	9.4	289	27	AQ063112	CIT-HSP-2355C22.TR CIT	4.72e-02
	89	16	9.4	291	24	AA321857	EST24595 Cerebellum II	4.72e-02
	90	16	9.4	293	25	AA358747	EST57684 Fetal lung II	4.72e-02
	91	16	9.4	294	22	N97506	0610M7 gmbPfHB3.1, G.	4.72e-02
	92	16	9.4	298	14	C90598	Dictyostelium discoide	4.72e-02
c	93	16	9.4	302	8	AA256208	zr80a08.s1 Soares NhHM	4.72e-02

c	94	16	9.4	308	12	AA765239	nz78g03.s1 NCI_CGAP_GC	4.72e-02
c	95	16	9.4	311	13	AA179342	zp48f02.s1 Stratagene	4.72e-02
	96	16	9.4	312	22	N97683	1008C3 czapPFDd2.1, De	4.72e-02
	97	16	9.4	314	27	B66956	CIT-HSP-2017A24.TRB CI	4.72e-02
	98	16	9.4	325	5	T69682	yc44g01.r1 Homo sapien	4.72e-02
c	99	16	9.4	328	19	H70904	yu70a02.s1 Homo sapien	4.72e-02
c	100	16	9.4	332	18	AI015688	oy28h07.s1 Soares_sene	4.72e-02
c	101	16	9.4	332	17	AI072419	UI-R-C2-ne-h-11-0-UI.s	4.72e-02
c	102	16	9.4	334	24	AA290204	vb34g06.r1 Soares mous	4.72e-02
c	103	16	9.4	337	12	AA814275	nz06f01.s1 NCI_CGAP_GC	4.72e-02
	104	16	9.4	350	12	AA743668	ny92a09.s1 NCI_CGAP_GC	4.72e-02
c	105	16	9.4	354	25	AA376795	EST89260 Small intesti	4.72e-02
	106	16	9.4	360	6	CELK026D5F	C.elegans cDNA clone y	4.72e-02
c	107	16	9.4	360	17	AI087898	oz95b09.x1 Soares_para	4.72e-02
	108	16	9.4	362	17	AI022092	ow56e03.x1 Soares_para	4.72e-02
c	109	16	9.4	362	23	AA164105	ms18a05.r1 Stratagene	4.72e-02
c	110	16	9.4	364	19	H70911	yu70b01.s1 Homo sapien	4.72e-02
	111	16	9.4	367	28	AQ120822	HS_3075_A1_D06_MR CIT	4.72e-02
c	112	16	9.4	367	13	AA809100	nv32f09.r1 NCI_CGAP_Br	4.72e-02
	113	16	9.4	368	13	AA813921	ny95f10.s1 NCI_CGAP_GC	4.72e-02
	114	16	9.4	371	13	AA805369	ob44a02.s1 NCI_CGAP_GC	4.72e-02
c	115	16	9.4	377	27	B87977	RPCI11-14K12.TVB RPCI1	4.72e-02
c	116	16	9.4	378	6	C06752	similar to none, mRNA	4.72e-02
	117	16	9.4	384	10	AA618351	nq18b03.s1 NCI_CGAP_Th	4.72e-02
	118	16	9.4	385	14	AA848026	od26g01.s1 NCI_CGAP_GC	4.72e-02
	119	16	9.4	386	28	AQ133799	HS_3047_A2_D08_MR CIT	4.72e-02
	120	16	9.4	393	28	AQ147876	HS_3059_B1_F06_MR CIT	4.72e-02
c	121	16	9.4	394	5	R42500	yg02h08.s1 Homo sapien	4.72e-02
	122	16	9.4	396	14	C93412	Dictyostelium discoide	4.72e-02
c	123	16	9.4	396	23	AA144098	mq66c03.r1 Soares 2NbM	4.72e-02
c	124	16	9.4	396	15	AA936064	oo65b01.s1 NCI_CGAP_GC	4.72e-02
	125	16	9.4	397	21	R85840	yq26h11.s1 Soares feta	4.72e-02
c	126	16	9.4	397	10	AA638427	vo53a06.r1 Barstead mo	4.72e-02
	127	16	9.4	399	13	AA844329	ai95e11.s1 Soares para	4.72e-02
	128	16	9.4	400	10	AA605039	no68h12.s1 NCI_CGAP_AA	4.72e-02
c	129	16	9.4	401	28	AQ214239	HS_3013_B2_A05_MR CIT	4.72e-02
c	130	16	9.4	401	16	H03252	yj44b12.s1 Homo sapien	4.72e-02
c	131	16	9.4	402	15	AI020474	ub27g02.r1 Soares 2NbM	4.72e-02
	132	16	9.4	406	11	AA699707	zi40e12.s1 Soares feta	4.72e-02
c	133	16	9.4	409	28	AQ127352	HS_3064_A2_B02_MF CIT	4.72e-02
	134	16	9.4	410	15	AI033628	ow14c04.x1 Soares_para	4.72e-02
c	135	16	9.4	410	15	AI043442	TENU1001 T. cruzi epim	4.72e-02
	136	16	9.4	415	10	AA425109	zw07c01.s1 Soares NhHM	4.72e-02
c	137	16	9.4	415	13	AA199826	zq76g06.s1 Stratagene	4.72e-02
c	138	16	9.4	416	11	T43548	6811 Lambda-PRL2 Arabi	4.72e-02
c	139	16	9.4	416	26	AQ021404	CIT-HSP-2311L19.TR CIT	4.72e-02
c	140	16	9.4	417	14	C91200	Dictyostelium discoide	4.72e-02
c	141	16	9.4	422	13	AA180056	zp40g10.s1 Stratagene	4.72e-02
c	142	16	9.4	425	5	T82118	yd30h03.r1 Homo sapien	4.72e-02
c	143	16	9.4	425	13	AA806467	oc27e09.s1 NCI_CGAP_GC	4.72e-02
	144	16	9.4	427	14	AA837255	od26a08.s1 NCI_CGAP_GC	4.72e-02

	145	16	9.4	427	5	T84639	yd53g05.r1 Homo sapien	4.72e-02
c	146	16	9.4	431	8	AA435314	vd49e02.s1 Knowles Sol	4.72e-02
	147	16	9.4	432	14	AA964040	UI-R-C0-gu-g-05-0-UI.s	4.72e-02
c	148	16	9.4	433	28	AQ183297	HS_3141_B1_G01_MR CIT	4.72e-02
c	149	16	9.4	434	28	AQ184889	HS_2216_B2_F12_MR CIT	4.72e-02
c	150	16	9.4	434	23	AA172732	mt06c09.r1 Soares mous	4.72e-02
c	151	16	9.4	436	11	AA680142	zi39e01.s1 Soares feta	4.72e-02
c	152	16	9.4	437	28	AQ168994	HS_3179_B2_D12_MR CIT	4.72e-02
c	153	16	9.4	438	21	AA052583	mc76e02.r1 Soares mous	4.72e-02
c	154	16	9.4	438	6	AA388807	vb25d10.r1 Soares mous	4.72e-02
	155	16	9.4	439	17	AI027458	ow06b12.x1 Soares_para	4.72e-02
c	156	16	9.4	442	11	AA071045	zm58d02.s1 Stratagene	4.72e-02
	157	16	9.4	447	17	AI127917	qc46g08.x1 Soares_preg	4.72e-02
c	158	16	9.4	447	16	H14516	ym18h09.s1 Homo sapien	4.72e-02
c	159	16	9.4	449	9	AA585721	28370 Lambda-PRL2 Arab	4.72e-02
c	160	16	9.4	449	27	AQ038317	CIT-HSP-2336C5.TF CIT-	4.72e-02
	161	16	9.4	451	23	AA154891	mq61g08.r1 Soares 2NbM	4.72e-02
c	162	16	9.4	453	29	AQ226939	HS_2016_B1_F11_MR CIT	4.72e-02
	163	16	9.4	455	20	N94842	yy62f05.r1 Homo sapien	4.72e-02
	164	16	9.4	455	10	AA424856	zw03c01.s1 Soares NhHM	4.72e-02
	165	16	9.4	458	26	B38465	HS-1047-B2-B02-MF.abi	4.72e-02
	166	16	9.4	458	5	T69605	yc44f08.s1 Homo sapien	4.72e-02
c	167	16	9.4	458	17	AI126447	qd78c11.x1 Soares_test	4.72e-02
	168	16	9.4	461	18	AI143906	qd99e08.x1 Soares_test	4.72e-02
c	169	16	9.4	461	7	AA511327	vj22d08.r1 Soares mous	4.72e-02
	170	16	9.4	462	5	T67996	yc28g01.s1 Homo sapien	4.72e-02
	171	16	9.4	462	15	AI060510	ub35h05.r1 Soares 2NbM	4.72e-02
c	172	16	9.4	462	22	AA104073	mo49a10.r1 Life Tech m	4.72e-02
c	173	16	9.4	463	15	AI011287	EST205738 Normalized r	4.72e-02
	174	16	9.4	466	17	AI095884	qb30h04.x1 Soares_preg	4.72e-02
c	175	16	9.4	467	12	AA742262	ny80d03.s1 NCI_CGAP_GC	4.72e-02
c	176	16	9.4	476	10	T13079	1145 Lambda-PRL2 Arabi	4.72e-02
c	177	16	9.4	477	18	AI148990	qc71c06.x1 Soares_plac	4.72e-02
c	178	16	9.4	477	14	AA833780	od61e09.s1 NCI_CGAP_GC	4.72e-02
c	179	16	9.4	488	21	AA049450	mj35c02.r1 Soares mous	4.72e-02
	180	16	9.4	502	19	N36081	yy32a06.s1 Homo sapien	4.72e-02
	181	16	9.4	503	5	T61857	yb92g12.s1 Homo sapien	4.72e-02
	182	16	9.4	504	28	AQ215931	HS_3243_B2_D02_T7 CIT	4.72e-02
c	183	16	9.4	505	28	AQ169066	HS_3171_B1_E02_T7 CIT	4.72e-02
c	184	16	9.4	508	10	AA425532	zw48d02.r1 Soares tota	4.72e-02
	185	16	9.4	514	28	AQ213047	HS_3243_A2_D08_T7 CIT	4.72e-02
	186	16	9.4	521	29	AQ227509	HS_2015_B2_H08_MR CIT	4.72e-02
c	187	16	9.4	523	12	AA773163	ac22f10.s1 Stratagene	4.72e-02
c	188	16	9.4	527	27	B59705	CIT-HSP-343P21.TPB CIT	4.72e-02
c	189	16	9.4	528	26	FR0004620	F.rubripes GSS sequenc	4.72e-02
	190	16	9.4	529	27	AQ050254	nbxb0003cG07r CUGI Ric	4.72e-02
	191	16	9.4	537	8	AA563137	vl74d10.r1 Knowles Sol	4.72e-02
c	192	16	9.4	537	8	AA546813	vk55g07.r1 Stratagene	4.72e-02
c	193	16	9.4	541	16	H19144	yn50a02.r1 Homo sapien	4.72e-02
c	194	16	9.4	544	18	AI148890	qc70d12.x1 Soares_plac	4.72e-02
c	195	16	9.4	544	18	AI103635	EST212924 Normalized r	4.72e-02

	196	16	9.4	561	13	C86523	Mus musculus fertilize	4.72e-02
	197	16	9.4	561	9	AA589511	vl48d03.s1 Stratagene	4.72e-02
c	198	16	9.4	566	23	AA198959	mu16a08.r1 Soares 2NbM	4.72e-02
c	199	16	9.4	567	10	AA659946	JM00M235.QM3 Miracidia	4.72e-02
	200	16	9.4	572	17	A00021471	Mus musculus 8-cell em	4.72e-02
	201	16	9.4	573	15	C81156	Mus musculus 3.5-dpc b	4.72e-02
c	202	16	9.4	574	21	W46364	sc30h06.s1 Soares sene	4.72e-02
c	203	16	9.4	582	26	B63603	RPCI11-2F19.TVB RPCI11	4.72e-02
	204	16	9.4	587	15	C81131	Mus musculus 3.5-dpc b	4.72e-02
	205	16	9.4	587	8	AA550222	1341m3 gmbPfHB3.1, G.	4.72e-02
c	206	16	9.4	588	28	AQ157189	nbxb0009G09f CUGI Rice	4.72e-02
	207	16	9.4	601	15	C78830	Mus musculus 3.5-dpc b	4.72e-02
c	208	16	9.4	602	26	FR0013679	F.rubripes GSS sequenc	4.72e-02
	209	16	9.4	614	27	AQ034241	CpG0537B CpIOWAgDNA1 C	4.72e-02
	210	16	9.4	619	26	FR0004639	F.rubripes GSS sequenc	4.72e-02
c	211	16	9.4	619	26	FR0004634	F.rubripes GSS sequenc	4.72e-02
c	212	16	9.4	621	14	AA888021	ng97g01.s1 NCI_CGAP_Co	4.72e-02
c	213	16	9.4	636	18	AI179439	EST223149 Normalized r	4.72e-02
	214	16	9.4	649	14	C90168	Dictyostelium discoide	4.72e-02
	215	16	9.4	652	8	AA549987	1066m3 gmbPfHB3.1, G.	4.72e-02
	216	16	9.4	664	14	C91022	Dictyostelium discoide	4.72e-02
	217	16	9.4	679	19	H99019	yx11c12.s1 Homo sapien	4.72e-02
c	218	16	9.4	693	27	AQ049777	nbxb0001aB08f CUGI Ric	4.72e-02
c	219	16	9.4	695	15	C94371	Dictyostelium discoide	4.72e-02
c	220	16	9.4	704	28	AQ156961	nbxb0008P21r CUGI Rice	4.72e-02
c	221	16	9.4	718	20	W25929	15b10 Human retina cDN	4.72e-02
	222	16	9.4	720	14	C90799	Dictyostelium discoide	4.72e-02
c	223	16	9.4	838	28	AQ163518	nbxb0007M03r CUGI Rice	4.72e-02
c	224	16	9.4	865	28	AF011025	Homo sapiens chromosom	4.72e-02
	225	16	9.4	963	26	B08101	T25C22-Sp6 TAMU Arabid	4.72e-02
c	226	16	9.4	964	26	B09240	T17M24-Sp6 TAMU Arabid	4.72e-02
c	227	16	9.4	1001	10	W13807	ma95d09.r1 Soares mous	4.72e-02
c	228	15	8.8	52	9	AA579703	nm80e07.s1 NCI_CGAP_Co	1.12e+00
	229	15	8.8	68	14	AA982545	uh09f12.r1 Soares mous	1.12e+00
c	230	15	8.8	80	14	AA937417	oj09a01.s1 NCI_CGAP_Me	1.12e+00
c	231	15	8.8	109	19	N30014	yx82f08.s1 Homo sapien	1.12e+00
	232	15	8.8	115	26	B02679	cSRL-158G1-u cSRL flow	1.12e+00
c	233	15	8.8	138	15	AA943920	EST199419 Normalized r	1.12e+00
c	234	15	8.8	138	8	AA503103	ne44h05.s1 NCI_CGAP_Co	1.12e+00
	235	15	8.8	154	11	AA700455	zj74d03.s1 Soares feta	1.12e+00
	236	15	8.8	160	10	AA623573	vn73b11.r1 Knowles Sol	1.12e+00
	237	15	8.8	161	29	AQ251329	T31B11-T7 TAMU Arabido	1.12e+00
c	238	15	8.8	171	21	R94740	yt71a06.s1 Soares feta	1.12e+00
	239	15	8.8	175	9	AA607177	vm96g02.r1 Knowles Sol	1.12e+00
c	240	15	8.8	179	27	B89328	RPCI11-24M18.TVB RPCI1	1.12e+00
	241	15	8.8	181	5	T09992	0546m7 Plasmodium falc	1.12e+00
	242	15	8.8	187	9	AA605932	fa25g08.s1 Ekkerpost s	1.12e+00
	243	15	8.8	195	20	N83770	KK2758F Homo sapiens c	1.12e+00
c	244	15	8.8	202	13	AA689558	nx06f05.s1 NCI_CGAP_GC	1.12e+00
c	245	15	8.8	204	15	AA886421	oj52a10.s1 NCI_CGAP_Ki	1.12e+00
c	246	15	8.8	206	13	AA620471	ae58f01.s1 Stratagene	1.12e+00

c	247	15	8.8	207	17	AI105968	bb01g02_p1 ZF adult he	1.12e+00
c	248	15	8.8	208	16	H21817	yl31d08.s1 Homo sapien	1.12e+00
	249	15	8.8	209	19	HSU25889	Human mRNA clone hsf14	1.12e+00
c	250	15	8.8	209	15	AI029589	UI-R-C0-ix-h-08-0-UI.s	1.12e+00
c	251	15	8.8	210	19	H81552	yu65f01.s1 Homo sapien	1.12e+00
c	252	15	8.8	211	27	B81983	RPCI11-19N8.TV RPCI11	1.12e+00
	253	15	8.8	213	19	H67233	yu65f01.r1 Homo sapien	1.12e+00
	254	15	8.8	215	5	T11790	A1151F Homo sapiens cD	1.12e+00
c	255	15	8.8	216	13	AA876224	nx25g09.s1 NCI_CGAP_GC	1.12e+00
	256	15	8.8	220	6	AA386351	EST32580 Embryo, 12 we	1.12e+00
	257	15	8.8	230	20	N83520	KK0764F Homo sapiens c	1.12e+00
	258	15	8.8	233	5	T18571	hbc2102 Homo sapiens c	1.12e+00
c	259	15	8.8	238	10	AA603872	np15g02.s1 NCI_CGAP_Pr	1.12e+00
	260	15	8.8	239	7	RICS10574A	Rice cDNA, partial seq	1.12e+00
	261	15	8.8	241	19	N47177	yy73e05.r1 Homo sapien	1.12e+00
	262	15	8.8	243	12	AA750633	ISGS0213 Rice Immature	1.12e+00
	263	15	8.8	244	29	AQ229758	HS_2033_A2_A08_T7 CIT	1.12e+00
c	264	15	8.8	244	9	AA554643	nl06b02.s1 NCI_CGAP_Pr	1.12e+00
	265	15	8.8	246	6	C02045	HUMGS0005362, Human Ge	1.12e+00
c	266	15	8.8	247	8	AA225172	nc22d10.s1 NCI_CGAP_Pr	1.12e+00
	267	15	8.8	250	6	AA386339	EST18395 Liver I Homo	1.12e+00
c	268	15	8.8	251	6	AA386338	EST18394 Liver I Homo	1.12e+00
c	269	15	8.8	252	5	T11791	A1151R Homo sapiens cD	1.12e+00
c	270	15	8.8	252	15	C94106	Dictyostelium discoide	1.12e+00
c	271	15	8.8	255	5	T12451	B001R Homo sapiens cDN	1.12e+00
	272	15	8.8	256	9	AA605653	fa25h05.s1 Ekkerpost s	1.12e+00
	273	15	8.8	262	13	AA811756	ob80f05.s1 NCI_CGAP_GC	1.12e+00
	274	15	8.8	263	20	HSC0AB072	H. sapiens partial cDN	1.12e+00
c	275	15	8.8	269	25	AA360978	EST70206 T-cell lympho	1.12e+00
	276	15	8.8	272	20	HSC3DH052	H. sapiens partial cDN	1.12e+00
	277	15	8.8	278	7	HUM302F03B	Human aorta cDNA 5'-en	1.12e+00
c	278	15	8.8	279	14	AA963013	UI-R-E1-gd-f-01-0-UI.s	1.12e+00
c	279	15	8.8	279	16	H44580	yp04d06.s1 Homo sapien	1.12e+00
c	280	15	8.8	280	10	AA588773	nn18f05.s1 NCI_CGAP_Co	1.12e+00
c	281	15	8.8	283	7	HUMGS01187	Human HL60 3'directed	1.12e+00
c	282	15	8.8	286	13	AA882750	TENS0325 T. cruzi epim	1.12e+00
c	283	15	8.8	286	8	AA546163	vk62d09.r1 Stratagene	1.12e+00
c	284	15	8.8	287	26	AQ011961	HS_2190_A1_MR_E08 CIT	1.12e+00
	285	15	8.8	289	6	AA386347	EST20781 Spleen I Homo	1.12e+00
c	286	15	8.8	289	19	N44080	yy32a01.r1 Homo sapien	1.12e+00
	287	15	8.8	291	6	C18687	Human placenta cDNA 5'	1.12e+00
c	288	15	8.8	291	15	AA996584	UI-R-C0-hd-e-08-0-UI.s	1.12e+00
	289	15	8.8	292	16	H01447	yj33a01.s1 Homo sapien	1.12e+00
c	290	15	8.8	292	12	AA757112	ah53g11.s1 Soares test	1.12e+00
c	291	15	8.8	293	6	AA386331	EST14954 Aorta endothe	1.12e+00
	292	15	8.8	295	5	T12450	B001F Homo sapiens cDN	1.12e+00
c	293	15	8.8	295	18	AI178084	EST221743 Normalized r	1.12e+00
	294	15	8.8	299	24	AA321720	EST24249 Corpus callos	1.12e+00
	295	15	8.8	300	9	C33647	C.elegans cDNA clone y	1.12e+00
	296	15	8.8	300	9	C55184	C.elegans cDNA clone y	1.12e+00
	297	15	8.8	300	9	C52900	C.elegans cDNA clone y	1.12e+00

	298	15	8.8	302	20	R58766	G5035 Fetal heart Homo	1.12e+00
	299	15	8.8	303	10	HUMNK480	Human keratinocyte cDN	1.12e+00
c	300	15	8.8	305	11	AA670248	af25a09.s1 Soares tota	1.12e+00
c	301	15	8.8	307	14	AA935430	ny39g10.s1 NCI_CGAP_Pr	1.12e+00
	302	15	8.8	308	5	T18204	0814c3 Plasmodium falc	1.12e+00
	303	15	8.8	309	14	AA837254	od26a07.s1 NCI_CGAP_GC	1.12e+00
c	304	15	8.8	309	9	AA564639	nj25e10.s1 NCI_CGAP_AA	1.12e+00
	305	15	8.8	310	5	T94541	ye36g08.r1 Homo sapien	1.12e+00
c	306	15	8.8	310	13	AA630588	ac11b10.s1 Stratagene	1.12e+00
c	307	15	8.8	310	22	AA180170	zp48h02.s1 Stratagene	1.12e+00
	308	15	8.8	311	5	T77965	yc97c02.r1 Homo sapien	1.12e+00
c	309	15	8.8	311	17	AA938443	oo98d04.s1 NCI_CGAP_Ki	1.12e+00
	310	15	8.8	313	26	B17358	345P20.TVB CIT9708KA1	1.12e+00
	311	15	8.8	315	7	HUMGS01239	Human HL60 3'directed	1.12e+00
c	312	15	8.8	316	22	AA151422	zo38f12.s1 Stratagene	1.12e+00
c	313	15	8.8	318	13	AA599399	ag23a06.s1 Jia bone ma	1.12e+00
c	314	15	8.8	318	27	AQ065856	HS_2239_A1_D06_MR CIT	1.12e+00
c	315	15	8.8	319	14	AA835855	oc76f07.s1 NCI_CGAP_GC	1.12e+00
c	316	15	8.8	320	12	AA731674	nw58e01.s1 NCI_CGAP_GC	1.12e+00
c	317	15	8.8	321	10	AA650149	na90h01.s1 NCI_CGAP_Pr	1.12e+00
c	318	15	8.8	322	23	AA218518	MAAD0204.M3R Schistoso	1.12e+00
c	319	15	8.8	322	11	AA707363	ah39b04.s1 Soares test	1.12e+00
c	320	15	8.8	322	8	AA492153	ng61g08.s1 NCI_CGAP_Li	1.12e+00
c	321	15	8.8	326	10	AA654129	nt10b12.s1 NCI_CGAP_Pr	1.12e+00
	322	15	8.8	328	6	C06035	similar to u-PA recept	1.12e+00
c	323	15	8.8	328	6	AA396738	vb29f01.r1 Soares mous	1.12e+00
	324	15	8.8	329	8	AA262010	zs18c02.r1 NCI_CGAP_GC	1.12e+00
c	325	15	8.8	330	5	R27004	yh45e04.s1 Homo sapien	1.12e+00
c	326	15	8.8	330	26	B34008	HS-1023-B2-C11-MF.abi	1.12e+00
c	327	15	8.8	332	26	B96097	F23N9TRB IGF Arabidops	1.12e+00
c	328	15	8.8	332	19	N74095	za75b03.s1 Homo sapien	1.12e+00
c	329	15	8.8	333	5	T47592	yb15f09.s1 Homo sapien	1.12e+00
	330	15	8.8	334	7	AA448254	zw83e07.s1 Soares test	1.12e+00
c	331	15	8.8	334	7	RICS3544A	Rice cDNA, partial seq	1.12e+00
c	332	15	8.8	334	25	AA349064	EST55623 Infant adrena	1.12e+00
c	333	15	8.8	334	25	AA362780	EST72625 Ovary II Homo	1.12e+00
	334	15	8.8	336	22	AA095731	15720.seq.F Fetal hear	1.12e+00
c	335	15	8.8	338	12	AA772581	ai45e12.s1 Soares para	1.12e+00
c	336	15	8.8	338	14	AA938478	oo56h07.s1 NCI_CGAP_Lu	1.12e+00
c	337	15	8.8	341	10	AA627796	nq50h04.s1 NCI_CGAP_Co	1.12e+00
c	338	15	8.8	344	14	AA862945	oq98c11.s1 NCI_CGAP_Ki	1.12e+00
c	339	15	8.8	345	29	AQ241630	CITB1-E1-2501E3.TF.1 C	1.12e+00
c	340	15	8.8	347	12	AA796546	vp14f09.r1 Soares mous	1.12e+00
	341	15	8.8	348	6	HSPD06033	H.sapiens EST sequence	1.12e+00
c	342	15	8.8	349	18	AI079721	oy55c08.s1 NCI_CGAP_Br	1.12e+00
	343	15	8.8	350	20	ATTS2287	A. thaliana transcribe	1.12e+00
c	344	15	8.8	350	18	AI168806	ox67b05.s1 Soares_NhHM	1.12e+00
c	345	15	8.8	351	18	AI079664	oy42c04.s1 Soares_para	1.12e+00
c	346	15	8.8	351	20	ATTS4292	A. thaliana transcribe	1.12e+00
c	347	15	8.8	351	18	AI143380	qa51d03.s1 Soares_NhHM	1.12e+00
c	348	15	8.8	352	13	AA840246	vw89c02.r1 Stratagene	1.12e+00

c	349	15	8.8	352	11	AA064620	zm13c03.s1 Stratagene	1.12e+00
c	350	15	8.8	353	20	W95726	ze07d01.s1 Soares feta	1.12e+00
	351	15	8.8	354	21	R84498	yq24b07.r1 Soares feta	1.12e+00
	352	15	8.8	355	8	AA556104	TENF0403 T.cruzi epima	1.12e+00
	353	15	8.8	357	29	AQ137672	HS_3065_B2_A12_MR CIT	1.12e+00
	354	15	8.8	360	9	C41592	C.elegans cDNA clone y	1.12e+00
	355	15	8.8	360	6	CELK044EXF	C.elegans cDNA clone y	1.12e+00
	356	15	8.8	361	19	H87034	yw15h09.r1 Homo sapien	1.12e+00
	357	15	8.8	362	16	R98906	yr31c03.r1 Homo sapien	1.12e+00
c	358	15	8.8	364	16	T28782	EST56165 Homo sapiens	1.12e+00
	359	15	8.8	366	12	AA773546	ab57h08.s1 Stratagene	1.12e+00
c	360	15	8.8	366	14	AA899361	UI-R-E0-c1-g-02-0-UI.s	1.12e+00
c	361	15	8.8	368	28	AQ212811	HS_3118_B2_E11_MR CIT	1.12e+00
	362	15	8.8	369	12	AA768165	oa61g04.s1 NCI_CGAP_GC	1.12e+00
c	363	15	8.8	369	12	AA778366	zl56g06.s1 Soares preg	1.12e+00
c	364	15	8.8	370	18	AI079161	oy44e04.s1 Soares_para	1.12e+00
	365	15	8.8	371	19	H94494	yv19f06.s1 Homo sapien	1.12e+00
c	366	15	8.8	371	8	AA228759	nc14f10.s1 NCI_CGAP_Pr	1.12e+00
	367	15	8.8	372	11	AA719315	zh35d01.s1 Soares pine	1.12e+00
c	368	15	8.8	372	9	W66293	TgESTzy72f10.r1 TgME49	1.12e+00
	369	15	8.8	374	22	W35324	zc71b09.r1 Soares feta	1.12e+00
	370	15	8.8	375	5	T92729	ye23d05.s1 Homo sapien	1.12e+00
c	371	15	8.8	375	27	AQ024541	HS_2182_A2_D12_MF CIT	1.12e+00
c	372	15	8.8	375	12	AA758070	ah61c08.s1 Soares test	1.12e+00
	373	15	8.8	376	17	AA886966	oi14f10.s1 NCI_CGAP_GC	1.12e+00
c	374	15	8.8	376	11	AA076646	zm20c06.s1 Stratagene	1.12e+00
c	375	15	8.8	376	9	C70992	C.elegans cDNA clone y	1.12e+00
	376	15	8.8	377	12	AA765722	oa09f12.s1 NCI_CGAP_GC	1.12e+00
c	377	15	8.8	377	21	W35368	zc07g05.s1 Soares para	1.12e+00
	378	15	8.8	378	14	C88911	Mus musculus early bla	1.12e+00
c	379	15	8.8	378	5	T94453	ye36g08.s1 Homo sapien	1.12e+00
	380	15	8.8	381	6	C12945	C.elegans cDNA clone y	1.12e+00
	381	15	8.8	381	9	AA535674	nj77h10.s1 NCI_CGAP_Pr	1.12e+00
c	382	15	8.8	381	5	R06957	yf12g01.r1 Homo sapien	1.12e+00
	383	15	8.8	382	15	AA891543	EST195346 Normalized r	1.12e+00
c	384	15	8.8	382	16	R62362	yi20c11.s1 Homo sapien	1.12e+00
c	385	15	8.8	382	28	AQ134545	HS_3058_A1_A11_MR CIT	1.12e+00
	386	15	8.8	383	6	AA406934	MBAPC27A06T3 Brugia ma	1.12e+00
	387	15	8.8	383	19	H87652	yw15a06.r1 Homo sapien	1.12e+00
c	388	15	8.8	383	14	AA958809	vw63g01.s1 Soares mous	1.12e+00
c	389	15	8.8	383	10	AA625552	zu90e02.s1 Soares test	1.12e+00
	390	15	8.8	385	15	AA943092	EST198591 Normalized r	1.12e+00
c	391	15	8.8	385	18	AI091959	qa59d11.s1 Soares_feta	1.12e+00
c	392	15	8.8	385	12	AA246609	LD05437.3prime LD Dros	1.12e+00
c	393	15	8.8	385	5	R24065	yh33f03.s1 Homo sapien	1.12e+00
c	394	15	8.8	385	21	N98773	zb85g05.s1 Soares sene	1.12e+00
	395	15	8.8	386	9	AA525196	ni52a04.s1 NCI_CGAP_Ov	1.12e+00
	396	15	8.8	387	11	AA701356	zi63e12.s1 Soares feta	1.12e+00
	397	15	8.8	387	12	AA740790	ob32f05.s1 NCI_CGAP_Ki	1.12e+00
c	398	15	8.8	387	10	AA627901	nq70b10.s1 NCI_CGAP_Pr	1.12e+00
c	399	15	8.8	387	8	AA132660	zo18e07.s1 Stratagene	1.12e+00

c	400	15	8.8	388	22	N98042	2104C3 czapPFDd2.1, De	1.12e+00
c	401	15	8.8	388	10	AA405724	zu66d12.s1 Soares test	1.12e+00
c	402	15	8.8	389	9	AA565316	nj37c09.s1 NCI_CGAP_AA	1.12e+00
	403	15	8.8	390	5	R18497	yf95h07.r1 Homo sapien	1.12e+00
c	404	15	8.8	390	18	AI152653	ud93c09.r1 Soares mous	1.12e+00
c	405	15	8.8	390	18	AI052280	oy58g01.x1 NCI_CGAP_Br	1.12e+00
c	406	15	8.8	390	22	AA123612	MBAPCE9C09T3 Brugia ma	1.12e+00
c	407	15	8.8	391	12	AA816254	LD01056.3prime LD Dros	1.12e+00
	408	15	8.8	392	27	AQ045416	RPCI11-32N12.TJ RPCI11	1.12e+00
c	409	15	8.8	392	18	AI175653	EST219212 Normalized r	1.12e+00
c	410	15	8.8	392	12	AA765520	oa07d07.s1 NCI_CGAP_GC	1.12e+00
	411	15	8.8	393	29	AQ228320	HS_2019_A2_F11_T7 CIT	1.12e+00
c	412	15	8.8	393	28	AQ125403	HS_2168_A2_H08_MF CIT	1.12e+00
	413	15	8.8	395	26	B40957	HS-1052-B2-A08-MF.abi	1.12e+00
	414	15	8.8	395	28	AQ102151	HS_3035_A1_C06_MF CIT	1.12e+00
c	415	15	8.8	395	23	AA152806	mq45c08.r1 Soares 2NbM	1.12e+00
c	416	15	8.8	396	12	AA778547	af86b06.s1 Soares test	1.12e+00
	417	15	8.8	397	7	HUM428H06B	Human fetal brain cDNA	1.12e+00
c	418	15	8.8	397	27	AQ027100	CIT-HSP-2323F13.TF CIT	1.12e+00
c	419	15	8.8	398	26	B40380	HS-1051-A2-E06-MR.abi	1.12e+00
c	420	15	8.8	399	10	AA632054	np62h10.s1 NCI_CGAP_Br	1.12e+00
c	421	15	8.8	399	26	B33409	HS-1017-B1-D07-MF.abi	1.12e+00
	422	15	8.8	400	28	AQ132304	HS_3012_A1_F11_MR CIT	1.12e+00
c	423	15	8.8	401	22	N97805	1280C3 czapPFDd2.1, De	1.12e+00
c	424	15	8.8	401	17	AU023386	Mus musculus unfertili	1.12e+00
c	425	15	8.8	402	18	AI168650	ox42h05.s1 Soares_NhHM	1.12e+00
c	426	15	8.8	402	27	AQ062293	CIT-HSP-2347P18.TR CIT	1.12e+00
c	427	15	8.8	402	13	AA868695	ak53f11.s1 Soares test	1.12e+00
c	428	15	8.8	402	28	AQ118338	HS_3004_A1_E01_MR CIT	1.12e+00
c	429	15	8.8	403	27	AQ004791	CIT-HSP-2294L13.TR CIT	1.12e+00
	430	15	8.8	404	10	AA588717	nn17h08.s1 NCI_CGAP_Co	1.12e+00
	431	15	8.8	404	19	H73595	ya11b08.s1 Homo sapien	1.12e+00
	432	15	8.8	404	22	AA157044	zo51d01.r1 Stratagene	1.12e+00
c	433	15	8.8	404	10	N80160	za65f03.s1 Homo sapien	1.12e+00
c	434	15	8.8	404	19	N27395	yw62f06.s1 Homo sapien	1.12e+00
	435	15	8.8	405	8	AA560319	vl67d08.r1 Knowles Sol	1.12e+00
	436	15	8.8	405	15	AA962531	or80g07.s1 NCI_CGAP_Lu	1.12e+00
c	437	15	8.8	405	10	AA424196	zv58d06.s1 Soares test	1.12e+00
c	438	15	8.8	405	11	AA128412	zm24e01.s1 Stratagene	1.12e+00
	439	15	8.8	406	28	AQ185099	HS_2229_A2_G03_MF CIT	1.12e+00
	440	15	8.8	406	22	AA157161	zo51f01.r1 Stratagene	1.12e+00
	441	15	8.8	406	14	C92209	Dictyostelium discoide	1.12e+00
c	442	15	8.8	406	19	N29702	yw78c08.s1 Homo sapien	1.12e+00
c	443	15	8.8	407	18	AI136680	UI-R-C2p-nq-d-09-0-UI.	1.12e+00
	444	15	8.8	408	15	AA962734	oo88c12.s1 NCI_CGAP_Ki	1.12e+00
	445	15	8.8	408	11	AA706464	ag95b10.s1 Stratagene	1.12e+00
	446	15	8.8	409	6	AA044007	zk58e02.r1 Soares preg	1.12e+00
c	447	15	8.8	409	18	AI142843	ox64b05.s1 Soares_NhHM	1.12e+00
c	448	15	8.8	409	27	AQ006715	CIT-HSP-2293B17.TF CIT	1.12e+00
	449	15	8.8	410	10	AA621700	af48h01.s1 Soares tota	1.12e+00
	450	15	8.8	410	15	AA912503	ol23f05.s1 Soares_NFL_	1.12e+00

c	451	15	8.8	410	26	B17345	345E20.TVB CIT978SKA1	1.12e+00
c	452	15	8.8	411	16	R79974	yi95a03.s1 Homo sapien	1.12e+00
c	453	15	8.8	412	5	T77071	yd72h06.s1 Homo sapien	1.12e+00
c	454	15	8.8	412	21	W73332	zd53a02.s1 Soares feta	1.12e+00
	455	15	8.8	413	7	AA463138	yg70h10.r1 Soares mous	1.12e+00
c	456	15	8.8	413	28	AQ180707	HS_3209_A2_H11_MR CIT	1.12e+00
c	457	15	8.8	413	18	AI104052	EST213341 Normalized r	1.12e+00
	458	15	8.8	414	18	AI161077	qc77h06.x1 Soares_plac	1.12e+00
	459	15	8.8	415	6	AA434538	zw53b09.r1 Soares tota	1.12e+00
c	460	15	8.8	415	13	AA641675	nr62g01.s1 NCI_CGAP_Ly	1.12e+00
c	461	15	8.8	416	7	AA508324	MAAD0414.M3R Schistoso	1.12e+00
c	462	15	8.8	416	12	AA746230	ob22d01.s1 NCI_CGAP_Ki	1.12e+00
c	463	15	8.8	417	11	AA694280	zi35b09.s1 Soares feta	1.12e+00
c	464	15	8.8	417	19	H96521	yw06g11.s1 Homo sapien	1.12e+00
c	465	15	8.8	417	5	R32416	yh76f11.r1 Homo sapien	1.12e+00
	466	15	8.8	418	23	AA116870	mq25g02.r1 Barstead MP	1.12e+00
	467	15	8.8	418	7	AA497990	vi69g01.r1 Stratagene	1.12e+00
	468	15	8.8	418	15	AA926966	om26a06.s1 Soares_NFL_	1.12e+00
	469	15	8.8	419	5	T66861	ya50d10.s2 Homo sapien	1.12e+00
	470	15	8.8	419	18	AI147709	qb43c05.x1 NCI_CGAP_Br	1.12e+00
c	471	15	8.8	419	13	AA824407	ai75e06.s1 Soares test	1.12e+00
c	472	15	8.8	419	27	AQ089803	HS_3008_B2_E06_T7 CIT	1.12e+00
c	473	15	8.8	419	28	AQ178476	HS_2222_B1_D09_MF CIT	1.12e+00
c	474	15	8.8	419	28	AQ128551	HS_2254_B1_D12_MF CIT	1.12e+00
	475	15	8.8	420	9	AA579711	nm80g05.s1 NCI_CGAP_Co	1.12e+00
c	476	15	8.8	421	28	AQ178472	HS_2222_B1_D04_MF CIT	1.12e+00
	477	15	8.8	422	26	B45528	HS-1062-A1-A01-MF.abi	1.12e+00
c	478	15	8.8	422	27	AQ000084	CIT-HSP-2285L1.TF CIT-	1.12e+00
c	479	15	8.8	422	28	AQ184547	HS_2205_A1_H01_MF CIT	1.12e+00
c	480	15	8.8	422	24	AA296459	EST11043 Umbilical vei	1.12e+00
c	481	15	8.8	423	6	AA412211	zt99h03.r1 Soares test	1.12e+00
c	482	15	8.8	424	21	AA027768	HPLA_CCLEE_102c10u HPL	1.12e+00
c	483	15	8.8	424	22	AA157073	zo51f01.s1 Stratagene	1.12e+00
	484	15	8.8	425	19	N68160	zaila08.s1 Homo sapien	1.12e+00
c	485	15	8.8	425	22	AA157060	zo51d01.s1 Stratagene	1.12e+00
c	486	15	8.8	425	17	AI034429	zw18d12.x1 Soares_para	1.12e+00
	487	15	8.8	427	9	AA603719	np24g08.s1 NCI_CGAP_GC	1.12e+00
c	488	15	8.8	427	15	AA983627	on44b11.s1 NCI_CGAP_Co	1.12e+00
	489	15	8.8	429	27	B86147	RPCI11-22C20.TV RPCI11	1.12e+00
c	490	15	8.8	429	18	AI170458	EST216385 Normalized r	1.12e+00
c	491	15	8.8	429	22	N51042	yv30b11.s1 Soares feta	1.12e+00
	492	15	8.8	430	19	N59251	yz82a12.s1 Homo sapien	1.12e+00
	493	15	8.8	432	5	T55199	yb43f07.s1 Homo sapien	1.12e+00
c	494	15	8.8	432	15	AI049906	an30g09.x1 Gessler Wil	1.12e+00
c	495	15	8.8	432	13	AA181152	zp69c12.s1 Stratagene	1.12e+00
c	496	15	8.8	432	11	AA069116	zm10e09.s1 Stratagene	1.12e+00
c	497	15	8.8	432	21	W68063	zd33a12.s1 Soares feta	1.12e+00
	498	15	8.8	433	12	AA779245	zj39g04.s1 Soares feta	1.12e+00
	499	15	8.8	433	11	AA692841	vr57h01.s1 Knowles Sol	1.12e+00
c	500	15	8.8	433	9	AA578369	n123c10.s1 NCI_CGAP_HS	1.12e+00
c	501	15	8.8	433	22	N51027	yv29f11.s1 Soares feta	1.12e+00

	502	15	8.8	434	17	AI088545	qb12b11.x1 Soares_preg	1.12e+00
c	503	15	8.8	434	16	H15231	ym30g06.r1 Homo sapien	1.12e+00
	504	15	8.8	435	10	AA605110	no71a10.s1 NCI_CGAP_AA	1.12e+00
c	505	15	8.8	435	16	H49465	yo23h12.s1 Homo sapien	1.12e+00
c	506	15	8.8	435	14	AA985609	oq84c03.s1 NCI_CGAP_Ki	1.12e+00
	507	15	8.8	436	15	AI052402	oy75f01.x1 NCI_CGAP_CL	1.12e+00
c	508	15	8.8	436	17	AU022658	Mus musculus unfertili	1.12e+00
c	509	15	8.8	436	22	N51035	yv29h11.s1 Soares feta	1.12e+00
c	510	15	8.8	436	19	H73760	ya16b05.s1 Homo sapien	1.12e+00
	511	15	8.8	437	12	AA754139	97GS1032 Rice Immature	1.12e+00
c	512	15	8.8	437	9	AA224803	nc33a05.s1 NCI_CGAP_Pr	1.12e+00
c	513	15	8.8	437	5	T47434	yb13g09.s1 Homo sapien	1.12e+00
c	514	15	8.8	437	19	N26655	yx93b04.s1 Homo sapien	1.12e+00
c	515	15	8.8	437	17	AI081868	ov24d04.x1 Soares_test	1.12e+00
	516	15	8.8	438	10	AA603718	np24g07.s1 NCI_CGAP_GC	1.12e+00
	517	15	8.8	440	14	AA828909	od76g09.s1 NCI_CGAP_Ov	1.12e+00
	518	15	8.8	441	21	W48837	zc42h03.s1 Soares sene	1.12e+00
	519	15	8.8	441	28	AQ180231	HS_3209_B2_C03_MR CIT	1.12e+00
c	520	15	8.8	441	29	AQ140603	HS_3078_B2_E10_MR CIT	1.12e+00
c	521	15	8.8	441	15	AA934594	oo69a03.s1 NCI_CGAP_GC	1.12e+00
c	522	15	8.8	442	22	AA011014	ze34a08.s1 Soares reti	1.12e+00
c	523	15	8.8	442	5	T92091	ye01e12.r1 Homo sapien	1.12e+00
c	524	15	8.8	442	26	AQ019022	CIT-HSP-2312O14.TR CIT	1.12e+00
c	525	15	8.8	443	13	AA599741	ag11e12.s1 Gessler Wil	1.12e+00
c	526	15	8.8	443	26	AA187082	HS_3128_A1_A02_T7 CIT	1.12e+00
c	527	15	8.8	444	13	AA846387	aj55e10.s1 Soares test	1.12e+00
c	528	15	8.8	445	16	H18871	yn52a11.s1 Homo sapien	1.12e+00
c	529	15	8.8	446	18	AI185644	qe65b07.x1 Soares_feta	1.12e+00
c	530	15	8.8	446	11	AA121401	zm23a07.s1 Stratagene	1.12e+00
	531	15	8.8	447	10	W08956	ma61b07.r1 Soares mous	1.12e+00
	532	15	8.8	448	13	AA858891	UI-R-A0-bd-d-01-0-UI.s	1.12e+00
c	533	15	8.8	448	8	AA470729	ne13f05.s1 NCI_CGAP_Co	1.12e+00
c	534	15	8.8	449	15	AI012176	EST206627 Normalized r	1.12e+00
c	535	15	8.8	449	20	W01745	za37g04.r1 Soares feta	1.12e+00
	536	15	8.8	450	18	AI167351	oz88c04.x1 Soares_sene	1.12e+00
c	537	15	8.8	451	19	N40916	yy52a03.s1 Homo sapien	1.12e+00
c	538	15	8.8	451	17	AI027302	ov93c06.x1 Soares_test	1.12e+00
	539	15	8.8	452	13	AA630533	ab96d01.s1 Stratagene	1.12e+00
	540	15	8.8	453	18	AI182729	ub75h07.r1 Soares mous	1.12e+00
c	541	15	8.8	453	29	AQ144456	HS_3084_A1_E12_MR CIT	1.12e+00
	542	15	8.8	454	26	AQ016262	CIT-HSP-2310P12.TR CIT	1.12e+00
c	543	15	8.8	454	5	R42857	yg05g09.s1 Homo sapien	1.12e+00
c	544	15	8.8	454	5	R49238	yg69b12.s1 Homo sapien	1.12e+00
	545	15	8.8	455	6	C05992	similar to none, mRNA	1.12e+00
c	546	15	8.8	455	28	AQ208883	HS_3056_B2_E03_MF CIT	1.12e+00
	547	15	8.8	456	6	HSPD07494	H.sapiens EST sequence	1.12e+00
c	548	15	8.8	456	17	AI083933	qf26d10.x1 NCI_CGAP_Br	1.12e+00
	549	15	8.8	457	5	T68418	yc42b02.s1 Homo sapien	1.12e+00
	550	15	8.8	458	26	B07311	G3547 MVAT4 sheared g	1.12e+00
c	551	15	8.8	458	17	AI067751	EST209434 Schistosoma	1.12e+00
	552	15	8.8	459	17	AI016720	ov03c10.x1 NCI_CGAP_Ki	1.12e+00

c	553	15	8.8	459	22	AA045557	zl66g11.s1 Stratagene	1.12e+00
c	554	15	8.8	460	10	AA654803	nt60e11.s1 NCI_CGAP_Pr	1.12e+00
	555	15	8.8	461	11	C83833	Dictyostelium discoide	1.12e+00
	556	15	8.8	463	13	AA845209	ak76c09.s1 Barstead sp	1.12e+00
	557	15	8.8	463	6	C06102	similar to none, mRNA	1.12e+00
c	558	15	8.8	463	19	N39238	yy50d06.s1 Homo sapien	1.12e+00
c	559	15	8.8	465	22	AA026634	ze93e07.s1 Soares feta	1.12e+00
c	560	15	8.8	465	28	AQ220608	HS_2006_B1_B10_MR CIT	1.12e+00
c	561	15	8.8	466	9	AA524108	ng35a11.s1 NCI_CGAP_Co	1.12e+00
c	562	15	8.8	466	16	H19616	yn54b09.r1 Homo sapien	1.12e+00
	563	15	8.8	467	10	AA643538	nq78c02.s1 NCI_CGAP_Co	1.12e+00
	564	15	8.8	468	8	AA225240	nc22d10.r1 NCI_CGAP_Pr	1.12e+00
c	565	15	8.8	468	12	AA781453	aj26f06.s1 Soares test	1.12e+00
c	566	15	8.8	468	11	AA722679	zg90h11.s1 Soares feta	1.12e+00
	567	15	8.8	469	19	N35123	yy20b01.s1 Homo sapien	1.12e+00
	568	15	8.8	470	10	AA422729	vd29b09.s1 Knowles Sol	1.12e+00
	569	15	8.8	470	18	AI087367	oz60f11.x1 Soares_sene	1.12e+00
c	570	15	8.8	470	17	AI085299	qf18g01.x1 NCI_CGAP_Br	1.12e+00
c	571	15	8.8	470	27	B89333	RPCI11-24M18.TVC RPCI1	1.12e+00
c	572	15	8.8	473	17	AA992588	ot97f09.s1 Soares_tota	1.12e+00
	573	15	8.8	474	11	AA567972	HL02065.5prime HL Dros	1.12e+00
	574	15	8.8	475	17	AI081561	ox59g09.s1 Soares_NhHM	1.12e+00
c	575	15	8.8	475	11	AA678591	ah06h02.s1 Gessler Wil	1.12e+00
	576	15	8.8	476	27	AQ000615	CIT-HSP-2290C1.TR CIT-	1.12e+00
c	577	15	8.8	477	28	AQ176526	HS_3207_A2_H11_MR CIT	1.12e+00
	578	15	8.8	479	15	AA976730	oq06f02.s1 NCI_CGAP_Lu	1.12e+00
c	579	15	8.8	479	16	H11885	ym15h05.r1 Homo sapien	1.12e+00
c	580	15	8.8	479	29	AQ142377	HS_3076_B1_H01_MF CIT	1.12e+00
c	581	15	8.8	479	17	AI067712	EST209395 Schistosoma	1.12e+00
	582	15	8.8	480	9	AA602573	no90e06.s1 NCI_CGAP_AA	1.12e+00
c	583	15	8.8	481	16	H27980	y159c01.s1 Homo sapien	1.12e+00
c	584	15	8.8	482	14	AA903062	ok51f10.s1 NCI_CGAP_Le	1.12e+00
	585	15	8.8	483	10	AA422849	vd28h09.s1 Knowles Sol	1.12e+00
	586	15	8.8	484	7	HUM509A07B	Human placenta cDNA 5'	1.12e+00
	587	15	8.8	485	15	AI017887	ou35f06.x1 Soares_NFL_	1.12e+00
	588	15	8.8	485	9	AA606198	fa16g02.s1 Appel Eisen	1.12e+00
	589	15	8.8	487	17	AA994804	ou54h07.s1 NCI_CGAP_Br	1.12e+00
	590	15	8.8	489	26	FR000531	F.rubripes GSS sequenc	1.12e+00
c	591	15	8.8	489	28	AQ174747	HS_3211_A2_E07_T7 CIT	1.12e+00
	592	15	8.8	490	21	W49706	sc43d11.s1 Soares sene	1.12e+00
c	593	15	8.8	490	16	H04153	yj45c12.r1 Homo sapien	1.12e+00
c	594	15	8.8	490	17	AA857448	oi03d03.s1 NCI_CGAP_HN	1.12e+00
c	595	15	8.8	491	13	AA609162	af12a03.s1 Soares test	1.12e+00
	596	15	8.8	492	14	C90605	Dictyostelium discoide	1.12e+00
c	597	15	8.8	492	17	AI022861	ow54c05.s1 Soares_para	1.12e+00
c	598	15	8.8	492	10	AA639294	nx19c09.s1 NCI_CGAP_Pr	1.12e+00
c	599	15	8.8	492	18	AI104036	EST213325 Normalized r	1.12e+00
	600	15	8.8	493	29	AF046563	Mus musculus clone OST	1.12e+00
c	601	15	8.8	493	26	B77765	T28P16TR TAMU Arabidop	1.12e+00
c	602	15	8.8	493	6	AA195026	zr35e07.s1 Soares NhHM	1.12e+00
	603	15	8.8	494	10	AA652552	ns72f02.s1 NCI_CGAP_Pr	1.12e+00

c	604	15	8.8	494	13	AA765519	oa07d05.s1 NCI_CGAP_GC	1.12e+00
	605	15	8.8	495	29	AQ251290	T29N20-Sp6 TAMU Arabid	1.12e+00
c	606	15	8.8	497	28	AQ121385	HS_3083_B1_H03_MF CIT	1.12e+00
c	607	15	8.8	497	16	H29377	ym32g05.r1 Homo sapien	1.12e+00
	608	15	8.8	498	14	C92913	Dictyostelium discoide	1.12e+00
c	609	15	8.8	499	10	AA622530	np21b05.s1 NCI_CGAP_GC	1.12e+00
c	610	15	8.8	499	29	AQ144849	HS_3092_A2_F12_MF CIT	1.12e+00
	611	15	8.8	500	28	AQ212680	HS_3115_B2_G03_T7 CIT	1.12e+00
c	612	15	8.8	501	10	AA426212	zw17c04.s1 Soares ovar	1.12e+00
c	613	15	8.8	502	17	AI093153	qa97g02.x1 Soares_feta	1.12e+00
	614	15	8.8	503	14	C90570	Dictyostelium discoide	1.12e+00
	615	15	8.8	504	6	AA007590	zh99a03.s1 Soares feta	1.12e+00
	616	15	8.8	504	5	R38636	yc88e05.s1 Homo sapien	1.12e+00
c	617	15	8.8	504	17	AA917067	ol46h01.s1 Soares_NFL_	1.12e+00
c	618	15	8.8	504	12	AA776760	ah49h09.s1 Soares test	1.12e+00
c	619	15	8.8	505	23	AA153421	mq66d07.r1 Soares 2NbM	1.12e+00
c	620	15	8.8	506	12	AA758245	zg51a07.s1 Soares pine	1.12e+00
	621	15	8.8	508	10	AA159481	zo74e02.s1 Stratagene	1.12e+00
c	622	15	8.8	508	22	AA025053	ze77f12.s1 Soares feta	1.12e+00
c	623	15	8.8	508	28	AQ204912	HS_3226_B2_F05_T7 CIT	1.12e+00
c	624	15	8.8	508	14	AA799820	EST189317 Normalized r	1.12e+00
c	625	15	8.8	510	9	AA558451	nk39b01.s1 NCI_CGAP_GC	1.12e+00
	626	15	8.8	511	15	RICC2778A	Rice cDNA, partial seq	1.12e+00
	627	15	8.8	512	7	AA454879	zx79h09.s1 Soares ovar	1.12e+00
c	628	15	8.8	512	27	AQ036962	CIT-HSP-2324H22.TR CIT	1.12e+00
c	629	15	8.8	512	17	AI122881	qb01e06.s1 Soares_preg	1.12e+00
c	630	15	8.8	512	13	AA859211	UI-R-E0-cd-a-04-0-UI.s	1.12e+00
c	631	15	8.8	512	26	B06754	cSRL-85a8-u cSRL flow	1.12e+00
	632	15	8.8	513	15	AI038867	ox48h12.x1 Soares_tota	1.12e+00
c	633	15	8.8	513	17	AI067869	EST209553 Schistosoma	1.12e+00
	634	15	8.8	514	19	N21378	yx54c01.s1 Homo sapien	1.12e+00
c	635	15	8.8	514	26	B72260	RPCI11-7020.TP RPCI11	1.12e+00
	636	15	8.8	515	28	AQ213058	HS_3243_A2_E08_T7 CIT	1.12e+00
	637	15	8.8	516	24	AA272146	vb63d07.r1 Barstead mo	1.12e+00
	638	15	8.8	516	11	AA699673	zi78c01.s1 Soares feta	1.12e+00
	639	15	8.8	516	23	AA240536	mv32a03.r1 GuayWoodfor	1.12e+00
	640	15	8.8	516	8	AA504245	aa61a01.s1 NCI_CGAP_GC	1.12e+00
c	641	15	8.8	516	14	AA888978	ak24e12.s1 Soares_test	1.12e+00
c	642	15	8.8	517	9	AA557203	nl74h11.s1 NCI_CGAP_Br	1.12e+00
c	643	15	8.8	517	10	AA582874	nn72b07.s1 NCI_CGAP_La	1.12e+00
	644	15	8.8	518	9	AA564472	ng50f11.s1 NCI_CGAP_Co	1.12e+00
c	645	15	8.8	518	28	AQ213512	HS_3021_B1_G11_T7 CIT	1.12e+00
c	646	15	8.8	518	12	AA803368	GM10737.5prime GM Dros	1.12e+00
c	647	15	8.8	519	12	AA776376	ah16f02.s1 Gessler Wil	1.12e+00
	648	15	8.8	520	18	AI128933	ox85e07.s1 Soares_sene	1.12e+00
	649	15	8.8	520	6	AA131849	zl35h05.r1 Soares preg	1.12e+00
c	650	15	8.8	521	11	H36943	15072 Lambda-PRL2 Arab	1.12e+00
c	651	15	8.8	522	13	AA812215	ob84g01.s1 NCI_CGAP_GC	1.12e+00
c	652	15	8.8	522	13	AA859161	UI-R-A0-bn-c-12-0-UI.s	1.12e+00
c	653	15	8.8	522	14	AA915884	oh82d02.s1 NCI_CGAP_Co	1.12e+00
c	654	15	8.8	522	7	AA494869	fa12c07.r1 Zebrafish I	1.12e+00

c	655	15	8.8	523	9	AA551667	nf98c04.s1 NCI_CGAP_Co	1.12e+00
c	656	15	8.8	523	8	AA548020	nk50f02.s1 NCI_CGAP_GC	1.12e+00
	657	15	8.8	524	11	AA678754	ah08f10.s1 Gessler Wil	1.12e+00
	658	15	8.8	525	17	AU014117	Schizosaccharomyces po	1.12e+00
	659	15	8.8	525	8	AA215802	zr98b03.r1 NCI_CGAP_GC	1.12e+00
	660	15	8.8	525	29	AQ252148	HS_3112_A1_C08_T7 CIT	1.12e+00
c	661	15	8.8	525	26	B29805	T17E13TRB TAMU Arabido	1.12e+00
c	662	15	8.8	525	5	T62523	yc04b07.s1 Homo sapien	1.12e+00
c	663	15	8.8	525	18	AI172081	EST218076 Normalized r	1.12e+00
	664	15	8.8	526	12	C84805	Dictyostelium discoide	1.12e+00
	665	15	8.8	527	10	AA641366	nr91f08.s1 NCI_CGAP_Pr	1.12e+00
c	666	15	8.8	527	6	AA410898	zv03d09.s1 Soares NhHM	1.12e+00
c	667	15	8.8	527	9	AA554083	nk33g12.s1 NCI_CGAP_Co	1.12e+00
	668	15	8.8	528	26	B15529	345F20.TV CIT978SKAl H	1.12e+00
	669	15	8.8	528	15	AI035176	TENGO267 T. cruzi epim	1.12e+00
c	670	15	8.8	528	11	AA675002	vm71a04.r1 Knowles Sol	1.12e+00
	671	15	8.8	529	12	AA264748	LD08438.5prime LD Dros	1.12e+00
c	672	15	8.8	530	26	B24926	F22L18TF IGF Arabidops	1.12e+00
	673	15	8.8	531	11	H37401	15530 Lambda-PRL2 Arab	1.12e+00
c	674	15	8.8	531	28	AQ179732	HS_3195_A1_E09_T7 CIT	1.12e+00
c	675	15	8.8	531	21	N95391	zb68b11.s1 Soares feta	1.12e+00
c	676	15	8.8	532	10	AA622695	np36b04.s1 NCI_CGAP_Lu	1.12e+00
	677	15	8.8	534	28	AQ195906	RPCI11-59A24.TJ RPCI11	1.12e+00
	678	15	8.8	535	28	AQ116388	CIT978SK-A-522A7.TV CI	1.12e+00
c	679	15	8.8	535	15	AA976233	oo30d01.s1 NCI_CGAP_Lu	1.12e+00
c	680	15	8.8	536	8	AA256653	za31b12.r1 NCI_CGAP_GC	1.12e+00
	681	15	8.8	537	9	AA605889	fa13e10.s1 Appel Eisen	1.12e+00
c	682	15	8.8	537	15	AI012962	EST207413 Normalized r	1.12e+00
c	683	15	8.8	538	15	AI032129	os76c02.s1 NCI_CGAP_GC	1.12e+00
c	684	15	8.8	539	18	AI102470	EST211759 Normalized r	1.12e+00
	685	15	8.8	541	27	B59156	CIT-HSP-2015C10.TR CIT	1.12e+00
	686	15	8.8	541	19	N50529	yy89c07.s1 Homo sapien	1.12e+00
c	687	15	8.8	541	8	AA515547	ni42b03.s1 NCI_CGAP_Lu	1.12e+00
c	688	15	8.8	541	13	AA859348	UI-R-E0-cl-b-11-0-UI.s	1.12e+00
c	689	15	8.8	541	13	AA819620	UI-R-A0-an-c-12-0-UI.s	1.12e+00
c	690	15	8.8	541	19	N63403	yy60b03.s1 Homo sapien	1.12e+00
	691	15	8.8	542	10	AA613359	nq31a01.s1 NCI_CGAP_Co	1.12e+00
c	692	15	8.8	542	28	AQ203793	HS_3105_A1_B05_T7 CIT	1.12e+00
c	693	15	8.8	546	15	AA983898	op63e08.s1 Soares_NFL_	1.12e+00
c	694	15	8.8	547	7	AA443736	zw95a02.s1 Soares tota	1.12e+00
	695	15	8.8	548	10	AA527216	ni27c04.s1 NCI_CGAP_Ew	1.12e+00
c	696	15	8.8	549	10	AA586319	nd18a10.s1 NCI_CGAP_Ov	1.12e+00
c	697	15	8.8	552	26	B77297	T32F6TF TAMU Arabidops	1.12e+00
	698	15	8.8	554	23	AA245413	my49d07.r1 Barstead mo	1.12e+00
	699	15	8.8	554	12	AA817218	LD22396.5prime LD Dros	1.12e+00
	700	15	8.8	554	14	C91348	Dictyostelium discoide	1.12e+00
	701	15	8.8	555	29	AQ244016	HS_2060_A2_B01_MR CIT	1.12e+00
c	702	15	8.8	555	27	B59831	CIT-HSP-345E20.TVC CIT	1.12e+00
	703	15	8.8	556	18	AU035084	Mus musculus cDNA, clo	1.12e+00
	704	15	8.8	556	13	AA626470	ab53g07.r1 Stratagene	1.12e+00
	705	15	8.8	557	23	AA224652	CpEST.087 pSKIIminusCp	1.12e+00

c	706	15	8.8	557	8	AA287801	zs54d10.r1 NCI_CGAP_GC	1.12e+00
c	707	15	8.8	558	9	AA554363	nk37a11.s1 NCI_CGAP_GC	1.12e+00
c	708	15	8.8	559	17	AI068223	EST209914 Schistosoma	1.12e+00
	709	15	8.8	560	26	FR0027894	F.rubripes GSS sequenc	1.12e+00
c	710	15	8.8	561	10	AA610560	np97g03.s1 NCI_CGAP_Lu	1.12e+00
c	711	15	8.8	561	9	AA564133	nk44e01.s1 NCI_CGAP_GC	1.12e+00
c	712	15	8.8	562	6	AA034065	zi05g08.s1 Soares feta	1.12e+00
c	713	15	8.8	563	29	AQ144332	HS_3080_A1_G11_MF CIT	1.12e+00
	714	15	8.8	564	6	AA436436	zv44g07.s1 Soares ovar	1.12e+00
	715	15	8.8	566	19	N30338	yw75f04.s1 Homo sapien	1.12e+00
c	716	15	8.8	566	17	AI067997	EST209686 Schistosoma	1.12e+00
c	717	15	8.8	566	18	AI101149	EST210438 Normalized r	1.12e+00
c	718	15	8.8	568	19	N28905	yx63f11.s1 Homo sapien	1.12e+00
c	719	15	8.8	568	23	AA243499	zr67d07.s1 Soares NhHM	1.12e+00
	720	15	8.8	570	20	W64752	md88d04.r1 Soares mous	1.12e+00
	721	15	8.8	570	12	C84690	Dictyostelium discoide	1.12e+00
c	722	15	8.8	571	17	AU016155	Mus musculus 2-cell em	1.12e+00
	723	15	8.8	574	9	AA605924	fa13h11.s1 Appel Eisen	1.12e+00
	724	15	8.8	574	6	AA401081	zv66d03.s1 Soares tota	1.12e+00
	725	15	8.8	575	6	AA436421	zv44e07.s1 Soares ovar	1.12e+00
c	726	15	8.8	575	15	C78440	Mus musculus 3.5-dpc b	1.12e+00
c	727	15	8.8	580	15	C78445	Mus musculus 3.5-dpc b	1.12e+00
c	728	15	8.8	580	11	AA100802	zm26d04.s1 Stratagene	1.12e+00
	729	15	8.8	581	18	AI172018	EST218013 Normalized r	1.12e+00
	730	15	8.8	581	14	AA911262	oe75a02.s1 NCI_CGAP_Lu	1.12e+00
	731	15	8.8	583	7	AA511462	vg14e10.r1 Soares mous	1.12e+00
	732	15	8.8	583	7	AA531859	TgESTzz32d08.r1 TgME49	1.12e+00
	733	15	8.8	584	19	N30331	yw75e04.s1 Homo sapien	1.12e+00
c	734	15	8.8	584	17	AU014682	Mus musculus 2-cell em	1.12e+00
	735	15	8.8	585	22	AA121507	zk88b12.s1 Soares preg	1.12e+00
c	736	15	8.8	586	26	FR0017946	F.rubripes GSS sequenc	1.12e+00
c	737	15	8.8	588	7	AA455470	aa17h11.s1 Soares NhHM	1.12e+00
c	738	15	8.8	588	18	AI172412	EST218419 Normalized r	1.12e+00
	739	15	8.8	589	26	B14115	A-892B8.TP CIT978SK Ho	1.12e+00
c	740	15	8.8	591	5	T63363	yc23c09.s1 Homo sapien	1.12e+00
c	741	15	8.8	591	6	AA034150	zi06e11.s1 Soares feta	1.12e+00
c	742	15	8.8	592	13	AA838587	oe91a11.s1 NCI_CGAP_Co	1.12e+00
c	743	15	8.8	595	22	W93595	zd95h03.s1 Soares feta	1.12e+00
	744	15	8.8	597	14	C91051	Dictyostelium discoide	1.12e+00
	745	15	8.8	598	28	AQ159115	nbxb0013A19f CUGI Rice	1.12e+00
	746	15	8.8	600	27	AQ056090	CIT-HSP-2338P8.TR CIT-	1.12e+00
c	747	15	8.8	605	28	AQ156926	nbxb0008L23f CUGI Rice	1.12e+00
	748	15	8.8	608	10	AA427933	zw34e01.s1 Soares ovar	1.12e+00
c	749	15	8.8	609	29	AQ235049	HS_2046_A2_G06_MR CIT	1.12e+00
	750	15	8.8	612	8	AA401831	zt44f10.s1 Soares ovar	1.12e+00
	751	15	8.8	614	22	AA166483	ms46f11.r1 Life Tech m	1.12e+00
	752	15	8.8	615	28	AQ163607	nbxb0007M10f CUGI Rice	1.12e+00
c	753	15	8.8	615	9	AA577704	nn22h12.s1 NCI_CGAP_Co	1.12e+00
c	754	15	8.8	616	9	AA573899	nk09a10.s1 NCI_CGAP_Co	1.12e+00
	755	15	8.8	619	26	FR0003208	F.rubripes GSS sequenc	1.12e+00
c	756	15	8.8	619	19	N49852	yz08g10.s1 Homo sapien	1.12e+00

c	757	15	8.8	619	26	FR0003400	F.rubripes GSS sequenc 1.12e+00
c	758	15	8.8	621	13	AA192331	zp97g01.s1 Stratagene 1.12e+00
	759	15	8.8	622	14	C92238	Dictyostelium discoide 1.12e+00
c	760	15	8.8	625	10	AA600951	ni46c09.s1 NCI_CGAP_Lu 1.12e+00
c	761	15	8.8	629	26	B67590	T23O14TF TAMU Arabidop 1.12e+00
c	762	15	8.8	633	10	AA618432	ng03g11.s1 NCI_CGAP_Co 1.12e+00
c	763	15	8.8	635	9	AA573820	nk08a05.s1 NCI_CGAP_Co 1.12e+00
c	764	15	8.8	643	27	B70078	CIT-HSP-2063G9.TF CIT- 1.12e+00
c	765	15	8.8	646	27	B70590	CIT-HSP-2059N15.TF CIT 1.12e+00
c	766	15	8.8	651	17	AI067882	EST209569 Schistosoma 1.12e+00
	767	15	8.8	652	10	AA660482	00368 MtRKE Medicago t 1.12e+00
c	768	15	8.8	653	28	AQ160732	nbxb0005L20f CUGI Rice 1.12e+00
c	769	15	8.8	657	28	AQ194292	RPCI11-61M4.TK RPCI11 1.12e+00
c	770	15	8.8	658	18	AI176666	EST220258 Normalized r 1.12e+00
c	771	15	8.8	659	26	B67676	T23N18TF TAMU Arabidop 1.12e+00
	772	15	8.8	661	7	AA522246	vi33g01.r1 Barstead mo 1.12e+00
	773	15	8.8	661	20	W28366	46g3 Human retina cDNA 1.12e+00
	774	15	8.8	663	28	AG014461	Homo sapiens genomic D 1.12e+00
c	775	15	8.8	664	26	AG004182	Homo sapiens genomic D 1.12e+00
c	776	15	8.8	664	26	B56178	CIT-HSP-2006D19.TR CIT 1.12e+00
c	777	15	8.8	664	27	AG010308	Homo sapiens genomic D 1.12e+00
	778	15	8.8	668	9	AA605679	fa22h07.s1 Ekkerpost s 1.12e+00
	779	15	8.8	669	28	AG014462	Homo sapiens genomic D 1.12e+00
c	780	15	8.8	672	27	AG010309	Homo sapiens genomic D 1.12e+00
c	781	15	8.8	672	27	AQ053730	RPCI11-49J21.TJ RPCI11 1.12e+00
c	782	15	8.8	672	26	AG004183	Homo sapiens genomic D 1.12e+00
c	783	15	8.8	674	15	AA996039	os30d03.s1 NCI_CGAP_Ki 1.12e+00
	784	15	8.8	675	14	AA940832	LD23088.5prime LD Dros 1.12e+00
	785	15	8.8	675	26	AG004202	Homo sapiens genomic D 1.12e+00
	786	15	8.8	675	27	AG010328	Homo sapiens genomic D 1.12e+00
	787	15	8.8	679	8	AA550232	1352m3 gmbPfHB3.1, G. 1.12e+00
	788	15	8.8	680	29	AQ240408	CIT-HSP-2386K2.TF.1 CI 1.12e+00
	789	15	8.8	681	27	AQ083647	CpG0540A CpIOWAgDNA1 C 1.12e+00
	790	15	8.8	681	18	AI148147	qb56a01.x1 NCI_CGAP_Br 1.12e+00
	791	15	8.8	684	14	C89839	Dictyostelium discoide 1.12e+00
	792	15	8.8	685	11	C83903	Dictyostelium discoide 1.12e+00
c	793	15	8.8	691	8	AA550494	1649m3 gmbPfHB3.1, G. 1.12e+00
	794	15	8.8	692	27	B91159	CIT-HSP-2173M18.TF CIT 1.12e+00
	795	15	8.8	698	22	AA117329	mn26d12.r1 Beddington 1.12e+00
c	796	15	8.8	699	26	B27520	F12F5TFB IGF Arabidops 1.12e+00
	797	15	8.8	710	26	B75160	RPCI11-13L9.TV RPCI11 1.12e+00
	798	15	8.8	713	28	AQ202869	RPCI11-65O14.TK RPCI11 1.12e+00
	799	15	8.8	716	15	AA929056	on50e07.s1 NCI_CGAP_Co 1.12e+00
	800	15	8.8	717	12	AA263976	LD06592.5prime LD Dros 1.12e+00
	801	15	8.8	719	28	AQ162735	mgxb0016K07r CUGI Rice 1.12e+00
	802	15	8.8	731	23	AA200538	mu36e12.r1 Soares 2NbM 1.12e+00
	803	15	8.8	750	14	AA951540	LD32088.5prime LD Dros 1.12e+00
	804	15	8.8	756	27	AQ055676	CIT-HSP-2341B11.TF CIT 1.12e+00
c	805	15	8.8	785	26	B20016	F14G10-T7.1 IGF Arabid 1.12e+00
c	806	15	8.8	835	14	AA829751	od35e12.s1 NCI_CGAP_GC 1.12e+00
c	807	15	8.8	847	27	AQ051148	nbxb0005aH09f CUGI Ric 1.12e+00

c	808	15	8.8	894	6	C22780	Dictyostelium discoide 1.12e+00
c	809	15	8.8	912	9	AA522527	ni38d09.s1 NCI_CGAP_Lu 1.12e+00
	810	15	8.8	935	26	B10881	F24H6-Sp6.1 IGF Arabid 1.12e+00
	811	15	8.8	1008	26	B08140	F8M2-T7 IGF Arabidopsi 1.12e+00
	812	15	8.8	1110	12	AA142308	CK02588.contig CK Dros 1.12e+00
c	813	15	8.8	1610	27	AF057117	Homo sapiens chromosom 1.12e+00
	814	14	8.2	73	11	AA690063	vr78d06.s1 Knowles Sol 2.17e+01
	815	14	8.2	76	7	MUSGS00919	Mouse 3'-directed cDN 2.17e+01
c	816	14	8.2	91	17	AI130303	ka12a03.y1 Onchocerca 2.17e+01
	817	14	8.2	115	11	AA733640	vu74c10.r1 Stratagene 2.17e+01
	818	14	8.2	150	10	AA402212	zu47e12.r1 Soares ovar 2.17e+01
	819	14	8.2	177	12	AA764903	nz76h11.s1 NCI_CGAP_GC 2.17e+01
c	820	14	8.2	190	18	AI145839	UI-R-BT0-pz-c-03-0-UI. 2.17e+01
c	821	14	8.2	196	17	AI111789	UI-R-Y0-mf-a-11-0-UI.s 2.17e+01
	822	14	8.2	200	27	AQ080088	CIT-HSP-2355O5.TF CIT- 2.17e+01
	823	14	8.2	240	20	N83874	KK3794F Homo sapiens c 2.17e+01
	824	14	8.2	252	27	AQ087512	HS_2200_A1_D02_MF CIT 2.17e+01
	825	14	8.2	254	22	AA161531	MBAFCE9H01T3 Brugia ma 2.17e+01
	826	14	8.2	257	21	W32466	zc68b01.r1 Soares feta 2.17e+01
	827	14	8.2	264	9	C74663	Rice cDNA, partial seq 2.17e+01
c	828	14	8.2	265	5	R34197	yh84e03.r1 Homo sapien 2.17e+01
c	829	14	8.2	270	13	AA497082	ae32b06.r1 Gessler Wil 2.17e+01
	830	14	8.2	276	12	AA732747	ai28b10.s1 Soares test 2.17e+01
c	831	14	8.2	280	5	T09731	0332m7 Plasmodium falc 2.17e+01
	832	14	8.2	282	19	H70273	yr95b03.r1 Homo sapien 2.17e+01
	833	14	8.2	287	5	T87286	yd91h02.r1 Homo sapien 2.17e+01
	834	14	8.2	288	5	R47232	ph4g2_19/1TV Homo sapi 2.17e+01
c	835	14	8.2	288	12	AA440999	LD15723.5prime LD Dros 2.17e+01
	836	14	8.2	291	28	AQ215283	HS_2262_A1_E05_T7 CIT 2.17e+01
c	837	14	8.2	296	14	AA836021	od28d01.s1 NCI_CGAP_GC 2.17e+01
c	838	14	8.2	297	26	B61073	T20G7TR TAMU Arabidops 2.17e+01
	839	14	8.2	300	9	C33034	C.elegans cDNA clone y 2.17e+01
	840	14	8.2	300	6	C07664	C.elegans cDNA clone y 2.17e+01
	841	14	8.2	300	6	C10251	C.elegans cDNA clone y 2.17e+01
	842	14	8.2	302	13	AA844562	aj20g01.s1 Soares para 2.17e+01
c	843	14	8.2	310	11	AA699771	zj81g07.s1 Soares feta 2.17e+01
c	844	14	8.2	312	22	N64120	yv60g07.r1 Soares feta 2.17e+01
	845	14	8.2	317	17	AI077019	SWOvAFCAP02E09SK Oncho 2.17e+01
	846	14	8.2	322	27	AQ004901	CIT-HSP-2288E18.TF CIT 2.17e+01
	847	14	8.2	328	6	CELK027H6R	C.elegans cDNA clone y 2.17e+01
	848	14	8.2	330	16	H02163	yj34c11.s1 Homo sapien 2.17e+01
	849	14	8.2	334	11	AA128964	zo08a09.r1 Stratagene 2.17e+01
c	850	14	8.2	338	7	AA499575	vi94h12.r1 Stratagene 2.17e+01
	851	14	8.2	340	12	C84904	Dictyostelium discoide 2.17e+01
c	852	14	8.2	341	12	C84867	Dictyostelium discoide 2.17e+01
c	853	14	8.2	342	7	AA449674	zx07b01.s1 Soares tota 2.17e+01
	854	14	8.2	343	20	ATTS3719	A. thaliana transcribe 2.17e+01
c	855	14	8.2	345	16	R92058	yp96d02.r1 Homo sapien 2.17e+01
	856	14	8.2	347	10	T04285	332 Lambda-PRL2 Arabid 2.17e+01
c	857	14	8.2	352	22	AA072480	SW3ICA1623SK Brugia ma 2.17e+01
	858	14	8.2	358	5	T08172	EST06063 Homo sapiens 2.17e+01

c	859	14	8.2	359	5	R11332	yf41g11.r1 Homo sapien 2.17e+01
	860	14	8.2	360	9	C60476	C.elegans cDNA clone y 2.17e+01
	861	14	8.2	360	7	CELK102E3F	C.elegans cDNA clone y 2.17e+01
c	862	14	8.2	362	7	AA454016	zx46c11.s1 Soares test 2.17e+01
c	863	14	8.2	365	15	AA821247	HL07982.5prime HL Dros 2.17e+01
c	864	14	8.2	371	10	AA642462	ns30d06.s1 NCI_CGAP_GC 2.17e+01
c	865	14	8.2	371	26	AQ019121	CIT-HSP-2303A19.TR CIT 2.17e+01
	866	14	8.2	372	5	R40264	yf80e03.s1 Homo sapien 2.17e+01
	867	14	8.2	374	16	H13287	yj10f06.s1 Homo sapien 2.17e+01
	868	14	8.2	377	22	N59689	yv55f06.r1 Soares feta 2.17e+01
c	869	14	8.2	377	14	AA964923	UI-R-C0-hb-g-09-0-UI.s 2.17e+01
c	870	14	8.2	377	28	AQ118218	HS_3003_A1_H11_MR CIT 2.17e+01
c	871	14	8.2	381	6	AA029246	zk12a07.r1 Soares preg 2.17e+01
	872	14	8.2	382	27	AQ001360	CIT-HSP-2283M23.TR CIT 2.17e+01
	873	14	8.2	383	20	W19740	zb36h01.r1 Soares para 2.17e+01
	874	14	8.2	384	28	AQ094475	HS_3023_B1_G06_T7 CIT 2.17e+01
c	875	14	8.2	389	18	AI136647	UI-R-C2p-ob-g-04-0-UI. 2.17e+01
	876	14	8.2	397	19	N69444	za35h09.s1 Homo sapien 2.17e+01
	877	14	8.2	398	13	AA829418	od27b09.s1 NCI_CGAP_GC 2.17e+01
	878	14	8.2	398	19	T26082	T221 Trypanosoma bruce 2.17e+01
	879	14	8.2	400	19	H90077	yu83a01.r1 Homo sapien 2.17e+01
	880	14	8.2	403	28	AQ194452	RPCI11-46P19.TK RPCI11 2.17e+01
	881	14	8.2	403	20	N99383	za38e08.r1 Homo sapien 2.17e+01
c	882	14	8.2	404	28	AQ192651	HS_2248_B1_F07_MF CIT 2.17e+01
	883	14	8.2	406	13	AA768710	oc86e11.s1 NCI_CGAP_GC 2.17e+01
	884	14	8.2	410	28	AQ195614	RPCI11-46D24.TK RPCI11 2.17e+01
c	885	14	8.2	411	15	AI037693	uh25e08.r1 Barstead mo 2.17e+01
	886	14	8.2	415	22	W96084	ze08f10.r1 Soares feta 2.17e+01
	887	14	8.2	415	23	AA058918	zf64b11.r1 Soares reti 2.17e+01
	888	14	8.2	417	6	CELK020F9R	C.elegans cDNA clone y 2.17e+01
	889	14	8.2	417	7	AA448789	zx09h03.r1 Soares tota 2.17e+01
c	890	14	8.2	418	27	AQ072341	HS_3011_A2_G12_T7 CIT 2.17e+01
c	891	14	8.2	418	5	R13045	yf71h08.r1 Homo sapien 2.17e+01
	892	14	8.2	419	11	AA704916	zj82c02.s1 Soares feta 2.17e+01
c	893	14	8.2	419	12	AA441201	LD15911.5prime LD Dros 2.17e+01
c	894	14	8.2	421	8	AA497058	aa42c11.s1 Soares NbHM 2.17e+01
	895	14	8.2	423	26	AQ015620	CIT-HSP-2309N15.TF CIT 2.17e+01
	896	14	8.2	423	7	AA443918	zv46h03.s1 Soares ovar 2.17e+01
c	897	14	8.2	423	5	R20236	yg19b09.r1 Homo sapien 2.17e+01
c	898	14	8.2	423	11	AA076233	zm90b10.s1 Stratagene 2.17e+01
c	899	14	8.2	425	11	AA687164	nv60a08.s1 NCI_CGAP_GC 2.17e+01
	900	14	8.2	426	27	AQ044419	CIT-HSP-2329C6.TF CIT- 2.17e+01
	901	14	8.2	431	10	AA595631	ni49e02.s1 NCI_CGAP_Lu 2.17e+01
	902	14	8.2	431	19	N62648	za14e09.s1 Homo sapien 2.17e+01
	903	14	8.2	433	27	B98939	CIT-HSP-2287A5.TR CIT- 2.17e+01
c	904	14	8.2	440	18	AI159776	qc90c12.x1 Soares_preg 2.17e+01
	905	14	8.2	441	19	N49251	yy83h11.s1 Homo sapien 2.17e+01
	906	14	8.2	444	27	AQ087138	HS_2187_B1_C05_MF CIT 2.17e+01
	907	14	8.2	444	12	AA775581	zf30a12.s1 Soares feta 2.17e+01
c	908	14	8.2	444	12	AA441195	LD15907.5prime LD Dros 2.17e+01
	909	14	8.2	445	28	AQ115872	RPCI11-55L20.TK RPCI11 2.17e+01

c	910	14	8.2	447	17	AI123649	oo16b05.x1 Soares_NSF_ 2.17e+01
	911	14	8.2	449	24	HSPD03503	H.sapiens EST sequence 2.17e+01
c	912	14	8.2	451	5	R21038	yg55g08.r1 Homo sapien 2.17e+01
c	913	14	8.2	453	27	AQ007428	CIT-HSP-2291A22.TF CIT 2.17e+01
	914	14	8.2	455	17	AI098085	ue28c11.x1 Sugano mous 2.17e+01
	915	14	8.2	456	11	AA677653	ae29h01.s1 Soares NbHF 2.17e+01
c	916	14	8.2	456	29	AQ243293	HS_2064_A1_H09_MR CIT 2.17e+01
c	917	14	8.2	458	27	B95176	CIT-HSP-2167L10.TF CIT 2.17e+01
	918	14	8.2	464	28	AQ208821	HS_3056_B2_H09_MF CIT 2.17e+01
c	919	14	8.2	465	11	T41903	5166 Lambda-PRL2 Arabi 2.17e+01
c	920	14	8.2	471	19	N50855	yy92a04.s1 Homo sapien 2.17e+01
c	921	14	8.2	474	28	AQ208881	HS_3056_B2_E04_MF CIT 2.17e+01
	922	14	8.2	476	12	AA141410	CK02001.3prime CK Dros 2.17e+01
c	923	14	8.2	477	14	AA833778	od61e07.s1 NCI_CGAP_GC 2.17e+01
	924	14	8.2	480	28	AQ214085	HS_2187_B1_A09_MR CIT 2.17e+01
c	925	14	8.2	480	19	N53405	yz25c05.s1 Homo sapien 2.17e+01
c	926	14	8.2	480	14	AA965101	UI-R-C0-gz-c-12-0-UI.s 2.17e+01
c	927	14	8.2	483	13	AA885448	am10a04.s1 Soares NFL 2.17e+01
c	928	14	8.2	483	11	N96857	21803 Lambda-PRL1 Arab 2.17e+01
c	929	14	8.2	484	28	AQ220895	HS_2193_A2_E04_MR CIT 2.17e+01
	930	14	8.2	485	7	C23861	Dictyostelium discoide 2.17e+01
c	931	14	8.2	485	29	AQ139205	HS_3090_A1_C05_MR CIT 2.17e+01
c	932	14	8.2	488	11	AA695230	GM02449.5prime GM Dros 2.17e+01
	933	14	8.2	489	29	AQ138807	HS_3075_A2_A09_MF CIT 2.17e+01
c	934	14	8.2	489	10	AA626131	vy75h11.s1 Knowles Sol 2.17e+01
c	935	14	8.2	491	26	B74027	T27I20TF TAMU Arabidop 2.17e+01
c	936	14	8.2	492	12	AA391511	LD10540.5prime LD Dros 2.17e+01
	937	14	8.2	495	26	AQ003641	CpG0297A CpIOWAgDNA1 C 2.17e+01
	938	14	8.2	499	13	AA868286	ak40c05.s1 Soares test 2.17e+01
c	939	14	8.2	500	17	AI082464	os71e10.x1 NCI_CGAP_GC 2.17e+01
	940	14	8.2	503	13	AA858306	ob13c06.s1 NCI_CGAP_Ki 2.17e+01
c	941	14	8.2	506	29	AQ148854	HS_3139_B1_H06_T7 CIT 2.17e+01
c	942	14	8.2	508	29	AQ149097	HS_3165_A1_C04_T7 CIT 2.17e+01
	943	14	8.2	509	28	AQ204611	HS_3229_B1_G05_T7 CIT 2.17e+01
	944	14	8.2	513	10	AA650514	nn86g08.s1 NCI_CGAP_Br 2.17e+01
	945	14	8.2	515	13	AA839460	vw50c12.r1 Soares mous 2.17e+01
c	946	14	8.2	517	15	AU009596	Schizosaccharomyces po 2.17e+01
c	947	14	8.2	521	28	AQ193116	HS_3060_B1_D06_MF CIT 2.17e+01
	948	14	8.2	524	23	AA061151	ml39g12.r1 Stratagene 2.17e+01
c	949	14	8.2	524	15	C80487	Mus musculus 3.5-dpc b 2.17e+01
c	950	14	8.2	528	12	AA141569	CK02096.5prime CK Dros 2.17e+01
c	951	14	8.2	528	28	AQ114978	CIT-HSP-2374C18.TF CIT 2.17e+01
	952	14	8.2	535	13	AA181179	zp57b04.s1 Stratagene 2.17e+01
c	953	14	8.2	536	11	AA594688	GM01292.5prime GM Dros 2.17e+01
c	954	14	8.2	539	23	AA124537	mp97g01.r1 Soares 2NbM 2.17e+01
c	955	14	8.2	541	8	AA280278	zt05g11.r1 NCI_CGAP_GC 2.17e+01
	956	14	8.2	553	14	AA895646	vx47h01.r1 Stratagene 2.17e+01
	957	14	8.2	554	26	B06803	cSRL-85f9-u cSRL flow 2.17e+01
c	958	14	8.2	555	12	AA768842	nz75e04.s1 NCI_CGAP_GC 2.17e+01
c	959	14	8.2	557	27	AQ038807	CIT-HSP-2326N4.TF CIT- 2.17e+01
	960	14	8.2	558	10	AA428142	zw57b03.s1 Soares tota 2.17e+01

c	961	14	8.2	559	18	AI175026	EST218544 Normalized r 2.17e+01
c	962	14	8.2	561	26	FR0012990	F.rubripes GSS sequenc 2.17e+01
	963	14	8.2	564	6	AA193291	zr39b02.r1 Soares NhHM 2.17e+01
	964	14	8.2	566	17	AI107248	GH06584.5prime GH Dros 2.17e+01
c	965	14	8.2	567	18	AI134662	GH12186.5prime GH Dros 2.17e+01
c	966	14	8.2	572	6	AA406551	zv12h06.r1 Soares NhHM 2.17e+01
c	967	14	8.2	576	27	AQ049462	cLM-14f3-t cLM Giardia 2.17e+01
	968	14	8.2	580	17	AU015154	Mus musculus 2-cell em 2.17e+01
	969	14	8.2	593	17	AI034263	ow15g04.x1 Soares_para 2.17e+01
c	970	14	8.2	594	19	N26123	yx90h08.s1 Homo sapien 2.17e+01
c	971	14	8.2	595	21	W84486	zd90a11.s1 Soares feta 2.17e+01
	972	14	8.2	596	21	W91238	mf72e12.r1 Soares mous 2.17e+01
c	973	14	8.2	596	28	AQ156616	nbxb0008M09f CUGI Rice 2.17e+01
c	974	14	8.2	597	18	AI101331	EST210620 Normalized r 2.17e+01
	975	14	8.2	604	14	C93038	Dictyostelium discoide 2.17e+01
	976	14	8.2	605	27	AQ013332	RPCI11-23D16.TKBF RPCI 2.17e+01
	977	14	8.2	623	14	C92234	Dictyostelium discoide 2.17e+01
	978	14	8.2	636	10	AA616192	vo96f12.r1 Soares mous 2.17e+01
c	979	14	8.2	639	12	AA263919	LD07164.5prime LD Dros 2.17e+01
	980	14	8.2	641	21	W63541	zc56c11.r1 Soares para 2.17e+01
c	981	14	8.2	643	26	AQ003566	CpG0256A CpIOWAgDNA1 C 2.17e+01
	982	14	8.2	658	26	AG008448	Homo sapiens genomic D 2.17e+01
	983	14	8.2	658	23	AA166136	ms28g04.r1 Stratagene 2.17e+01
c	984	14	8.2	672	12	AA739699	464 PtIPG2 Pinus taeda 2.17e+01
c	985	14	8.2	675	27	AQ008140	RPCI11-20N13.TPK RPCI1 2.17e+01
c	986	14	8.2	679	17	AI107724	GH05552.5prime GH Dros 2.17e+01
	987	14	8.2	680	23	AA105693	mm68h01.r1 Stratagene 2.17e+01
	988	14	8.2	681	18	AI172489	EST218502 Normalized r 2.17e+01
c	989	14	8.2	683	28	AQ194912	RPCI11-47N4.TJ RPCI11 2.17e+01
	990	14	8.2	685	7	C23855	Dictyostelium discoide 2.17e+01
	991	14	8.2	691	26	B03398	cSRL-178D2-u cSRL flow 2.17e+01
	992	14	8.2	701	26	AG003175	Homo sapiens genomic D 2.17e+01
	993	14	8.2	702	28	AQ200089	RPCI11-59P22.TK RPCI11 2.17e+01
	994	14	8.2	719	20	W22873	74H9 Human retina cDNA 2.17e+01
c	995	14	8.2	729	26	AG006039	Homo sapiens genomic D 2.17e+01
	996	14	8.2	753	26	B71929	RPCI11-13G5.TV RPCI11 2.17e+01
c	997	14	8.2	809	26	B10155	F4P14-Sp6.1 IGF Arabid 2.17e+01
c	998	14	8.2	979	26	B09905	T17I18-T7 TAMU Arabido 2.17e+01
	999	14	8.2	1014	7	C25704	Dictyostelium discoide 2.17e+01
	1000	14	8.2	1017	26	B11791	T23F3-T7 TAMU Arabidop 2.17e+01

ALIGNMENTS

```
RESULT    1
LOCUS       AI004135       281 bp    mRNA            EST       27-AUG-1998
DEFINITION  ot43g02.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1619570
            3', mRNA sequence.
ACCESSION   AI004135
NID         g3213645
KEYWORDS    EST.
```



```
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 281)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 583   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 152.
FEATURES              Location/Qualifiers
     source           1. .281
                      /organism="Homo sapiens"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                      was prepared from mRNA obtained from Clontech
                      Laboratories, Inc., and primed with  a Not I - oligo(dT)
                      primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3'].
                      Double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Not I and cloned into the Not I
                      and Eco RI sites of the modified pT7T3 vector. Library
                      went through one round of normalization to Cot5, and was
                      constructed  by Bento Soares and M. Fatima Bonaldo."
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1619570"
                      /clone_lib="Soares_testis_NHT"
                      /sex="male"
                      /lab_host="DH10B"
BASE COUNT       69 a      67 c      52 g      93 t
ORIGIN

  Query Match          100.0%;  Score 171;  DB 17;  Length 281;
  Best Local Similarity  99.4%;  Pred. No. 0.00e+00;
  Matches   171;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db      28 TTTCACTTATTATTTATTTTATCTTCCAATTTCCTCTTGCCAGACTCCCATCCAAAGAGT 87
```



```
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     172 TTTCACTTATTATTTATTTTATCTTCCAATTTCCTCTTGCCAGACTCCCATCCAAAGAGT 113

Db      88 CATAGCAGCCTTCTTCCACCTTCTACATGAAATACATCCCCACCTGAACAAAGGCACACG 147
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     112 NATAGCAGCCTTCTTCCACCTTCTACATGAAATACATCCCCACCTGAACAAAGGCACACG 53

Db     148 ACAGGAGGAGGGGAATAGGACTTCGCAAACTGGACACGGCATCGTTCAGATC 199
            |||||||||||||||||||||||||||||||||||||||||||||||||||
Cp      52 ACAGGAGGAGGGGAATAGGACTTCGCAAACTGGACACGGCATCGTTCAGATC 1


RESULT    2
LOCUS       AA724792       314 bp    mRNA            EST       08-JAN-1998
DEFINITION  al05e02.s1 Soares parathyroid tumor NbHPA Homo sapiens cDNA clone
            1341914 3', mRNA sequence.
ACCESSION   AA724792
NID         g2742499
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 314)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 244.
FEATURES              Location/Qualifiers
     source           1. .314
                      /organism="Homo sapiens"
                      /note="Organ: parathyroid gland; Vector: pT7T3D
                      (Pharmacia) with a modified polylinker; Site_1: Not I;
                      Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                      oligo(dT) primer
                      [5'-
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCACCAATTTTTTTTTTTTTTTTTTTTTTT
```



```
                      T-3'], double-stranded cDNA was size selected, ligated to
                      Eco RI adapters (Pharmacia), digested with Not I and
                      cloned into the Not I and Eco RI sites of a modified pT7T3
                      vector (Pharmacia). Library went through one round of
                      normalization to a Cot = 5. Library constructed by Bento
                      Soares and M.Fatima Bonaldo. RNA from sporadic parathyroid
                      adenomas was kindly provided by Dr. Stephen Marx, National
                      Institute of Diabetes and Digestive and Kidney Diseases,
                      NIH."
                      /db_xref="taxon:9606"
                      /clone="1341914"
                      /clone_lib="Soares parathyroid tumor NbHPA"
                      /tissue_type="parathyroid tumor"
                      /dev_stage="adult"
                      /lab_host="DH10B (ampicillin resistant)"
BASE COUNT       72 a      80 c      58 g     104 t
ORIGIN

  Query Match          100.0%;  Score 171;  DB 12;  Length 314;
  Best Local Similarity  99.4%;  Pred. No. 0.00e+00;
  Matches   171;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db      11 TTTCACTTATTATTTATTTTATCTTCCAATTTCCTCTTGCCAGACTCCCATCCAAAGAGT 70
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     172 TTTCACTTATTATTTATTTTATCTTCCAATTTCCTCTTGCCAGACTCCCATCCAAAGAGT 113

Db      71 CATAGCAGCCTTCTTCCACCTTCTACATGAAATACATCCCCACCTGAACAAAGGCACACG 130
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     112 NATAGCAGCCTTCTTCCACCTTCTACATGAAATACATCCCCACCTGAACAAAGGCACACG 53

Db     131 ACAGGAGGAGGGGAATAGGACTTCGCAAACTGGACACGGCATCGTTCAGATC 182
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp      52 ACAGGAGGAGGGGAATAGGACTTCGCAAACTGGACACGGCATCGTTCAGATC 1


RESULT    3
LOCUS       AA626315       341 bp    mRNA            EST       15-OCT-1997
DEFINITION  zu93e06.s1 Soares testis NHT Homo sapiens cDNA clone 745570 3',
            mRNA sequence.
ACCESSION   AA626315
NID         g2538702
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 341)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
            Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
            Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
```

```
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40ml3 fwd. ET from Amersham.
FEATURES              Location/Qualifiers
     source           1..341
                      /organism="Homo sapiens"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                      was prepared from mRNA obtained from Clontech
                      Laboratories, Inc., and primed with  a Not I - oligo(dT)
                      primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3']."
                      Double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Not I and cloned into the Not I
                      and Eco RI sites of the modified pT7T3 vector. Library
                      went through one round of normalization to Cot5, and was
                      constructed  by Bento Soares and M. Fatima Bonaldo."
                      /db_xref="GDB:5933552"
                      /db_xref="taxon:9606"
                      /clone="745570"
                      /clone_lib="Soares testis NHT"
                      /sex="male"
                      /lab_host="DH10B"
BASE COUNT       84 a     85 c     65 g    107 t
ORIGIN

  Query Match          100.0%;  Score 171;  DB 10;  Length 341;
  Best Local Similarity  99.4%;  Pred. No. 0.00e+00;
  Matches   171;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db       5 TTTCACTTATTATTTATTTTATCTTCCAATTTCCTCTTGCCAGACTCCCATCCAAAGAGT 64
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     172 TTTCACTTATTATTTATTTTATCTTCCAATTTCCTCTTGCCAGACTCCCATCCAAAGAGT 113

Db      65 CATAGCAGCCTTCTTCCACCTTCTACATGAAATACATCCCCCACCTGAACAAAGGCACACG 124
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     112 NATAGCAGCCTTCTTCCACCTTCTACATGAAATACATCCCCCACCTGAACAAAGGCACACG 53

Db     125 ACAGGAGGAGGGGAATAGGACTTCGCAAACTGGACACGGCATCGTTCAGATC 176
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp      52 ACAGGAGGAGGGGAATAGGACTTCGCAAACTGGACACGGCATCGTTCAGATC 1
```



```
RESULT    4
LOCUS       AA448861      394 bp    mRNA            EST       04-JUN-1997
DEFINITION  zm11401.s1 Soares total fetus Nb2HF8 9w Homo sapiens cDNA clone
            786145 3', mRNA sequence.
ACCESSION   AA448861
NID         g2162531
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 394)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J.,
            Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-Merck EST Project 1997
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -41ml3 fwd. ET from Amersham.
FEATURES              Location/Qualifiers
     source           1..394
                      /organism="Homo sapiens"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                      was prepared from mRNA obtained from pooled 8-9 week
                      (total) fetus material with a Not I - oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCTTAATTTTTTTTTTTTTTTT 3']."
                      Double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Not I and cloned into the Not I
                      and Eco RI sites of the modified pT7T3 vector. Library
                      went through one round of normalization, and was
                      constructed  by Bento Soares and M. Fatima Bonaldo."
                      /db_xref="taxon:9606"
                      /clone="786145"
                      /clone_lib="Soares total fetus Nb2HF8 9w"
                      /dev_stage="8-9 weeks"
                      /lab_host="DH10B"
     mRNA             complement(<1. .>394)
                      /db_xref="GDB:5983764"
```



```
BASE COUNT      104 a     95 c     76 g    119 t
ORIGIN

  Query Match          100.0%;  Score 171;  DB 7;  Length 394;
  Best Local Similarity  99.4%;  Pred. No. 0.00e+00;
  Matches   171;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db       2 TTTCACTTATTATTTATTTTATCTTCCAATTTCCTCTTGCCAGACTCCCATCCAAAGAGT 61
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     172 TTTCACTTATTATTTATTTTATCTTCCAATTTCCTCTTGCCAGACTCCCATCCAAAGAGT 113

Db      62 CATAGCAGCCTTCTTCCACCTTCTACATGAAATACATCCCCCACCTGAACAAAGGCACACG 121
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     112 NATAGCAGCCTTCTTCCACCTTCTACATGAAATACATCCCCCACCTGAACAAAGGCACACG 53

Db     122 ACAGGAGGAGGGGAATAGGACTTCGCAAACTGGACACGGCATCGTTCAGATC 173
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp      52 ACAGGAGGAGGGGAATAGGACTTCGCAAACTGGACACGGCATCGTTCAGATC 1


RESULT    5
LOCUS       AI077370      420 bp    mRNA            EST       29-SEP-1998
DEFINITION  oy87f05.x1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA
            clone IMAGE:1672833 3', mRNA sequence.
ACCESSION   AI077370
NID         g3411778
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 420)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 561   Std Error: 0.00
            Seq primer: -40ml3 fwd. ET from Amersham
            High quality sequence stop: 256.
FEATURES              Location/Qualifiers
     source           1..420
                      /organism="Homo sapiens"
                      /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                      with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                      This is a subtracted version of the original Soares fetal
```



```
                      liver spleen 1NFLS library.   1st strand cDNA was primed
                      with a Pac I - oligo(dT) primer (5'
                      AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTT 3'],
                      double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Pac I and cloned into the Pac I
                      and Eco RI sites of the modified pT7T3 vector.  Library
                      went through one round of normalization. Library
                      constructed by Bento Soares and M.Fatima Bonaldo."
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1672833"
                      /clone_lib="Soares_fetal_liver_spleen_1NFLS_S1"
                      /sex="male"
                      /dev_stage="20 week-post conception fetus"
                      /lab_host="DH10B (ampicillin resistant)"
BASE COUNT      108 a     105 c     77 g    130 t
ORIGIN

  Query Match          100.0%;  Score 171;  DB 18;  Length 420;
  Best Local Similarity  99.4%;  Pred. No. 0.00e+00;
  Matches   171;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db       2 TTTCACTTATTATTTATTTTATCTTCCAATTTCCTCTTGCCAGACTCCCATCCAAAGAGT 61
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     172 TTTCACTTATTATTTATTTTATCTTCCAATTTCCTCTTGCCAGACTCCCATCCAAAGAGT 113

Db      62 CATAGCAGCCTTCTTCCACCTTCTACATGAAATACATCCCCCACCTGAACAAAGGCACACG 121
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     112 NATAGCAGCCTTCTTCCACCTTCTACATGAAATACATCCCCCACCTGAACAAAGGCACACG 53

Db     122 ACAGGAGGAGGGGAATAGGACTTCGCAAACTGGACACGGCATCGTTCAGATC 173
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp      52 ACAGGAGGAGGGGAATAGGACTTCGCAAACTGGACACGGCATCGTTCAGATC 1


RESULT    6
LOCUS       N78952        423 bp    mRNA            EST       29-MAR-1996
DEFINITION  zb18c05.s1 Homo sapiens cDNA clone 302408 3'.
ACCESSION   N78952
NID         g1241653
KEYWORDS    EST.
SOURCE      human clone=302408 primer=m13 -40 forward library=Soares fetal lung
            NbHL19W vector=pT7T3D (Pharmacia) with a modified polylinker
            host=DH10B (ampicillin resistant) Rsite1=Not I Rsite2=Eco RI 19
            week fetus.  1st strand cDNA was primed with a Not I - oligo(dT)
            primer  [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCATTTTTTTTTTTTTTTT-3'],
            double-stranded cDNA was size selected, ligated to Eco RI adapters
            (Pharmacia), digested with Not I and cloned into the Not I and Eco
            RI sites of a modified pT7T3 vector (Pharmacia). Library went
            through one round of normalization to a Cot = 5. Library
            constructed by Bento Soares and M.Fatima Bonaldo.
  ORGANISM  Homo sapiens
```

```
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 423)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 255
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES             Location/Qualifiers
     source          1. .423
                     /organism="Homo sapiens"
                     /clone="302408"
     mRNA            <1. .>423
BASE COUNT      107 a    105 c     77 g    133 t      1 others
ORIGIN

  Query Match          100.0%;  Score 171;  DB 20;  Length 423;
  Best Local Similarity  99.4%;  Pred. No. 0.00e+00;
  Matches   171;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db       3 TTTCACTTATTATTTATTTTATCTTCCAATTTCCTCTTGCCAGACTCCCATCCAAAGAGT 62
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     172 TTTCACTTATTATTTATTTTATCTTCCAATTTCCTCTTGCCAGACTCCCATCCAAAGAGT 113

Db      63 CATAGCAGCCTTCTTCCACCTTCTACATGAAATACATCCCCACCTGAACAAAGGCACACG 122
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     112 NATAGCAGCCTTCTTCCACCTTCTACATGAAATACATCCCCACCTGAACAAAGGCACACG 53

Db     123 ACAGGAGGAGGGGAATAGGACTTCGCAAACTGGACACGGCATCGTTCAGATC 174
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp      52 ACAGGAGGAGGGGAATAGGACTTCGCAAACTGGACACGGCATCGTTCAGATC 1


RESULT    7
LOCUS       AA435896      428 bp    mRNA            EST       09-NOV-1997
```



```
DEFINITION  zt80e12.s1 Soares testis NHT Homo sapiens cDNA clone 728686 3',
            mRNA sequence.
ACCESSION   AA435896
NID         g2140810
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 428)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
            Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
            Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 698   Std Error: 0.00
            Seq primer: -41m13 fwd. ET from Amersham
            High quality sequence stop: 392.
FEATURES             Location/Qualifiers
     source          1. .428
                     /organism="Homo sapiens"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from mRNA obtained from Clontech
                     Laboratories, Inc., and primed with  a Not I - oligo(dT)
                     primer (5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3']. 
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization to Cot5, and was
                     constructed  by Bento Soares and M. Fatima Bonaldo."
                     /db_xref="GDB:5925599"
                     /db_xref="taxon:9606"
                     /clone="728686"
                     /clone_lib="Soares testis NHT"
                     /sex="male"
                     /lab_host="DH10B"
BASE COUNT      106 a    104 c     77 g    141 t
ORIGIN
```



```
  Query Match          100.0%;  Score 171;  DB 10;  Length 428;
  Best Local Similarity  99.4%;  Pred. No. 0.00e+00;
  Matches   171;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db      14 TTTCACTTATTATTTATTTTATCTTCCAATTTCCTCTTGCCAGACTCCCATCCAAAGAGT 73
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     172 TTTCACTTATTATTTATTTTATCTTCCAATTTCCTCTTGCCAGACTCCCATCCAAAGAGT 113

Db      74 CATAGCAGCCTTCTTCCACCTTCTACATGAAATACATCCCCACCTGAACAAAGGCACACG 133
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     112 NATAGCAGCCTTCTTCCACCTTCTACATGAAATACATCCCCACCTGAACAAAGGCACACG 53

Db     134 ACAGGAGGAGGGGAATAGGACTTCGCAAACTGGACACGGCATCGTTCAGATC 185
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp      52 ACAGGAGGAGGGGAATAGGACTTCGCAAACTGGACACGGCATCGTTCAGATC 1


RESULT    8
LOCUS       W68342         437 bp    mRNA            EST       15-OCT-1996
DEFINITION  zd34c01.s1 Soares fetal heart NbHH19W Homo sapiens cDNA clone
            342528 3', mRNA sequence.
ACCESSION   W68342
NID         g1377212
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 437)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 521   Std Error: 0.00
            Seq primer: mob.REGA+ET
            High quality sequence stop: 254.
FEATURES             Location/Qualifiers
     source          1. .437
```



```
                     /organism="Homo sapiens"
                     /note="Organ: heart; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer (5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATCTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by
                     M.Fatima Bonaldo. This library was constructed from the
                     same fetus as the fetal lung library, Soares fetal lung
                     NbHL19W."
                     /db_xref="taxon:9606"
                     /clone="342528"
                     /clone_lib="Soares fetal heart NbHH19W"
                     /sex="unknown"
                     /dev_stage="19 weeks"
                     /lab_host="DH10B (ampicillin resistant)"
     mRNA            complement(<1. .>437)
BASE COUNT      108 a    110 c     83 g    134 t      2 others
ORIGIN

  Query Match          100.0%;  Score 171;  DB 21;  Length 437;
  Best Local Similarity  99.4%;  Pred. No. 0.00e+00;
  Matches   171;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db       2 TTTCACTTATTATTTATTTTATCTTCCAATTTCCTCTTGCCAGACTCCCATCCAAAGAGT 61
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     172 TTTCACTTATTATTTATTTTATCTTCCAATTTCCTCTTGCCAGACTCCCATCCAAAGAGT 113

Db      62 CATAGCAGCCTTCTTCCACCTTCTACATGAAATACATCCCCACCTGAACAAAGGCACACG 121
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     112 NATAGCAGCCTTCTTCCACCTTCTACATGAAATACATCCCCACCTGAACAAAGGCACACG 53

Db     122 ACAGGAGGAGGGGAATAGGACTTCGCAAACTGGACACGGCATCGTTCAGATC 173
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp      52 ACAGGAGGAGGGGAATAGGACTTCGCAAACTGGACACGGCATCGTTCAGATC 1


RESULT    9
LOCUS       AI147031       447 bp    mRNA            EST       29-SEP-1998
DEFINITION  qa82c01.x1 Soares_fetal_heart_NbHH19W Homo sapiens cDNA clone
            IMAGE:1693248 3', mRNA sequence.
ACCESSION   AI147031
NID         g3674713
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE   1  (bases 1 to 447)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40m13 fwd. ET from Amersham.
FEATURES              Location/Qualifiers
     source           1. .447
                      /organism="Homo sapiens"
                      /note="Organ: heart; Vector: pT7T3D (Pharmacia) with a
                      modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                      strand cDNA was primed with a Not I - oligo(dT) primer [5'-
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCATCTTTTTTTTTTTTTTTT 3'],
                      double-stranded cDNA was size selected, ligated to Eco RI
                      adapters (Pharmacia), digested with Not I and cloned into
                      the Not I and Eco RI sites of a modified pT7T3 vector
                      (Pharmacia). Library went through one round of
                      normalization to a Cot = 5. Library constructed by
                      M.Fatima Bonaldo. This library was constructed from the
                      same fetus as the fetal lung library, Soares fetal lung
                      NbHL19W."
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1693248"
                      /clone_lib="Soares_fetal_heart_NbHH19W"
                      /sex="unknown"
                      /dev_stage="19 weeks"
                      /lab_host="DH10B (ampicillin resistant)"
BASE COUNT      114 a    112 c     80 g    141 t
ORIGIN

  Query Match          100.0%;  Score 171;  DB 18;  Length 447;
  Best Local Similarity  99.4%;  Pred. No. 0.00e+00;
  Matches   171;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db      4 TTTCACTTATTATTTATTTTATCTTCCAATTTCCTCTTGCCAGACTCCCATCCAAAGAGT 63
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp    172 TTTCACTTATTATTTATTTTATCTTCCAATTTCCTCTTGCCAGACTCCCATCCAAAGAGT 113

Db     64 CATAGCAGCCTTCTTCCACCTTCTACATGAAATACATCCCCACCTGAACAAAGGCACACG 123
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp    112 NATAGCAGCCTTCTTCCACCTTCTACATGAAATACATCCCCACCTGAACAAAGGCACACG 53

Db    124 ACAGGAGGAGGGGAATAGGACTTCGCAAACTGGACACGGCATCGTTCAGATC 175
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     52 ACAGGAGGAGGGGAATAGGACTTCGCAAACTGGACACGGCATCGTTCAGATC 1
```



```
RESULT   10
LOCUS       AI092949      449 bp    mRNA            EST       05-OCT-1998
DEFINITION  qa81c06.x1 Soares_fetal_heart_NbHH19W Homo sapiens cDNA clone
            IMAGE:1693162 3', mRNA sequence.
ACCESSION   AI092949
NID         g3431925
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 449)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 525    Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham.
FEATURES              Location/Qualifiers
     source           1. .449
                      /organism="Homo sapiens"
                      /note="Organ: heart; Vector: pT7T3D (Pharmacia) with a
                      modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                      strand cDNA was primed with a Not I - oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCATCTTTTTTTTTTTTTTTT 3'],
                      double-stranded cDNA was size selected, ligated to Eco RI
                      adapters (Pharmacia), digested with Not I and cloned into
                      the Not I and Eco RI sites of a modified pT7T3 vector
                      (Pharmacia). Library went through one round of
                      normalization to a Cot = 5. Library constructed by
                      M.Fatima Bonaldo. This library was constructed from the
                      same fetus as the fetal lung library, Soares fetal lung
                      NbHL19W."
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1693162"
                      /clone_lib="Soares_fetal_heart_NbHH19W"
                      /sex="unknown"
                      /dev_stage="19 weeks"
                      /lab_host="DH10B (ampicillin resistant)"
BASE COUNT      114 a    112 c     82 g    141 t
ORIGIN

  Query Match          100.0%;  Score 171;  DB 18;  Length 449;
```



```
  Best Local Similarity  99.4%;  Pred. No. 0.00e+00;
  Matches   171;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db      4 TTTCACTTATTATTTATTTTATCTTCCAATTTCCTCTTGCCAGACTCCCATCCAAAGAGT 63
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp    172 TTTCACTTATTATTTATTTTATCTTCCAATTTCCTCTTGCCAGACTCCCATCCAAAGAGT 113

Db     64 CATAGCAGCCTTCTTCCACCTTCTACATGAAATACATCCCCACCTGAACAAAGGCACACG 123
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp    112 NATAGCAGCCTTCTTCCACCTTCTACATGAAATACATCCCCACCTGAACAAAGGCACACG 53

Db    124 ACAGGAGGAGGGGAATAGGACTTCGCAAACTGGACACGGCATCGTTCAGATC 175
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     52 ACAGGAGGAGGGGAATAGGACTTCGCAAACTGGACACGGCATCGTTCAGATC 1


RESULT   11
LOCUS       N78829        468 bp    mRNA            EST       29-MAR-1996
DEFINITION  zb17a05.s1 Homo sapiens cDNA clone 302288 3'.
ACCESSION   N78829
NID         g1241530
KEYWORDS    EST.
SOURCE      human clone=302288 primer=ml3 -40 forward library=Soares fetal lung
            NbHL19W vector=pT7T3D (Pharmacia) with a modified polylinker
            host=DH10B (ampicillin resistant) Rsite1=Not I Rsite2=Eco RI 19
            week fetus.  1st strand cDNA was primed with a Not I - oligo(dT)
            primer  [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCATTTTTTTTTTTTTTTT-3'],
            double-stranded cDNA was size selected, ligated to Eco RI adapters
            (Pharmacia), digested with Not I and cloned into the Not I and Eco
            RI sites of a modified pT7T3 vector (Pharmacia). Library went
            through one round of normalization to a Cot = 5. Library
            constructed by Bento Soares and M.Fatima Bonaldo.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 468)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
```



```
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 415
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES              Location/Qualifiers
     source           1. .468
                      /organism="Homo sapiens"
                      /clone="302288"
     mRNA             <1. .>468
BASE COUNT      121 a    116 c     89 g    142 t
ORIGIN

  Query Match          100.0%;  Score 171;  DB 20;  Length 468;
  Best Local Similarity  99.4%;  Pred. No. 0.00e+00;
  Matches   171;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db      2 TTTCACTTATTATTTATTTTATCTTCCAATTTCCTCTTGCCAGACTCCCATCCAAAGAGT 61
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp    172 TTTCACTTATTATTTATTTTATCTTCCAATTTCCTCTTGCCAGACTCCCATCCAAAGAGT 113

Db     62 CATAGCAGCCTTCTTCCACCTTCTACATGAAATACATCCCCACCTGAACAAAGGCACACG 121
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp    112 NATAGCAGCCTTCTTCCACCTTCTACATGAAATACATCCCCACCTGAACAAAGGCACACG 53

Db    122 ACAGGAGGAGGGGAATAGGACTTCGCAAACTGGACACGGCATCGTTCAGATC 173
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     52 ACAGGAGGAGGGGAATAGGACTTCGCAAACTGGACACGGCATCGTTCAGATC 1


RESULT   12
LOCUS       AA824370      549 bp    mRNA            EST       18-FEB-1998
DEFINITION  aj29a06.s1 Soares testis NHT Homo sapiens cDNA clone 1391698 3',
            mRNA sequence.
ACCESSION   AA824370
NID         g2896254
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 549)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
```

```
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40ml3 fwd. ET from Amersham
            High quality sequence stop: 465.
FEATURES             Location/Qualifiers
     source          1. .549
                     /organism="Homo sapiens"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI: 1st strand cDNA
                     was prepared from mRNA obtained from Clontech
                     Laboratories, Inc., and primed with  a Not I - oligo(dT)
                     primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3'].
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization to Cot5, and was
                     constructed  by Bento Soares and M. Fatima Bonaldo."
                     /db_xref="taxon:9606"
                     /clone="1391698"
                     /clone_lib="Soares testis NHT"
                     /sex="male"
                     /lab_host="DH10B"
BASE COUNT      139 a    140 c    101 g    169 t
ORIGIN

  Query Match          100.0%;  Score 171;  DB 13;  Length 549;
  Best Local Similarity  99.4%;  Pred. No. 0.00e+00;
  Matches   171;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db       6 TTTCACTTATTATTTATTTTATCTTCCAATTTCCTCTTGCCAGACTCCCATCCAAAGAGT 65
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     172 TTTCACTTATTATTTATTTTATCTTCCAATTTCCTCTTGCCAGACTCCCATCCAAAGAGT 113

Db      66 CATAGCAGCCTTCTTCCACCTTCTACATGAAATACATCCCCCACCTGAACAAAGGCACACG 125
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     112 NATAGCAGCCTTCTTCCACCTTCTACATGAAATACATCCCCCACCTGAACAAAGGCACACG 53

Db     126 ACAGGAGGAGGGGAATAGGACTTCGCAAACTGGACACGGCATCGTTCAGATC 177
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp      52 ACAGGAGGAGGGGAATAGGACTTCGCAAACTGGACACGGCATCGTTCAGATC 1


RESULT   13
LOCUS       AA862620      379 bp    mRNA            EST       11-MAR-1998
```



```
DEFINITION  oh07c11.s1 NCI_CGAP_Kid3 Homo sapiens cDNA clone IMAGE:1457108 3',
            mRNA sequence.
ACCESSION   AA862620
NID         g2955099
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 379)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40ml3 fwd. ET from Amersham
            High quality sequence stop: 325.
FEATURES             Location/Qualifiers
     source          1. .379
                     /organism="Homo sapiens"
                     /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer,
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia),  digested with Not I and cloned into the Not
                     I and Eco RI sites  of the modified pT7T3 vector.  mRNA
                     source: 2 pooled kidneys. Library went through one round
                     of normalization. Library constructed by Bento Soares and
                     M. Fatima Bonaldo."
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1457108"
                     /clone_lib="NCI_CGAP_Kid3"
                     /lab_host="DH10B"
BASE COUNT       99 a      92 c      75 g     113 t
ORIGIN

  Query Match          98.2%;  Score 168;  DB 13;  Length 379;
  Best Local Similarity  99.4%;  Pred. No. 0.00e+00;
  Matches   168;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;
```



```
Db       1 CACTTATTATTTATTTTATCTTCCAATTTCCTCTTGCCAGACTCCCATCCAAAGAGTCAT 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Cp     169 CACTTATTATTTATTTTATCTTCCAATTTCCTCTTGCCAGACTCCCATCCAAAGAGTNAT 110

Db      61 AGCAGCCTTCTTCCACCTTCTACATGAAATACATCCCCACCTGAACAAAGGCACACGACA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     109 AGCAGCCTTCTTCCACCTTCTACATGAAATACATCCCCACCTGAACAAAGGCACACGACA 50

Db     121 GGAGGAGGGGAATAGGACTTCGCAAACTGGACACGGCATCGTTCAGATC 169
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Cp      49 GGAGGAGGGGAATAGGACTTCGCAAACTGGACACGGCATCGTTCAGATC 1


RESULT   14
LOCUS       T98518        407 bp    mRNA            EST       31-MAR-1995
DEFINITION  ye60h05.s1 Homo sapiens cDNA clone 122169 3'.
ACCESSION   T98518
NID         g748255
KEYWORDS    EST.
SOURCE      human clone=122169 library=Soares fetal liver spleen 1NFLS
            vector=pT7T3D (Pharmacia) with a modified polylinker host=DH10B
            (ampicillin resistant) primer=-21m13 Rsitel=Pac I Rsite2=Eco RI
            Liver and spleen from a 20 week-post conception male fetus. 1st
            strand cDNA was primed with a Pac I - oligo(dT) primer [5'
            AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTT 3'], double-stranded
            cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac
            I and cloned into the Pac I and Eco RI sites of the modified pT7T3
            vector.  Library went through one round of normalization. Library
            constructed by Bento Soares and M.Fatima Bonaldo.
  ORGANISM  Homo sapiens
            Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 407)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High qality sequence stops: 241
            Source: IMAGE Consortium, LLNL
```



```
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES             Location/Qualifiers
     source          1. .407
                     /organism="Homo sapiens"
                     /clone="122169"
BASE COUNT       97 a      99 c      98 g     110 t      3 others
ORIGIN

  Query Match          98.2%;  Score 168;  DB 5;  Length 407;
  Best Local Similarity  99.4%;  Pred. No. 0.00e+00;
  Matches   168;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db       1 CACTTATTATTTATTTTATCTTCCAATTTCCTCTTGCCAGACTCCCATCCAAAGAGTCAT 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Cp     169 CACTTATTATTTATTTTATCTTCCAATTTCCTCTTGCCAGACTCCCATCCAAAGAGTNAT 110

Db      61 AGCAGCCTTCTTCCACCTTCTACATGAAATACATCCCCACCTGAACAAAGGCACACGACA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     109 AGCAGCCTTCTTCCACCTTCTACATGAAATACATCCCCACCTGAACAAAGGCACACGACA 50

Db     121 GGAGGAGGGGAATAGGACTTCGCAAACTGGACACGGCATCGTTCAGATC 169
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Cp      49 GGAGGAGGGGAATAGGACTTCGCAAACTGGACACGGCATCGTTCAGATC 1


RESULT   15
LOCUS       AI015595      450 bp    mRNA            EST       27-AUG-1998
DEFINITION  ov09h08.x1 NCI_CGAP_Kid3 Homo sapiens cDNA clone IMAGE:1636863 3',
            mRNA sequence.
ACCESSION   AI015595
NID         g3229931
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 450)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
```

```
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 568   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .450
                     /organism="Homo sapiens"
                     /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer,
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not
                     I and Eco RI sites of the modified pT7T3 vector. mRNA
                     source: 2 pooled kidneys. Library went through one round
                     of normalization. Library constructed by Bento Soares and
                     M. Fatima Bonaldo."
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1636863"
                     /clone_lib="NCI_CGAP_Kid3"
                     /lab_host="DH10B"
BASE COUNT      115 a    114 c     84 g    137 t
ORIGIN

  Query Match           98.2%;  Score 168;  DB 17;  Length 450;
  Best Local Similarity  99.4%;  Pred. No. 0.00e+00;
  Matches   168;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db       1 CACTTATTATTTATTTTATCTTCCAATTTCCTCTTGCCAGACTCCCATCCAAAGAGTCAT 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Cp     169 CACTTATTATTTATTTTATCTTCCAATTTCCTCTTGCCAGACTCCCATCCAAAGAGTNAT 110

Db      61 AGCAGCCTTCTTCCACCTTCTACATGAAATACATCCCCACCTGAACAAAGGCACACGACA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     109 AGCAGCCTTCTTCCACCTTCTACATGAAATACATCCCCACCTGAACAAAGGCACACGACA 50

Db     121 GGAGGAGGGGAATAGGACTTCGCAAACTGGACACGGCATCGTTCAGATC 169
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Cp      49 GGAGGAGGGGAATAGGACTTCGCAAACTGGACACGGCATCGTTCAGATC 1


RESULT   16
LOCUS       AA447925      470 bp    mRNA            EST       04-JUN-1997
DEFINITION  zm11d01.r1 Soares total fetus Nb2HF8 9w Homo sapiens cDNA clone
            786145 5', mRNA sequence.
ACCESSION   AA447925
NID         g2161595
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
```



```
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 470)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J.,
            Moore,B., Schallenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,T., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-Merck EST Project 1997
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 438.
FEATURES             Location/Qualifiers
     source          1. .470
                     /organism="Homo sapiens"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from mRNA obtained from pooled 8-9 week
                     (total) fetus material with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCTTAATTTTTTTTTTTTTTTT 3']."
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo."
                     /db_xref="taxon:9606"
                     /clone="786145"
                     /clone_lib="Soares total fetus Nb2HF8 9w"
                     /dev_stage="8-9 weeks"
                     /lab_host="DH10B"
     mRNA            <1. .>470
                     /db_xref="GDB:5983764"
BASE COUNT      135 a     91 c    120 g    124 t
ORIGIN

  Query Match           88.3%;  Score 151;  DB 7;  Length 470;
  Best Local Similarity  99.3%;  Pred. No. 0.00e+00;
  Matches   151;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db     319 GATCTGAACGATGCCGTGTCCAGTTTGCGAAGTCCTATTCCCCTCCTCCTGTCGTGTGCC 378
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```



```
Qy       1 GATCTGAACGATGCCGTGTCCAGTTTGCGAAGTCCTATTCCCCTCCTCCTGTCGTGTGCC 60

Db     379 TTTGTTCAGGTGGGGATGTATTTCATGTAGAAGGTGGAAGAAGGCTGCTATGACTCTTTG 438
           ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Qy      61 TTTGTTCAGGTGGGGATGTATTTCATGTAGAAGGTGGAAGAAGGCTGCTATNACTCTTTG 120

Db     439 GATGGGAGTCTGGCAAGAGGAAATTGGAAGAT 470
           ||||||||||||||||||||||||||||||||
Qy     121 GATGGGAGTCTGGCAAGAGGAAATTGGAAGAT 152


RESULT   17
LOCUS       W93038        381 bp    mRNA            EST       25-NOV-1996
DEFINITION  zd93b04.s1 Soares fetal heart NbHH19W Homo sapiens cDNA clone
            357007 3', mRNA sequence.
ACCESSION   W93038
NID         g1422189
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 381)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 465   Std Error: 0.00
            Seq primer: mob.REGA+ET
            High quality sequence stop: 330.
FEATURES             Location/Qualifiers
     source          1. .381
                     /organism="Homo sapiens"
                     /note="Organ: heart; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATCTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
```



```
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by
                     M.Fatima Bonaldo. This library was constructed from the
                     same fetus as the fetal lung library, Soares fetal lung
                     NbHL19W."
                     /db_xref="taxon:9606"
                     /clone="357007"
                     /clone_lib="Soares fetal heart NbHH19W"
                     /sex="unknown"
                     /dev_stage="19 weeks"
                     /lab_host="DH10B (ampicillin resistant)"
     mRNA            complement(<1. .>381)
BASE COUNT       98 a     94 c     76 g    113 t
ORIGIN

  Query Match           87.7%;  Score 150;  DB 22;  Length 381;
  Best Local Similarity  99.3%;  Pred. No. 0.00e+00;
  Matches   150;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db       3 TTTCACTTATTATTTATTTTATCTTCCAATTTCCTCTTGCCAGACTCCCATCCAAAGAGT 62
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     172 TTTCACTTATTATTTATTTTATCTTCCAATTTCCTCTTGCCAGACTCCCATCCAAAGAGT 113

Db      63 CATAGCAGCCTTCTTCCACCTTCTACATGAAATACATCCCCACCTGAACAAAGGCACACG 122
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     112 NATAGCAGCCTTCTTCCACCTTCTACATGAAATACATCCCCACCTGAACAAAGGCACACG 53

Db     123 ACAGGAGGAGGGGAATAGGACTTCGCAAACT 153
           ||||||||||||||||||||||||||||||
Cp      52 ACAGGAGGAGGGGAATAGGACTTCGCAAACT 22


RESULT   18
LOCUS       AA991266      405 bp    mRNA            EST       27-JUL-1998
DEFINITION  os31c03.s1 NCI_CGAP_Br2 Homo sapiens cDNA clone IMAGE:1606948 3',
            mRNA sequence.
ACCESSION   AA991266
NID         g3177755
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 405)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
```

```
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 572   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 386.
FEATURES              Location/Qualifiers
     source           1. .405
                      /organism="Homo sapiens"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; 1st strand cDNA was prepared from pooled bulk
                      breast tumor tissue, and was then primed with a Not I -
                      oligo(dT) primer. Double-stranded cDNA was ligated to Eco
                      RI adaptors (Pharmacia), digested with Not I and cloned
                      into the Not I and Eco RI sites of the modified pT7T3
                      vector.  This library is the normalized version of
                      NCI_CGAP_Br1.1. Library was constructed by Bento Soares
                      and M. Fatima Bonaldo."
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1606948"
                      /clone_lib="NCI_CGAP_Br2"
                      /sex="female, pooled"
                      /tissue_type="breast"
                      /lab_host="DH10B"
BASE COUNT      107 a      96 c      76 g     126 t
ORIGIN

  Query Match           86.0%;  Score 147;  DB 15;  Length 405;
  Best Local Similarity  99.3%;  Pred. No. 0.00e+00;
  Matches  147;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db      45 TCCAATTTCCTCTTGCCAGACTCCCATCCAAAGAGTCATAGCAGCCTTCTTCCACCTTCT 104
           ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Cp     148 TCCAATTTCCTCTTGCCAGACTCCCATCCAAAGAGTNATAGCAGCCTTCTTCCACCTTCT 89

Db     105 ACATGAAATACATCCCCACCTGAACAAAGGCACACGACAGGAGGAGGGGAATAGGACTTC 164
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp      88 ACATGAAATACATCCCCACCTGAACAAAGGCACACGACAGGAGGAGGGGAATAGGACTTC 29

Db     165 GCAAACTGGACACGGCATCGTTCAGATC 192
           ||||||||||||||||||||||||||||
Cp      28 GCAAACTGGACACGGCATCGTTCAGATC 1
```



```
RESULT   19
LOCUS       AA652731      253 bp    mRNA            EST       13-NOV-1997
DEFINITION  ns65b03.s1 NCI_CGAP_Pr22 Homo sapiens cDNA clone IMAGE:1188461,
            mRNA sequence.
ACCESSION   AA652731
NID         g2584383
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 253)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 364   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham.
FEATURES              Location/Qualifiers
     source           1. .253
                      /organism="Homo sapiens"
                      /note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
                      with a modified polylinker; 1st strand cDNA was prepared
                      from normal prostate bulk tissue, and was then primed with
                      a Not I - oligo(dT) primer. Double-stranded cDNA was
                      ligated to Eco RI adaptors (Pharmacia), digested with Not
                      I and cloned into the Not I  and Eco RI sites of the
                      modified pT7T3 vector.  Library is normalized, and was
                      constructed by Bento Soares and M. Fatima Bonaldo."
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1188461"
                      /clone_lib="NCI_CGAP_Pr22"
                      /sex="male"
                      /tissue_type="normal prostate"
                      /lab_host="DH10B"
BASE COUNT       61 a      68 c      55 g      69 t
```



```
ORIGIN

  Query Match           85.4%;  Score 146;  DB 10;  Length 253;
  Best Local Similarity  99.3%;  Pred. No. 0.00e+00;
  Matches  146;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db       1 CCAATTTCCTCTTGCCAGACTCCCATCCAAAGAGTCATAGCAGCCTTCTTCCACCTTCTA 60
           ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Cp     147 CCAATTTCCTCTTGCCAGACTCCCATCCAAAGAGTNATAGCAGCCTTCTTCCACCTTCTA 88

Db      61 CATGAAATACATCCCCACCTGAACAAAGGCACACGACAGGAGGAGGGGAATAGGACTTCG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp      87 CATGAAATACATCCCCACCTGAACAAAGGCACACGACAGGAGGAGGGGAATAGGACTTCG 28

Db     121 CAAACTGGACACGGCATCGTTCAGATC 147
           |||||||||||||||||||||||||||
Cp      27 CAAACTGGACACGGCATCGTTCAGATC 1


RESULT   20
LOCUS       W17346        481 bp    mRNA            EST       29-APR-1996
DEFINITION  zb18c05.r1 Soares fetal lung NbHL19W Homo sapiens cDNA clone 302408
            5', mRNA sequence.
ACCESSION   W17346
NID         g1291746
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 481)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Putative full length read
            Seq primer: ETPrimer
            High quality sequence stop: 456.
```



```
FEATURES              Location/Qualifiers
     source           1. .481
                      /organism="Homo sapiens"
                      /note="Organ: lung; Vector: pT7T3D (Pharmacia) with a
                      modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                      strand cDNA was primed with a Not I - oligo(dT) primer
                      [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCAATTTTTTTTTTTTTTTT-3'],
                      double-stranded cDNA was size selected, ligated to Eco RI
                      adapters (Pharmacia), digested with Not I and cloned into
                      the Not I and Eco RI sites of a modified pT7T3 vector
                      (Pharmacia). Library went through one round of
                      normalization to a Cot = 5. Library constructed by Bento
                      Soares and M.Fatima Bonaldo. This library was constructed
                      from the same fetus as the fetal heart library, Soares
                      fetal heart NbHH19W."
                      /db_xref="taxon:9606"
                      /clone="302408"
                      /clone_lib="Soares fetal lung NbHL19W"
                      /dev_stage="19 weeks"
                      /lab_host="DH10B (ampicillin resistant)"
     mRNA             <1. .>481
BASE COUNT      144 a      89 c     121 g     125 t      2 others
ORIGIN

  Query Match           73.7%;  Score 126;  DB 20;  Length 481;
  Best Local Similarity  98.4%;  Pred. No. 2.00e-25;
  Matches  126;  Conservative    0;  Mismatches   2;  Indels   0;  Gaps   0;

Db     313 GATCTGAACGATGCCGTGTCCAGTTTGCGAAGTCCTATTCCCCTCCTCCTGTCGTGTGCC 372
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy       1 GATCTGAACGATGCCGTGTCCAGTTTGCGAAGTCCTATTCCCCTCCTCCTGTCGTGTGCC 60

Db     373 TTTGTTCAGGTGGGGATGTATTNCATGTAGAAGGTGGAAGAAGGCTGCTATGACTCTTTG 432
           |||||||||||||||||||| |||||||||||||||||||||||||||||| ||||||||
Qy      61 TTTGTTCAGGTGGGGATGTATTTCATGTAGAAGGTGGAAGAAGGCTGCTATNACTCTTTG 120

Db     433 GATGGGAG 440
           ||||||||
Qy     121 GATGGGAG 128


RESULT   21
LOCUS       AA401790      424 bp    mRNA            EST       12-AUG-1997
DEFINITION  zt57g10.r1 Soares testis NHT Homo sapiens cDNA clone 726498 5',
            mRNA sequence.
ACCESSION   AA401790
NID         g2057274
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
```

```
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 424)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J.,
            Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-Merck EST Project 1997
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 538   Std Error: 0.00
            Seq primer: -28ml3 rev2 ET from Amersham
            High quality sequence stop: 404.
FEATURES            Location/Qualifiers
     source         1. .424
                    /organism="Homo sapiens"
                    /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                    polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                    was prepared from mRNA obtained from Clontech
                    Laboratories, Inc., and primed with a Not I - oligo(dT)
                    primer [5'
                    TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3']."
                    Double-stranded cDNA was ligated to Eco RI adaptors
                    (Pharmacia), digested with Not I and cloned into the Not I
                    and Eco RI sites of the modified pT7T3 vector. Library
                    went through one round of normalization to Cot5, and was
                    constructed by Bento Soares and M. Fatima Bonaldo."
                    /db_xref="taxon:9606"
                    /clone="726498"
                    /clone_lib="Soares testis NHT"
                    /sex="male"
                    /lab_host="DH10B"
     mRNA           <1. .>424
                    /db_xref="GDB:5923410"
BASE COUNT      120 a      86 c     105 g     113 t
ORIGIN

  Query Match           66.1%;  Score 113;  DB 8;  Length 424;
  Best Local Similarity  99.1%;  Pred. No. 2.87e-225;
  Matches   113;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db    311 GATCTGAACGATGCCGTGTCCAGTTTGCGAAGTCCTATTCCCCTCCTCCTGTCGTGTGCC 370
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```



```
Qy      1 GATCTGAACGATGCCGTGTCCAGTTTGCGAAGTCCTATTCCCCTCCTCCTGTCGTGTGCC 60

Db    371 TTTGTTCAGGTGGGGATGTATTTCATGTAGAAGGTGGAAGAAGGCTGCTATGAC 424
          ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Qy     61 TTTGTTCAGGTGGGGATGTATTTCATGTAGAAGGTGGAAGAAGGCTGCTATNAC 114


RESULT   22
LOCUS       AA027214       520 bp    mRNA            EST       01-FEB-1997
DEFINITION  ze97b09.s1 Soares fetal heart NbHH19W Homo sapiens cDNA clone
            366905 3', mRNA sequence.
ACCESSION   AA027214
NID         g1492916
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 520)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
            Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 575   Std Error: 0.00
            Seq primer: -40M13 fwd. from Amersham
            High quality sequence stop: 155.
FEATURES            Location/Qualifiers
     source         1. .520
                    /organism="Homo sapiens"
                    /note="Organ: heart; Vector: pT7T3D (Pharmacia) with a
                    modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                    strand cDNA was primed with a Not I - oligo(dT) primer [5'
                    TGTTACCAATCTGAAGTGGGAGCGGCCGCAATTTTTTTTTTTTTTTT 3'],
                    double-stranded cDNA was size selected, ligated to Eco RI
                    adapters (Pharmacia), digested with Not I and cloned into
                    the Not I and Eco RI sites of a modified pT7T3 vector
                    (Pharmacia). Library went through one round of
                    normalization to a Cot = 5. Library constructed by
                    M.Fatima Bonaldo. This library was constructed from the
```



```
                    same fetus as the fetal lung library, Soares fetal lung
                    NbHL19W."
                    /db_xref="taxon:9606"
                    /clone="366905"
                    /clone_lib="Soares fetal heart NbHH19W"
                    /sex="unknown"
                    /dev_stage="19 weeks"
                    /lab_host="DH10B (ampicillin resistant)"
     mRNA           complement(<1. .>520)
BASE COUNT      130 a     126 c     103 g     151 t      10 others
ORIGIN

  Query Match           63.7%;  Score 109;  DB 22;  Length 520;
  Best Local Similarity 100.0%;  Pred. No. 4.04e-215;
  Matches   109;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     64 AGCAGCCTTCTTCCACCTTCTACATGAAATACATCCCCACCTGAACAAAGGCACACGACA 123
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp    109 AGCAGCCTTCTTCCACCTTCTACATGAAATACATCCCCACCTGAACAAAGGCACACGACA 50

Db    124 GGAGGAGGGGAATAGGACTTCGCAAACTGGACACGGCATCGTTCAGATC 172
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     49 GGAGGAGGGGAATAGGACTTCGCAAACTGGACACGGCATCGTTCAGATC 1


RESULT   23
LOCUS       W17259         477 bp    mRNA            EST       29-APR-1996
DEFINITION  zb17a05.r1 Soares fetal lung NbHL19W Homo sapiens cDNA clone 302288
            5', mRNA sequence.
ACCESSION   W17259
NID         g1291694
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 477)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
```



```
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Putative full length read
            Seq primer: ETPrimer
            High quality sequence stop: 367.
FEATURES            Location/Qualifiers
     source         1. .477
                    /organism="Homo sapiens"
                    /note="Organ: lung; Vector: pT7T3D (Pharmacia) with a
                    modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                    strand cDNA was primed with a Not I - oligo(dT) primer
                    [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCAATTTTTTTTTTTTTTTT-3'],
                    double-stranded cDNA was size selected, ligated to Eco RI
                    adapters (Pharmacia), digested with Not I and cloned into
                    the Not I and Eco RI sites of a modified pT7T3 vector
                    (Pharmacia). Library went through one round of
                    normalization to a Cot = 5. Library constructed by Bento
                    Soares and M.Fatima Bonaldo. This library was constructed
                    from the same fetus as the fetal heart library, Soares
                    fetal heart NbHH19W."
                    /db_xref="taxon:9606"
                    /clone="302288"
                    /clone_lib="Soares fetal lung NbHL19W"
                    /dev_stage="19 weeks"
                    /lab_host="DH10B (ampicillin resistant)"
     mRNA           <1. .>477
BASE COUNT      139 a      89 c     120 g     127 t       2 others
ORIGIN

  Query Match           47.4%;  Score 81;  DB 20;  Length 477;
  Best Local Similarity  98.8%;  Pred. No. 4.94e-145;
  Matches    81;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db    313 GATCTGAACGATGCCGTGTCCAGTTTGCGAAGTNCTATTCCCCTCCTCCTGTCGTGTGCC 372
          ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Qy      1 GATCTGAACGATGCCGTGTCCAGTTTGCGAAGTCCTATTCCCCTCCTCCTGTCGTGTGCC 60

Db    373 TTTGTTCAGGTGGGGATGTATT 394
          |||||||||||||||||||||||
Qy     61 TTTGTTCAGGTGGGGATGTATT 82


RESULT   24
LOCUS       AA878662       262 bp    mRNA            EST       19-MAY-1998
DEFINITION  oj22c11.s1 NCI_CGAP_Kid5 Homo sapiens cDNA clone IMAGE:1492916 3',
            mRNA sequence.
ACCESSION   AA878662
NID         g2987627
KEYWORDS    EST.
SOURCE      human.
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 262)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 435   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 116.
FEATURES              Location/Qualifiers
     source           1. .262
                      /organism="Homo sapiens"
                      /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                      a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                      strand cDNA was primed with a Not I - oligo(dT) primer [5'
                      AACTGGAAGAATTCGCGGCCGCAATATTTTTTTTTTTTTTTTT 3'],
                      double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Not I and cloned into the Not I
                      and Eco RI sites of the modified pT7T3 vector. Library
                      went through one round of normalization. Library
                      constructed by Bento Soares and M. Fatima Bonaldo."
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1492916"
                      /clone_lib="NCI_CGAP_Kid5"
                      /tissue_type="2 pooled tumors (clear cell type)"
                      /lab_host="DH10B"
BASE COUNT       60 a      63 c      52 g      87 t
ORIGIN

  Query Match           46.8%;  Score 80;  DB 14;  Length 262;
  Best Local Similarity  98.8%;  Pred. No. 1.44e-142;
  Matches    80;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Db      6 TTTCACTTATTATTTATTTTATCTTCCAATTTCCTCTTGCCAGACTCCCATCCAAAGAGT 65
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp    172 TTTCACTTATTATTTATTTTATCTTCCAATTTCCTCTTGCCAGACTCCCATCCAAAGAGT 113
```



```
Db     66 CATAGCAGCCTTCTTCCACCT 86
          ||||||||||||||||||||||
Cp    112 NATAGCAGCCTTCTTCCACCT 92


RESULT   25
LOCUS       R07662        334 bp    mRNA            EST       05-APR-1995
DEFINITION  ye98b02.r1 Homo sapiens cDNA clone 125739 5'.
ACCESSION   R07662
NID         g759585
KEYWORDS    EST.
SOURCE      human clone=125739 library=Soares fetal liver spleen 1NFLS
            vector=pT7T3D (Pharmacia) with a modified polylinker host=DH10B
            (ampicillin resistant) primer=M13RP1 Rsite1=Pac I Rsite2=Eco RI
            Liver and spleen from a 20 week-post conception male fetus. 1st
            strand cDNA was primed with a Pac I - oligo(dT) primer [5'
            AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTT 3'], double-stranded
            cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac
            I and cloned into the Pac I and Eco RI sites of the modified pT7T3
            vector.  Library went through one round of normalization. Library
            constructed by Bento Soares and M.Fatima Bonaldo.
  ORGANISM  Homo sapiens
            Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 334)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High qality sequence stops: 158
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES              Location/Qualifiers
     source           1. .334
                      /organism="Homo sapiens"
                      /clone="125739"
BASE COUNT       89 a      58 c      94 g      88 t       5 others
ORIGIN
```



```
  Query Match           36.8%;  Score 63;  DB 5;  Length 334;
  Best Local Similarity  95.5%;  Pred. No. 2.91e-101;
  Matches    63;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Db    248 GATCTNAACGATNCCGTNTCCAGTTTGCGAAGTCCTATTCCCCTCCTCCTGTCGTGTGCC 307
          ||||| |||||| |||| ||||||||||||||||||||||||||||||||||||||||||
Qy      1 GATCTGAACGATGCCGTGTCCAGTTTGCGAAGTCCTATTCCCCTCCTCCTGTCGTGTGCC 60

Db    308 TTTGTT 313
          ||||||
Qy     61 TTTGTT 66


RESULT   26
LOCUS       W68383        442 bp    mRNA            EST       15-OCT-1996
DEFINITION  zd34c01.r1 Soares fetal heart NbHH19W Homo sapiens cDNA clone
            342528 5', mRNA sequence.
ACCESSION   W68383
NID         g1377324
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 442)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 521   Std Error: 0.00
            Seq primer: mob.REGA+ET
            High quality sequence stop: 383.
FEATURES              Location/Qualifiers
     source           1. .442
                      /organism="Homo sapiens"
                      /note="Organ: heart; Vector: pT7T3D (Pharmacia) with a
                      modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
```



```
                      strand cDNA was primed with a Not I - oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCATCTTTTTTTTTTTTTTTTT 3'],
                      double-stranded cDNA was size selected, ligated to Eco RI
                      adapters (Pharmacia), digested with Not I and cloned into
                      the Not I and Eco RI sites of a modified pT7T3 vector
                      (Pharmacia). Library went through one round of
                      normalization to a Cot = 5. Library constructed by
                      M.Fatima Bonaldo. This library was constructed from the
                      same fetus as the fetal lung library, Soares fetal lung
                      NbHL19W."
                      /db_xref="taxon:9606"
                      /clone="342528"
                      /clone_lib="Soares fetal heart NbHH19W"
                      /sex="unknown"
                      /dev_stage="19 weeks"
                      /lab_host="DH10B (ampicillin resistant)"
     mRNA             <1. .>442
BASE COUNT      122 a      90 c     111 g     119 t
ORIGIN

  Query Match           31.0%;  Score 53;  DB 21;  Length 442;
  Best Local Similarity 100.0%;  Pred. No. 1.19e-77;
  Matches    53;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Db    309 GATCTGAACGATGCCGTGTCCAGTTTGCGAAGTCCTATTCCCCTCCTCCTGTC 361
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      1 GATCTGAACGATGCCGTGTCCAGTTTGCGAAGTCCTATTCCCCTCCTCCTGTC 53


RESULT   27
LOCUS       AI143531      388 bp    mRNA            EST       28-SEP-1998
DEFINITION  qc03a03.x1 Soares_fetal_heart_NbHH19W Homo sapiens cDNA clone
            IMAGE:1708492 3', mRNA sequence.
ACCESSION   AI143531
NID         g3665340
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 388)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
```

```
            Seq primer: -40m13 fwd. ET from Amersham.
  FEATURES              Location/Qualifiers
       source           1. .388
                        /organism="Homo sapiens"
                        /note="Organ: heart; Vector: pT7T3D (Pharmacia) with a
                        modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                        strand cDNA was primed with a Not I - oligo(dT) primer [5'
                        TGTTACCAATCTGAAGTGGGAGCGGCCGCATCTTTTTTTTTTTTTTTT 3'],
                        double-stranded cDNA was size selected, ligated to Eco RI
                        adapters (Pharmacia), digested with Not I and cloned into
                        the Not I and Eco RI sites of a modified pT7T3 vector
                        (Pharmacia). Library went through one round of
                        normalization to a Cot = 5. Library constructed by
                        M.Fatima Bonaldo. This library was constructed from the
                        same fetus as the fetal lung library, Soares fetal lung
                        NbHL19W."
                        /db_xref="taxon:9606"
                        /clone="IMAGE:1708492"
                        /clone_lib="Soares_fetal_heart_NbHH19W"
                        /sex="unknown"
                        /dev_stage="19 weeks"
                        /lab_host="DH10B (ampicillin resistant)"
  BASE COUNT       97 a      93 c      67 g     131 t
  ORIGIN

    Query Match           27.5%;  Score 47;  DB 18;  Length 388;
    Best Local Similarity  97.5%;  Pred. No. 7.81e-64;
    Matches    77;  Conservative    0;  Mismatches   2;  Indels   0;  Gaps   0;

  Db       4 TTTCACTTATTATTTATTTTATCTTCCAATTTTCTCTTGCCAGACTCCCATCCAAAGAGT 63
             ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
  Cp     172 TTTCACTTATTATTTATTTTATCTTCCAATTTCCTCTTGCCAGACTCCCATCCAAAGAGT 113

  Db      64 CATAGCAGCCTTCTTCCAC 82
             ||||||||||||||||||
  Cp     112 NATAGCAGCCTTCTTCCAC 94


  RESULT   28
  LOCUS       W25674         326 bp    mRNA            EST       07-MAY-1996
  DEFINITION  zb72d07.r1 Soares fetal lung NbHL19W Homo sapiens cDNA clone 309133
              5', mRNA sequence.
  ACCESSION   W25674
  NID         g1303565
  KEYWORDS    EST.
  SOURCE      human.
    ORGANISM  Homo sapiens
              Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
              Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE   1  (bases 1 to 326)
    AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
```



```
              Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
              Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
              Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
              Wilson,R.
    TITLE     The WashU-Merck EST Project
    JOURNAL   Unpublished (1995)
  COMMENT
              Contact: Wilson RK
              WashU-Merck EST Project
              Washington University School of Medicine
              4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
              Tel: 314 286 1800
              Fax: 314 286 1810
              Email: est@watson.wustl.edu
              This clone is available royalty-free through LLNL ; contact the
              IMAGE Consortium (info@image.llnl.gov) for further information.
              Seq primer: mob.REGA+ET.
  FEATURES              Location/Qualifiers
       source           1. .326
                        /organism="Homo sapiens"
                        /note="Organ: lung; Vector: pT7T3D (Pharmacia) with a
                        modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                        strand cDNA was primed with a Not I - oligo(dT) primer
                        [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCAATTTTTTTTTTTTTTTT-3'],
                        double-stranded cDNA was size selected, ligated to Eco RI
                        adapters (Pharmacia), digested with Not I and cloned into
                        the Not I and Eco RI sites of a modified pT7T3 vector
                        (Pharmacia). Library went through one round of
                        normalization to a Cot = 5. Library constructed by Bento
                        Soares and M.Fatima Bonaldo. This library was constructed
                        from the same fetus as the fetal heart library, Soares
                        fetal heart NbHH19W."
                        /db_xref="taxon:9606"
                        /clone="309133"
                        /clone_lib="Soares fetal lung NbHL19W"
                        /dev_stage="19 weeks"
                        /lab_host="DH10B (ampicillin resistant)"
       mRNA             <1. .>326
  BASE COUNT      103 a      73 c      67 g      82 t       1 others
  ORIGIN

    Query Match           22.2%;  Score 38;  DB 20;  Length 326;
    Best Local Similarity  97.4%;  Pred. No. 8.59e-44;
    Matches    38;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

  Db     271 GATCTGAACGATGCCGTGTCCAGTTTGCGAAGTCCTANT 309
             |||||||||||||||||||||||||||||||||||||| |
  Qy       1 GATCTGAACGATGCCGTGTCCAGTTTGCGAAGTCCTATT 39


  RESULT   29
```



```
  LOCUS       AA070633       504 bp    mRNA            EST       23-DEC-1997
  DEFINITION  zm53a01.r1 Stratagene fibroblast (#937212) Homo sapiens cDNA clone
              529320 5', mRNA sequence.
  ACCESSION   AA070633
  NID         g1577993
  KEYWORDS    EST.
  SOURCE      human.
    ORGANISM  Homo sapiens
              Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
              Primates; Catarrhini; Hominidae; Homo.
  REFERENCE   1  (bases 1 to 504)
    AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
              Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
              Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
              Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
    TITLE     WashU-Merck EST Project
    JOURNAL   Unpublished (1995)
  COMMENT
              Contact: Wilson RK
              Washington University School of Medicine
              4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
              Tel: 314 286 1800
              Fax: 314 286 1810
              Email: est@watson.wustl.edu
              WARNING: There is evidence that suggests that the 384-well parent
              plate of this clone contains both human and mouse derived clones.
              Thus, the origin of this clone is uncertain. This caution should be
              kept in mind should you use this clone.

              This clone is available royalty-free through LLNL ; contact the
              IMAGE Consortium (info@image.llnl.gov) for further information.
              Insert Length: 651   Std Error: 0.00
              Seq primer: -28M13 rev2 from Amersham
              High quality sequence stop: 411.
  FEATURES              Location/Qualifiers
       source           1. .504
                        /organism="Homo sapiens"
                        /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                        XhoI; Cloned unidirectionally.  Primer: Oligo dT. W138
                        cell line. Average insert size: 0.8 kb; Uni-ZAP XR Vector;
                        -5' adaptor sequence: 5' GAATTCGGCACGAG 3' -3' adaptor
                        sequence: 5' CTCGAGTTTTTTTTTTTTTTTTT 3'"
                        /db_xref="GDB:3918831"
                        /db_xref="taxon:9606"
                        /clone="529320"
                        /clone_lib="Stratagene fibroblast (#937212)"
                        /lab_host="SOLR cells (kanamycin resistant)"
  BASE COUNT      133 a     101 c     121 g     131 t      18 others
  ORIGIN

    Query Match           21.1%;  Score 36;  DB 11;  Length 504;
```



```
    Best Local Similarity  92.3%;  Pred. No. 1.75e-39;
    Matches    36;  Conservative    0;  Mismatches   3;  Indels   0;  Gaps   0;

  Db     465 TTTGCGAAGTCCNATTCCCCTNCTCCTGTCGNGTGCCTT 503
             ||||||||| |||||||| ||||||||| ||||||| |||
  Qy      24 TTTGCGAAGTCCTATTCCCCTCCTCCTGTCGTGTGCCTT 62


  RESULT   30
  LOCUS       R07618         352 bp    mRNA            EST       05-APR-1995
  DEFINITION  ye98b02.s1 Homo sapiens cDNA clone 125739 3' similar to contains
              Alu repetitive element;.
  ACCESSION   R07618
  NID         g759541
  KEYWORDS    EST.
  SOURCE      human clone=125739 library=Soares fetal liver spleen 1NFLS
              vector=pT7T3D (Pharmacia) with a modified polylinker host=DH10B
              (ampicillin resistant) primer=-21m13 Rsitel=Pac I Rsite2=Eco RI
              Liver and spleen from a 20 week-post conception male fetus. 1st
              strand cDNA was primed with a Pac I - oligo(dT) primer [5'
              AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTT 3'], double-stranded
              cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac
              I and cloned into the Pac I and Eco RI sites of the modified pT7T3
              vector.  Library went through one round of normalization. Library
              constructed by Bento Soares and M.Fatima Bonaldo.
    ORGANISM  Homo sapiens
              Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
              Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE   1  (bases 1 to 352)
    AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
              Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
              Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
              Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
              Wilson,R.
    TITLE     The WashU-Merck EST Project
    JOURNAL   Unpublished (1995)
  COMMENT
              Contact: Wilson RK
              WashU-Merck EST Project
              Washington University School of Medicine
              4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
              Tel: 314 286 1800
              Fax: 314 286 1810
              Email: est@watson.wustl.edu
              High qality sequence stops: 291
              Source: IMAGE Consortium, LLNL
              This clone is available royalty-free through LLNL ; contact the
              IMAGE Consortium (info@image.llnl.gov) for further information.
  FEATURES              Location/Qualifiers
       source           1. .352
                        /organism="Homo sapiens"
```

/clone="125739"
BASE COUNT 97 a 78 c 71 g 103 t 3 others
ORIGIN

Query Match 12.9%; Score 22; DB 5; Length 352;
Best Local Similarity 91.7%; Pred. No. 1.02e-11;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 329 CATCCCCNCCTGANCAAAGGCACA 352
||||||| ||||| ||||||||||
Cp 78 CATCCCCACCTGAACAAAGGCACA 55

RESULT 31
LOCUS AI137425 441 bp mRNA EST 18-SEP-1998
DEFINITION UI-R-C2p-ok-d-10-0-UI.s1 UI-R-C2p Rattus norvegicus cDNA clone UI-R-C2p-ok-d-10-0-UI 3', mRNA sequence.
ACCESSION AI137425
NID g3638202
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 441)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized adult Spleen library. cDNA Library Preparation: M. Fatima Bonaldo, Ph.D. Clone distribution: clones will be available through Research Genetics
Seq primer: M13 Forward.
FEATURES Location/Qualifiers
source 1..441
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-C2p library is a subtracted library derived from the UI-R-C1

library, which is a subtracted library derived from the UI-R-C0 library. The UI-R-C0 library consisted of a mixture of individually tagged normalized libraries constructed from rat placenta, adult lung, brain, liver, kidney, heart, spleen, ovary, muscle, 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-C2p) was constructed as follows: PCR amplified cDNA inserts from UI-R-C1 clones from which 3' ESTs had been derived was used as a driver in a hybridization with the UI-R-C1 library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-C2p library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996)"
/db_xref="taxon:10116"
/clone="UI-R-C2p-ok-d-10-0-UI"
/clone_lib="UI-R-C2p"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
BASE COUNT 102 a 102 c 89 g 148 t
ORIGIN

Query Match 11.1%; Score 19; DB 18; Length 441;
Best Local Similarity 95.0%; Pred. No. 1.26e-06;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 82 AAGAGTCATAGCAGCCTTCT 101
|||||| ||||||||||||||
Cp 118 AAGAGTNATAGCAGCCTTCT 99

RESULT 32
LOCUS AA980297 551 bp mRNA EST 27-MAY-1998
DEFINITION ua33a08.r1 Soares mouse mammary gland NbMMG Mus musculus cDNA clone 1348502 5' similar to gb:J04694 Mus musculus alpha-1 type IV collagen (MOUSE);, mRNA sequence.
ACCESSION AA980297
NID g3158833
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 551)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,

Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT
Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:697294
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 103.
FEATURES Location/Qualifiers
source 1..551
/organism="Mus musculus"
/strain="C57BL/6J"
/note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer (5' TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATGGTTTTTTTTTTTTTTTTTTTTTT T 3'); double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. RNA provided by Dr. Minoru Ko, Wayne State Univ. Library constructed and normalized by Bento Soares and M.Fatima Bonaldo."
/db_xref="taxon:10090"
/clone="1348502"
/clone_lib="Soares mouse mammary gland NbMMG"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
BASE COUNT 158 a 127 c 130 g 136 t
ORIGIN

Query Match 11.1%; Score 19; DB 14; Length 551;
Best Local Similarity 100.0%; Pred. No. 1.26e-06;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 489 AAAATAAATAATAAGTGAA 507
|||||||||||||||||||
Qy 153 AAAATAAATAATAAGTGAA 171

RESULT 33
LOCUS AQ101466 291 bp DNA GSS 26-AUG-1998
DEFINITION HS_3028_B1_H04_MR CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3028 Col=7 Row=P, genomic survey sequence.
ACCESSION AQ101466
NID g3472495
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 291)
AUTHORS Mahairas,G.G., Zackrone,K.D., Smith,T., Tipton,S., Schmidt,S., Traicoff,R., Abajian,C., Blanchard,A., West,A. and Hood,L.E.
TITLE Construction of a Characterized Clone Resource for Genomic Sequencing: Generation and Preliminary Analysis of 20,000 Sequence Tagged Connectors
JOURNAL Unpublished (1997)
COMMENT
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3028 row: P column: 7
Class: BAC ends
High quality sequence stop: 291.
FEATURES Location/Qualifiers
source 1..291
/organism="Homo sapiens"
/note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in E-Coli DH10B"
/db_xref="taxon:9606"
/clone="Plate=3028 Col=7 Row=P"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
BASE COUNT 98 a 57 c 33 g 102 t 1 others
ORIGIN

Query Match 10.5%; Score 18; DB 28; Length 291;
Best Local Similarity 100.0%; Pred. No. 4.92e-05;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 193 TTATTTATTTTATCTTCC 210
||||||||||||||||||

```
Cp    163 TTATTTATTTTATCTTCC 146


RESULT   34
LOCUS       N68090         393 bp    mRNA            EST       13-MAR-1996
DEFINITION  za10e03.s1 Homo sapiens cDNA clone 292156 3'.
ACCESSION   N68090
NID         g1224251
KEYWORDS    EST.
SOURCE      human clone=292156 primer=m13 -40 forward library=Soares fetal
            liver spleen 1NFLS vector=pT7T3D (Pharmacia) with a modified
            polylinker host=DH10B (ampicillin resistant) Rsitel=Pac I
            Rsite2=Eco RI Liver and spleen from a 20 week-post conception male
            fetus. 1st strand cDNA was primed with a Pac I - oligo(dT) primer
            [5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTT 3'],
            double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia),
            digested with Pac I and cloned into the Pac I and Eco RI sites of
            the modified pT7T3 vector.  Library went through one round of
            normalization. Library constructed by Bento Soares and M.Fatima
            Bonaldo.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 393)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES              Location/Qualifiers
     source           1. .393
                      /organism="Homo sapiens"
                      /clone="292156"
     mRNA             <1. .>393
BASE COUNT      110 a     61 c     84 g    136 t      2 others
ORIGIN
```



```
  Query Match           10.5%;  Score 18;  DB 19;  Length 393;
  Best Local Similarity 100.0%;  Pred. No. 4.92e-05;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    136 CTTATTATTTATTTTATC 153
          ||||||||||||||||||
Cp    167 CTTATTATTTATTTTATC 150


RESULT   35
LOCUS       AQ246068       440 bp    DNA             GSS       05-OCT-1998
DEFINITION  HS_2057_B2_A10_T7 CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=2057 Col=20 Row=B, genomic survey
            sequence.
ACCESSION   AQ246068
NID         g3696250
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 440)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Furlong,J.,
            Shaker,R., Schmidt,S., Traicoff,R. and Hood,L.E.
  TITLE     Construction of a Characterized Clone Resource for Genomic
            Sequencing
  JOURNAL   Unpublished (1998)
COMMENT
            Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 2057  row: B  column: 20
            Class: BAC ends
            High quality sequence stop: 440.
FEATURES              Location/Qualifiers
     source           1. .440
                      /organism="Homo sapiens"
                      /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                      E-Coli DH10B"
                      /db_xref="taxon:9606"
                      /clone="Plate=2057 Col=20 Row=B"
                      /clone_lib="CIT Approved Human Genomic Sperm Library D"
                      /sex="male"
BASE COUNT      115 a    104 c     81 g    140 t
ORIGIN
```



```
  Query Match           10.5%;  Score 18;  DB 29;  Length 440;
  Best Local Similarity 100.0%;  Pred. No. 4.92e-05;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     16 TTATTATTTATTTTATCT 33
          ||||||||||||||||||
Cp    166 TTATTATTTATTTTATCT 149


RESULT   36
LOCUS       C06551         475 bp    mRNA            EST       08-AUG-1996
DEFINITION  similar to hypothetical protein B2, mRNA sequence.
ACCESSION   C06551
NID         g1503327
KEYWORDS    EST; EST(expressed sequence tag).
SOURCE      Rattus norvegicus rat pancreatic islets cDNA to mRNA, clone_lib:Rat
            pancreatic islet cDNA clone:RBC059.
  ORGANISM  Rattus norvegicus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae;
            Murinae; Rattus.
REFERENCE   1  (sites)
  AUTHORS   Takeda,J.
  TITLE     Large scale collection of expressed sequence tags (ESTs) from rat
            pancr eatic islet cDNA library
  JOURNAL   Unpublished (1996)
REFERENCE   2  (bases 1 to 475)
  AUTHORS   Takeda,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAY-1996) to the DDBJ/EMBL/GenBank databases. Jun
            Takeda, Institute for Molecular and cellular Regulation,Gunma
            University, Dept. of Molecular Medicine; 3-39-15 Showa-machi,
            Maebashi 371, japan (E-mail:jtakeda@new.sb.gunma-u.ac.jp,
            Tel:81-272-20-8856, Fax:81-272-20-8896)
COMMENT     Project='Rat pancreatic islet cDNA'
            vector=Lambda ZAPII
            primer=SK primer
            Rsitel=Eco RI
            Rsite2=Xho I
            mRNA was prepared from normal rat islets.  cDNA was directionally
            synthesized from the Xho I in the vector to the EcoRI site.
FEATURES              Location/Qualifiers
     source           1. .475
                      /organism="Rattus norvegicus"
                      /db_xref="taxon:10116"
                      /clone="RBC059"
                      /clone_lib="Rat pancreatic islet cDNA"
                      /tissue_type="rat pancreatic islets"
BASE COUNT      157 a     88 c     84 g    146 t
ORIGIN
```



```
  Query Match           10.5%;  Score 18;  DB 6;  Length 475;
  Best Local Similarity 100.0%;  Pred. No. 4.92e-05;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    216 AGATAAAATAAATAATAA 233
          ||||||||||||||||||
Qy    149 AGATAAAATAAATAATAA 166


RESULT   37
LOCUS       AA532018       572 bp    mRNA            EST       22-JUL-1997
DEFINITION  TgESTzx46g01.r1 TgME49 invivo Bradyzoite cDNA size selected
            Toxoplasma gondii cDNA clone tgzx46g01.r1 5', mRNA sequence.
ACCESSION   AA532018
NID         g2274724
KEYWORDS    EST.
SOURCE      Toxoplasma gondii.
  ORGANISM  Toxoplasma gondii
            Eukaryotae; mitochondrial eukaryotes; Alveolata; Apicomplexa;
            Coccidia; Eimeriida; Sarcocystidae; Toxoplasma.
REFERENCE   1  (bases 1 to 572)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,S.,
            Kucaba,T., Lacy,M., Le,N., Jost,S., Martin,J., Moore,B.,
            Schellenberg,K., Steptoe,M., Tan,F., Theising,B., Bowers,Y.,
            Wylie,T., Ajioka,J.A., Aslett,M.A., Sibley,L.D., Wan,K.L.,
            Waterston,R. and Wilson,R.
  TITLE     WashU-Stanford-PAMF-NIH Toxoplasma EST project
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Marra M
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: toxo@watson.wustl.edu
            Contact Steve Parmley, PAMF (76424.16@compuserve.com) for
            information on clone and library availability.
            High quality sequence stop: 399.
FEATURES              Location/Qualifiers
     source           1. .572
                      /organism="Toxoplasma gondii"
                      /strain="ME49"
                      /note="Vector: Bluescript II SK-; Site_1: EcoRI; Site_2:
                      NotI; Mature bradyzoites were obtained from infected mouse
                      brains by percoll density centrifugation.  The library was
                      constructed by Steve Parmley, Palo Alto Medical
                      Foundation.  cDNAs were synthesized by priming with oligo
                      d(T) and directionally cloned into the EcoRI/NotI sites of
                      lambda gt11.  Warning: the library contains a small
```

```
                     percentage of host cDNAs derived from mouse cells."
                     /db_xref="taxon:5811"
                     /clone="tgzz46g01.rl"
                     /clone_lib="TgME49 invivo Bradyzoite cDNA size selected"
                     /dev_stage="Bradyzoite"
                     /lab_host="BB4"
                     <1. .>572
     mRNA
BASE COUNT      133 a    130 c    151 g    158 t
ORIGIN

  Query Match           10.5%;  Score 18;  DB 7;  Length 572;
  Best Local Similarity 100.0%;  Pred. No. 4.92e-05;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    412 TGCCGTGTCCAGTTTGCG 429
          ||||||||||||||||||
Qy     12 TGCCGTGTCCAGTTTGCG 29


RESULT   38
LOCUS       AA531880       572 bp    mRNA            EST       22-JUL-1997
DEFINITION  TgESTzz32e10.rl TgME49 invivo Bradyzoite cDNA size selected
            Toxoplasma gondii cDNA clone tgzz32e10.rl 5', mRNA sequence.
ACCESSION   AA531880
NID         g2274586
KEYWORDS    EST.
SOURCE      Toxoplasma gondii.
  ORGANISM  Toxoplasma gondii
            Eukaryotae; mitochondrial eukaryotes; Alveolata; Apicomplexa;
            Coccidia; Eimeriida; Sarcocystidae; Toxoplasma.
REFERENCE   1  (bases 1 to 572)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,S.,
            Kucaba,T., Lacy,M., Le,N., Jost,S., Martin,J., Moore,B.,
            Schellenberg,K., Steptoe,M., Tan,P., Theising,B., Bowers,Y.,
            Wylie,T., Ajioka,J.A., Aslett,M.A., Sibley,L.D., Wan,K.L.,
            Waterston,R. and Wilson,R.
  TITLE     WashU-Stanford-PAMF-NIH Toxoplasma EST project
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Marra M
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: toxo@watson.wustl.edu
            Contact Steve Parmley, PAMF (76424.16@compuserve.com) for
            information on clone and library availability.
            High quality sequence stop: 423.
FEATURES             Location/Qualifiers
     source          1. .572
```



```
                     /organism="Toxoplasma gondii"
                     /strain="ME49"
                     /note="Vector: Bluescript II SK-; Site_1: EcoRI; Site_2:
                     NotI; Mature bradyzoites were obtained from infected mouse
                     brains by percoll density centrifugation.  The library was
                     constructed by Steve Parmley, Palo Alto Medical
                     Foundation.  cDNAs were synthesized by priming with oligo
                     d(T) and directionally cloned into the EcoRI/NotI sites of
                     lambda gt11.  Warning: the library contains a small
                     percentage of host cDNAs derived from mouse cells."
                     /db_xref="taxon:5811"
                     /clone="tgzz32e10.rl"
                     /clone_lib="TgME49 invivo Bradyzoite cDNA size selected"
                     /dev_stage="Bradyzoite"
                     /lab_host="BB4"
                     <1. .>572
     mRNA
BASE COUNT      127 a    124 c    158 g    163 t
ORIGIN

  Query Match           10.5%;  Score 18;  DB 7;  Length 572;
  Best Local Similarity 100.0%;  Pred. No. 4.92e-05;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    412 TGCCGTGTCCAGTTTGCG 429
          ||||||||||||||||||
Qy     12 TGCCGTGTCCAGTTTGCG 29


RESULT   39
LOCUS       B72809         604 bp    DNA             GSS       06-JUN-1998
DEFINITION  RPCI11-10F20.TP RPCI11 Homo sapiens genomic clone R-10F20, genomic
            survey sequence.
ACCESSION   B72809
NID         g2711960
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 604)
  AUTHORS   Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K.,
            Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences for Sequence-Ready Map Building
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
```



```
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            For clone availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .604
                     /organism="Homo sapiens"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
                     /db_xref="taxon:9606"
                     /clone="R-10F20"
                     /clone_lib="RPCI11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
BASE COUNT      240 a     68 c     78 g    218 t
ORIGIN

  Query Match           10.5%;  Score 18;  DB 26;  Length 604;
  Best Local Similarity 100.0%;  Pred. No. 4.92e-05;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    366 TACATGAAATACATCCCC 383
          ||||||||||||||||||
Cp     89 TACATGAAATACATCCCC 72


RESULT   40
LOCUS       AQ017441       680 bp    DNA             GSS       08-JUN-1998
DEFINITION  CIT-HSP-2306F1.TF CIT-HSP Homo sapiens genomic clone 2306F1,
            genomic survey sequence.
ACCESSION   AQ017441
NID         g3196177
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 680)
  AUTHORS   Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K.,
            Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
            Simon,M. and Venter,J.C.
  TITLE     Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: CIT-HSP-2306F1.TR
            Contact: Mark Adams
            Department of Eukaryotic Genomics
```



```
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13-21
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .680
                     /organism="Homo sapiens"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII"
                     /db_xref="taxon:9606"
                     /clone="2306F1"
                     /clone_lib="CIT-HSP"
                     /sex="Male"
                     /cell_type="Sperm"
BASE COUNT      279 a     81 c     88 g    232 t
ORIGIN

  Query Match           10.5%;  Score 18;  DB 26;  Length 680;
  Best Local Similarity 100.0%;  Pred. No. 4.92e-05;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    250 AAATAAATAATAAGTGAA 267
          ||||||||||||||||||
Qy    154 AAATAAATAATAAGTGAA 171


RESULT   41
LOCUS       B13244        1148 bp    DNA             GSS       13-MAY-1997
DEFINITION  T18C16-T7 TAMU Arabidopsis thaliana genomic clone T18C16, genomic
            survey sequence.
ACCESSION   B13244
NID         g2088986
KEYWORDS    GSS.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryotae; mitochondrial eukaryotes; Viridiplantae;
            Charophyta/Embryophyta group; Embryophyta; vascular plants; seed
            plants; Magnoliophyta; Magnoliopsida; Capparales; Brassicaceae;
            Arabidopsis.
REFERENCE   1  (bases 1 to 1148)
  AUTHORS   Feng,J., Dewar,K., Buehler,E., Kim,C., Li,Y., Shinn,P., Sun,H. and
            Ecker,J.
  TITLE     BAC End Sequences at ATGC
  JOURNAL   Unpublished (1997)
COMMENT
```

```
            Contact: Ecker J.
            Arabidopsis Thaliana Genome Center
            University of Pennsylvania
            Dept. of Biology, University of Pennsylvania, Philadelphia, PA
            19104
            Tel: 215-898-9384
            Fax: 215-898-8780
            Email: jecker@atgenome.bio.upenn.edu
            Seq primer: T7
            Class: BAC ends
            High quality sequence start: 73
            High quality sequence stop: 657.
FEATURES            Location/Qualifiers
     source         1. .1148
                    /organism="Arabidopsis thaliana"
                    /strain="Columbia"
                    /note="Vector: BeloBACII; Site_1: HindIII; Site_2:
                    HindIII; Produced by Rod Wing"
                    /db_xref="taxon:3702"
                    /clone="T18C16"
                    /clone_lib="TAMU"
                    /sex="hermaphrodite"
BASE COUNT      208 a     347 c     208 g     370 t      15 others
ORIGIN

  Query Match           10.5%;  Score 18;  DB 26;  Length 1148;
  Best Local Similarity 100.0%;  Pred. No. 4.92e-05;
  Matches    18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    780 TTCCCCTCCTCCTGTCGT 797
          ||||||||||||||||||
Qy     38 TTCCCCTCCTCCTGTCGT 55


RESULT   42
LOCUS       AA259656      225 bp    mRNA            EST       18-MAR-1997
DEFINITION  va36a09.r1 Soares mouse 3NME12 5 Mus musculus cDNA clone 733432 5',
            mRNA sequence.
ACCESSION   AA259656
NID         g1896177
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Mus.
REFERENCE   1  (bases 1 to 225)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
```



```
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:450480
            Putative full length read
            vector to vector length is 226
            Seq primer: -28m13 rev2 ET from Amersham.
FEATURES            Location/Qualifiers
     source         1. .225
                    /organism="Mus musculus"
                    /strain="C57BL/6J"
                    /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                    polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                    was primed with a Not I - oligo(dT) primer [5'
                    TGTTACCAATCTGAAGTGGGAGCGCGCCGCCTTATTTTTTTTTTTTTTTT 3'],
                    on total  mouse RNA [provided by Minoru Ko, Wayne State
                    Univ.]; double-stranded cDNA was ligated to Eco RI
                    adaptors (Pharmacia), digested with Not I and cloned into
                    the Not I and Eco RI sites of the modified pT7T3 vector.
                    Library went through one round of normalization, and was
                    constructed by Bento Soares and M. Fatima Bonaldo."
                    /db_xref="taxon:10090"
                    /clone="733432"
                    /clone_lib="Soares mouse 3NME12 5"
                    /sex="unknown"
                    /dev_stage="12.5dpc total fetus"
                    /lab_host="DH10B"
     mRNA           <1. .>225
BASE COUNT       55 a      47 c      52 g      71 t
ORIGIN

  Query Match           9.9%;  Score 17;  DB 23;  Length 225;
  Best Local Similarity 100.0%;  Pred. No. 1.65e-03;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    204 AATAAATAATAAGTGAA 220
          |||||||||||||||||
Qy    155 AATAAATAATAAGTGAA 171


RESULT   43
```



```
LOCUS       AA098029      273 bp    mRNA            EST       15-FEB-1997
DEFINITION  mn82b06.r1 Stratagene mouse Tcell 937311 Mus musculus cDNA clone
            550547 5', mRNA sequence.
ACCESSION   AA098029
NID         g1643582
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Mus.
REFERENCE   1  (bases 1 to 273)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:331339
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 237.
FEATURES            Location/Qualifiers
     source         1. .273
                    /organism="Mus musculus"
                    /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                    XhoI; Cloned unidirectionally.  Primer: Oligo dT. M30 CD4+
                    cells. Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'
                    adaptor sequence: 5' GAATTCGGCACGAG 3' -3' adaptor
                    sequence: 5' CTCGAGTTTTTTTTTTTTTTTTT 3'"
                    /db_xref="taxon:10090"
                    /clone="550547"
                    /clone_lib="Stratagene mouse Tcell 937311"
                    /dev_stage="M30 CD4+ cells"
                    /lab_host="SOLR (kanamycin resistant)"
     mRNA           <1. .>273
BASE COUNT       92 a      51 c      48 g      82 t
ORIGIN

  Query Match           9.9%;  Score 17;  DB 23;  Length 273;
  Best Local Similarity 100.0%;  Pred. No. 1.65e-03;
```



```
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    114 TAAAATAAATAATAAGT 130
          |||||||||||||||||
Qy    152 TAAAATAAATAATAAGT 168


RESULT   44
LOCUS       T36476        301 bp    mRNA            EST       11-JAN-1995
DEFINITION  EST101417 Saccharomyces cerevisiae cDNA 5' end.
ACCESSION   T36476
NID         g620293
KEYWORDS    EST.
SOURCE      baker's yeast library=S. cerevisiae strain X2180-1A strain=X2180-1A
            vector=pBluescript SK- primer=M13 Reverse Rsite1=EcoRI Rsite2=XhoI.
  ORGANISM  Saccharomyces cerevisiae
            Eucaryotae; Eumycota; Ascomycotina; Hemiascomycetes; Endomycetales;
            Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 301)
  AUTHORS   Weinstock,K.
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Weinstock,K. and Venter,J.C.
            The Institute for Genomic Research
            932 Clopper Rd, Gaithersburg, MD 20878
            Tel: 3018699056
            Fax: 3018699423
            Email: tdbinfo@tdb.tigr.org
            For clone availability please contact the TIGR Database
            (tdbinfo@tdb.tigr.org).
FEATURES            Location/Qualifiers
     source         1. .301
                    /organism="Saccharomyces cerevisiae"
                    /strain="X2180-1A"
BASE COUNT       97 a      57 c      53 g      90 t       4 others
ORIGIN

  Query Match           9.9%;  Score 17;  DB 5;  Length 301;
  Best Local Similarity 89.5%;  Pred. No. 1.65e-03;
  Matches    17;  Conservative    0;  Mismatches   2;  Indels   0;  Gaps   0;

Db    185 ATNCAAAGAGTTATAGCAG 203
          || |||||||| |||||||
Cp    123 ATCCAAAGAGTNATAGCAG 105


RESULT   45
LOCUS       AA476276      316 bp    mRNA            EST       08-AUG-1997
DEFINITION  zw29a11.s1 Soares ovary tumor NbHOT Homo sapiens cDNA clone 770684
            3', mRNA sequence.
ACCESSION   AA476276
```

```
NID         g2204487
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 316)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J.,
            Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-Merck EST Project 1997
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -41m13 fwd. ET from Amersham
            High quality sequence stop: 287.
FEATURES              Location/Qualifiers
     source           1..316
                      /organism="Homo sapiens"
                      /note="Organ: ovary; Vector: pT7T3D (Pharmacia) with a
                      modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                      strand cDNA was primed with a Not I - oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCGGTTTTTTTTTTTTTTTTT 3'],
                      double-stranded cDNA was size selected, ligated to Eco RI
                      adapters (Pharmacia), digested with Not I and cloned into
                      the Not I and Eco RI sites of a modified pT7T3 vector
                      (Pharmacia). Library constructed by Bento Soares and
                      M.Fatima Bonaldo."
                      /db_xref="taxon:9606"
                      /clone="770684"
                      /clone_lib="Soares ovary tumor NbHOT"
                      /sex="Female"
                      /tissue_type="ovarian tumor"
                      /lab_host="DH10B (ampicillin resistant)"
     mRNA             complement(<1..>316)
                      /db_xref="GDB:5980557"
BASE COUNT       84 a      78 c      74 g      80 t
ORIGIN

  Query Match             9.9%;  Score 17;  DB 8;  Length 316;
  Best Local Similarity 100.0%;  Pred. No. 1.65e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```



```
Db      85 CAGACTCCCATCCAAAG 101
           |||||||||||||||||
Cp     132 CAGACTCCCATCCAAAG 116


RESULT   46
LOCUS       AA686340       318 bp    mRNA            EST       03-APR-1998
DEFINITION  EST109943 Rat PC-12 cells, NGF-treated (9 days) Rattus sp. cDNA
            clone RPNAX46 5' end, mRNA sequence.
ACCESSION   AA686340
NID         g2672938
KEYWORDS    EST.
SOURCE      Rattus sp.
  ORGANISM  Rattus sp.
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 318)
  AUTHORS   Lee,N.H., Weinstock,K.G., Kirkness,E.F., Earle-Hughes,J.A.,
            Fuldner,R.A., Marmaras,S., Glodek,A., Gocayne,J.D., Adams,M.D.,
            Kerlavage,A.R., Fraser,C.M. and Venter,J.C.
  TITLE     Comparative expressed-sequence-tag analysis of differential gene
            expression profiles in PC-12 cells before and after nerve growth
            factor treatment
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 92, 8303-8307 (1995)
  MEDLINE   95396786
COMMENT
            Contact: Lee, NH
            ATCC
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
            Seq primer: M13 Reverse.
FEATURES              Location/Qualifiers
     source           1..318
                      /organism="Rattus sp."
                      /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                      XhoI; poly(A)+ RNA was purified from 9-day NGF treated
                      PC12 cells. cDNA was constructed using an oligo-dT primer
                      and directionally cloned using the Lambda ZAP II Vector
                      Kit by Stratagene"
                      /db_xref="ATCC (inhost):2004176"
                      /db_xref="taxon:10118"
                      /clone="RPNAX46"
                      /clone_lib="Rat PC-12 cells, NGF-treated (9 days)"
BASE COUNT       98 a      39 c      85 g      93 t       3 others
ORIGIN

  Query Match             9.9%;  Score 17;  DB 14;  Length 318;
```



```
  Best Local Similarity 100.0%;  Pred. No. 1.65e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      54 TTATTTATTTTATCTTC 70
           |||||||||||||||||
Cp     163 TTATTTATTTTATCTTC 147


RESULT   47
LOCUS       AA270724       329 bp    mRNA            EST       26-MAR-1997
DEFINITION  va67e04.r1 Soares mouse 3NME12 5 Mus musculus cDNA clone 736446 5',
            mRNA sequence.
ACCESSION   AA270724
NID         g1909056
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Mus.
REFERENCE   1  (bases 1 to 329)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:453494
            Putative full length read
            vector to vector length is 330
            Seq primer: -28m13 rev2 ET from Amersham.
FEATURES              Location/Qualifiers
     source           1..329
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                      was primed with a Not I - oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCCTTATTTTTTTTTTTTTTTT 3'],
                      on total  mouse RNA [provided by Minoru Ko, Wayne State
```



```
                      Univ.]; double-stranded cDNA was ligated to Eco RI
                      adaptors (Pharmacia), digested with Not I and cloned into
                      the Not I and Eco RI sites of the modified pT7T3 vector.
                      Library went through one round of normalization, and was
                      constructed by Bento Soares and M. Fatima Bonaldo."
                      /db_xref="taxon:10090"
                      /clone="736446"
                      /clone_lib="Soares mouse 3NME12 5"
                      /sex="unknown"
                      /dev_stage="12.5dpc total fetus"
                      /lab_host="DH10B"
     mRNA             <1..>329
BASE COUNT       66 a      95 c      69 g      99 t
ORIGIN

  Query Match             9.9%;  Score 17;  DB 24;  Length 329;
  Best Local Similarity 100.0%;  Pred. No. 1.65e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     308 AATAAATAATAAGTGAA 324
           |||||||||||||||||
Qy     155 AATAAATAATAAGTGAA 171


RESULT   48
LOCUS       C46755         360 bp    mRNA            EST       29-AUG-1997
DEFINITION  C.elegans cDNA clone yk424c5 : 5' end, single read, mRNA sequence.
ACCESSION   C46755
NID         g2383008
KEYWORDS    EST; EST(expressed sequence tag).
SOURCE      Caenorhabditis elegans (strain:N2) embryo hermaphrodite embryo cDNA
            to mRNA, clone_lib:Yuji Kohara unpublished cDNA library
            clone:yk424c5.
  ORGANISM  Caenorhabditis elegans
            Eukaryotae; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (sites)
  AUTHORS   Kohara,Y., Motohashi,T., Tabara,H., Shin-i,T., Watanabe,H.,
            Sano,M., Miyata,A., Ohba,T., Mitani,Y., Uesugi,H., Sugiura,I.,
            Obara,M., Sugimoto,A., Iida,K. and Nishigaki,A.
  TITLE     Expression map of the C.elegans genome
  JOURNAL   Unpublished (1997)
REFERENCE   2  (bases 1 to 360)
  AUTHORS   Kohara,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1997) to the DDBJ/EMBL/GenBank databases. Yuji
            Kohara, National Institute of Genetics, Gene Network Lab; Yata
            1111, Mishima, Shizuoka 411, Japan (E-mail:ykohara@lab.nig.ac.jp,
            Tel:+81-559-81-6854, Fax:+81-559-81-6855)
FEATURES              Location/Qualifiers
     source           1..360
```

```
                     /organism="Caenorhabditis elegans"
                     /strain="N2"
                     /db_xref="taxon:6239"
                     /clone="yk424c5"
                     /clone_lib="Yuji Kohara unpublished cDNA library"
                     /dev_stage="embryo"
                     /sex="hermaphrodite"
                     /tissue_type="embryo"
BASE COUNT       97 a      85 c      76 g     102 t
ORIGIN

  Query Match           9.9%;  Score 17;  DB 9;  Length 360;
  Best Local Similarity 100.0%;  Pred. No. 1.65e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      97 TTTCACTTATTATTTAT 113
           |||||||||||||||||
Cp     172 TTTCACTTATTATTTAT 156


RESULT   49
LOCUS       R11080         373 bp    mRNA             EST       11-APR-1995
DEFINITION  yf39b03.r1 Homo sapiens cDNA clone 129197 5' similar to contains
            Alu repetitive element;.
ACCESSION   R11080
NID         g763815
KEYWORDS    EST.
SOURCE      human clone=129197 library=Soares fetal liver spleen 1NFLS
            vector=pT7T3D (Pharmacia) with a modified polylinker host=DH10B
            (ampicillin resistant) primer=M13RP1 Rsitel=Pac I Rsite2=Eco RI
            Liver and spleen from a 20 week-post conception male fetus. 1st
            strand cDNA was primed with a Pac I - oligo(dT) primer [5'
            AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTT 3'], double-stranded
            cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac
            I and cloned into the Pac I and Eco RI sites of the modified pT7T3
            vector.  Library went through one round of normalization. Library
            constructed by Bento Soares and M.Fatima Bonaldo.
  ORGANISM  Homo sapiens
            Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 373)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
```



```
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 345
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES             Location/Qualifiers
     source          1. .373
                     /organism="Homo sapiens"
                     /clone="129197"
BASE COUNT       90 a      89 c      85 g     106 t      3 others
ORIGIN

  Query Match           9.9%;  Score 17;  DB 5;  Length 373;
  Best Local Similarity 100.0%;  Pred. No. 1.65e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      30 CAGACTCCCATCCAAAG 46
           |||||||||||||||||
Cp     132 CAGACTCCCATCCAAAG 116


RESULT   50
LOCUS       AQ238631       393 bp    DNA              GSS       29-SEP-1998
DEFINITION  RPCI11-71A5.TK RPCI11 Homo sapiens genomic clone R-71A5, genomic
            survey sequence.
ACCESSION   AQ238631
NID         g3670922
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 393)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
            Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
  TITLE     Use of human BAC End Sequences for Sequence-Ready Map Building
  JOURNAL   Unpublished (1998)
COMMENT
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            For clone availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
```



```
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .393
                     /organism="Homo sapiens"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
                     /db_xref="taxon:9606"
                     /clone="R-71A5"
                     /clone_lib="RPCI11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
BASE COUNT      110 a      88 c      83 g     111 t      1 others
ORIGIN

  Query Match           9.9%;  Score 17;  DB 29;  Length 393;
  Best Local Similarity 100.0%;  Pred. No. 1.65e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     155 GGGAGTCTGGCAAGAGG 171
           |||||||||||||||||
Qy     124 GGGAGTCTGGCAAGAGG 140


Search completed: Tue Dec 22 21:54:14 1998
Job time : 382 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Tue Dec 22 21:54:33 1998; MasPar time 39.27 Seconds
595.902 Million cell updates/sec

Tabular output not generated.

Title: >US-08-530-112A-2965
Description: (1-172) from US08530112A.seq
Perfect Score: 171
N.A. Sequence: 1 GATCTGAACGATGCCGTGTC.....AAAATAAATAATAAGTGAAA 172
Comp: CTAGACTTGCTACGGCACAG.....TTTTATTTATTATTCACTTT

Scoring table: TABLE jmetric
Gap 60

Kmatch STD : Dbase 0; Query 0

Searched: 188442 seqs, 68026449 bases x 2

Post-processing: Minimum Match 0%
Listing first 1000 summaries

Database: n-geneseq32
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39 40:part40

Statistics: Mean 6.413; Variance 2.841; scale 2.258

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
	1	171	100.0	172	20	T21965	Human gene signature	2.85e-138
c	2	17	9.9	3331	2	N91248	Sequence of pLH1 or p	1.26e+00
c	3	17	9.9	6295	15	Q98818	Arabidopsis CTR1-2 mu	1.26e+00
c	4	17	9.9	6295	30	T67128	Constitutive triple r	1.26e+00
c	5	17	9.9	6312	30	T67129	Constitutive triple r	1.26e+00
c	6	17	9.9	6312	15	Q98821	Arabidopsis CTR1-4 mu	1.26e+00
c	7	17	9.9	6312	15	Q98819	Arabidopsis CTR1-3 mu	1.26e+00
c	8	17	9.9	6312	30	T67127	Constitutive triple r	1.26e+00
c	9	17	9.9	6312	30	T67131	Constitutive triple r	1.26e+00
c	10	17	9.9	6312	30	T67130	Constitutive triple r	1.26e+00
c	11	17	9.9	6312	15	Q98820	Arabidopsis CTR1-1 mu	1.26e+00
c	12	17	9.9	6312	15	Q98817	Arabidopsis CTR1 geno	1.26e+00
c	13	16	9.4	1234	12	Q70415	Full length pre-pro-v	5.49e+00
	14	16	9.4	1542	1	N90025	DNA encoding antigeni	5.49e+00
	15	16	9.4	2344	36	T91638	Castorbean ricin gene	5.49e+00
c	16	16	9.4	2475	1	Q004307	Coding region of huma	5.49e+00
c	17	16	9.4	2475	30	T61551	Human interleukin-4 r	5.49e+00
	18	16	9.4	4698	1	N90844	Tobacco basic beta-1,	5.49e+00
	19	16	9.4	4699	1	Q03254	Sequence of clone BSG	5.49e+00
c	20	16	9.4	7753	4	Q26727	Encodes exons XV to X	5.49e+00
c	21	15	8.8	15	29	T56308	Mouse TNF-a hammerhea	2.28e+01
	22	15	8.8	50	14	Q88983	VEGF 2'-NH2-RNA nucle	2.28e+01
	23	15	8.8	247	20	T23517	Human gene signature	2.28e+01
	24	15	8.8	315	19	T20096	Human gene signature	2.28e+01
c	25	15	8.8	648	16	Q91404	Human olfactory recep	2.28e+01
c	26	15	8.8	648	16	Q91405	Human olfactory recep	2.28e+01
c	27	15	8.8	648	16	Q91406	Human olfactory recep	2.28e+01
c	28	15	8.8	1400	23	T29191	Human urokinase plasm	2.28e+01
c	29	15	8.8	1400	1	Q06354	Recombinant urokinase	2.28e+01
c	30	15	8.8	1407	36	T89675	Human phospholipase i	2.28e+01
c	31	15	8.8	1407	12	Q71647	Human phospholipase i	2.28e+01
c	32	15	8.8	1407	2	N60374	Sequence encoding hum	2.28e+01
c	33	15	8.8	1407	9	Q52515	Sequence encoding hum	2.28e+01
c	34	15	8.8	1644	17	T15423	Mouse tumour necrosis	2.28e+01
c	35	15	8.8	1644	2	N60445	Sequence encoding tum	2.28e+01
c	36	15	8.8	1685	9	Q55058	Sequence encoding ami	2.28e+01
c	37	15	8.8	1735	36	T77840	Human melanoma associ	2.28e+01
c	38	15	8.8	1827	39	V15514	Hamster oral papillom	2.28e+01
c	39	15	8.8	1849	15	Q98564	N. crassa cellobiohyd	2.28e+01
	40	15	8.8	2444	1	N81254	cDNA sequence in plas	2.28e+01
	41	15	8.8	2454	10	Q58987	Human PDI.	2.28e+01
	42	15	8.8	2454	5	Q29618	PDI.	2.28e+01

	43	15	8.8	2808	5	Q32665	Nearly complete p107	2.28e+01
c	44	15	8.8	2828	3	Q15171	Bacillus thuringiensi	2.28e+01
c	45	15	8.8	2924	11	Q63261	CD26 cDNA clone.	2.28e+01
c	46	15	8.8	2924	8	Q46089	Sequence of human CD2	2.28e+01
	47	15	8.8	3077	1	N90068	Ricin D gene	2.28e+01
c	48	15	8.8	3500	1	N90229	Mouse tumour necrosis	2.28e+01
	49	15	8.8	6854	15	Q99852	Yeast TIH2 gene.	2.28e+01
c	50	15	8.8	7647	39	V15519	Hamster oral papillom	2.28e+01
c	51	15	8.8	10684	24	T33758	Control region isolat	2.28e+01
c	52	15	8.8	13633	7	Q40862	VAC-beta (genomic - p	2.28e+01
c	53	15	8.8	134525	1	Q04525	Total base sequence o	2.28e+01
c	54	15	8.8	580073	27	T58840	Mycoplasma genitalium	2.28e+01
c	55	14	8.2	15	29	T56310	Mouse TNF-a hammerhea	8.96e+01
c	56	14	8.2	117	22	T25509	Human gene signature	8.96e+01
c	57	14	8.2	287	19	T23809	Human gene signature	8.96e+01
	58	14	8.2	337	8	Q59409	Human brain Expressed	8.96e+01
	59	14	8.2	353	19	T20001	Human gene signature	8.96e+01
c	60	14	8.2	427	35	T65068	Canine genomic micros	8.96e+01
c	61	14	8.2	482	8	Q59560	Human brain Expressed	8.96e+01
	62	14	8.2	573	22	T18413	Flea saliva protein (	8.96e+01
	63	14	8.2	591	22	T18415	Flea saliva protein (	8.96e+01
c	64	14	8.2	612	32	T58248	5' end of CH14-2a16-1	8.96e+01
c	65	14	8.2	827	29	T47669	Human G protein gamma	8.96e+01
c	66	14	8.2	867	13	Q74397	Isoform lg12 of the g	8.96e+01
	67	14	8.2	1007	37	T92823	Flea saliva protein f	8.96e+01
	68	14	8.2	1109	9	Q49736	Pcp A gene.	8.96e+01
	69	14	8.2	1109	21	T34183	PcpA coding sequence.	8.96e+01
	70	14	8.2	1150	12	Q70736	TATA-binding protein-	8.96e+01
	71	14	8.2	1183	25	T42222	Human TATA-binding pr	8.96e+01
	72	14	8.2	1183	32	T79592	TATA-binding protein	8.96e+01
c	73	14	8.2	1259	18	T11026	DNA encoding the mous	8.96e+01
c	74	14	8.2	1284	3	Q15361	NF-YB.	8.96e+01
	75	14	8.2	1406	11	Q64831	Human brain serotonin	8.96e+01
	76	14	8.2	1461	1	N91621	Beta-mannase gene of	8.96e+01
	77	14	8.2	1470	30	T72063	Sequence flanking mar	8.96e+01
	78	14	8.2	1470	30	T43943	Sequence flanking mar	8.96e+01
	79	14	8.2	1618	17	T05503	Leishmania sp. antige	8.96e+01
	80	14	8.2	1618	30	T62139	Leishmania braziliens	8.96e+01
	81	14	8.2	1618	32	T80384	DNA encoding LbeIF4A.	8.96e+01
c	82	14	8.2	1685	10	Q54819	Oat globin A2B coding	8.96e+01
c	83	14	8.2	1704	20	T03694	Human alpha 2,8-sialy	8.96e+01
c	84	14	8.2	1723	37	T91745	Tobacco calcium/calmo	8.96e+01
c	85	14	8.2	1810	2	Q13573	E75B specific exon B1	8.96e+01
	86	14	8.2	1866	29	T41492	DNA encoding MAGE-Xp1	8.96e+01
c	87	14	8.2	2009	7	Q40109	Enocdes a mutant muco	8.96e+01
c	88	14	8.2	2009	1	N81804	Gene encoding Mucor r	8.96e+01
c	89	14	8.2	2021	32	T58247	CH14-2a16-1, over abu	8.96e+01
c	90	14	8.2	2094	37	T91356	Staphylococcus aureus	8.96e+01
c	91	14	8.2	2117	13	Q77831	Human alpha-2,8-sialy	8.96e+01
c	92	14	8.2	2430	20	T08125	Human syndecan-1 gene	8.96e+01
	93	14	8.2	2492	37	T89890	Mouse frizzled gene 6	8.96e+01

c	94	14	8.2	2607	6	Q31985	Pisium sativum L PAL	8.96e+01
	95	14	8.2	2809	15	Q91287	ORC2 gene encoding a	8.96e+01
	96	14	8.2	2809	31	T73280	S. cerevisiae origin	8.96e+01
	97	14	8.2	2829	38	T85374	Tetrahymena 95 kD tel	8.96e+01
	98	14	8.2	2829	25	T30868	95 kD protein coding	8.96e+01
	99	14	8.2	2861	24	T28639	RNase L inhibitor *H2	8.96e+01
	100	14	8.2	2930	25	T42174	Maize Id gene.	8.96e+01
	101	14	8.2	2945	6	Q38591	Plasmid J6B7.	8.96e+01
	102	14	8.2	3337	23	T34620	P. vivax ESP-1 blood	8.96e+01
	103	14	8.2	3425	16	T01485	Tobacco root specific	8.96e+01
	104	14	8.2	3426	7	Q38965	Sequence of TobRB7-5A	8.96e+01
	105	14	8.2	3426	12	Q67725	Nematode-inducible tr	8.96e+01
	106	14	8.2	3568	24	T28638	RNase L inhibitor *RL	8.96e+01
c	107	14	8.2	3812	1	N92734	Partial nucleotide se	8.96e+01
c	108	14	8.2	3913	17	T05573	MEKK4 cDNA.	8.96e+01
	109	14	8.2	4157	40	V19709	5-aminolevulinic acid	8.96e+01
	110	14	8.2	4157	40	V04102	Aspergillus oryzae 5-	8.96e+01
c	111	14	8.2	4415	7	Q46293	Phenylalanine ammonia	8.96e+01
c	112	14	8.2	4592	17	T05574	MEKK5 cDNA.	8.96e+01
	113	14	8.2	4796	38	T76897	Cauliflower floral me	8.96e+01
c	114	14	8.2	4887	7	Q41290	Ubiquitin-specific pr	8.96e+01
	115	14	8.2	5253	37	T64877	Neuroblastoma indicat	8.96e+01
	116	14	8.2	5359	17	T12251	Cytochrome P450 isoen	8.96e+01
c	117	14	8.2	5458	39	V03674	Human Jagged gene.	8.96e+01
	118	14	8.2	5534	6	Q55988	Tomato hap80 genomic	8.96e+01
	119	14	8.2	5595	37	T64850	Microtubule-associate	8.96e+01
	120	14	8.2	5778	3	N40033	Sequence of portion o	8.96e+01
	121	14	8.2	5810	37	T91323	Arabidopsis thaliana	8.96e+01
	122	14	8.2	5924	13	Q77790	Cobra CVF1 coding seq	8.96e+01
	123	14	8.2	6265	8	Q49460	Urea amidolyase gene.	8.96e+01
c	124	14	8.2	6464	24	T40090	Human Serrate-1 (HJ1)	8.96e+01
c	125	14	8.2	8022	29	T61982	Yeast rad3 gene.	8.96e+01
	126	14	8.2	9048	25	T43225	Brassica napus PCA ge	8.96e+01
	127	14	8.2	9299	29	T47716	Mouse bone morphogene	8.96e+01
	128	14	8.2	9601	1	N80860	Sequence corresp. to	8.96e+01
	129	14	8.2	13585	17	T11549	Tumour rejection anti	8.96e+01
	130	14	8.2	15464	2	Q11415	Ryanodine receptor ge	8.96e+01
	131	14	8.2	17350	12	Q72708	hOP1 human osteogenic	8.96e+01
	132	14	8.2	17410	9	Q53142	Sequence encoding ost	8.96e+01
	133	14	8.2	17410	18	T18381	hOP-1 genomic DNA.	8.96e+01
	134	14	8.2	17410	40	V15205	Human osteogenic prot	8.96e+01
	135	14	8.2	17415	17	T11639	Human osteogenic prot	8.96e+01
c	136	14	8.2	19124	32	T72882	Plasmodium var-7 gene	8.96e+01
	137	14	8.2	40352	40	V02032	MAGE-B cluster DNA se	8.96e+01
	138	14	8.2	580073	27	T58840	Mycoplasma genitalium	8.96e+01
c	139	13	7.6	15	29	T56326	Mouse TNF-a hammerhea	3.31e+02
c	140	13	7.6	15	26	T55797	Human TNF-alpha hamme	3.31e+02
c	141	13	7.6	15	26	T55813	Human TNF-alpha hamme	3.31e+02
c	142	13	7.6	15	26	T55815	Human TNF-alpha hamme	3.31e+02
c	143	13	7.6	15	26	T55799	Human TNF-alpha hamme	3.31e+02
c	144	13	7.6	15	29	T56332	Mouse TNF-a hammerhea	3.31e+02

c	145	13	7.6	24	4	Q26842	Primer 1830-N; mutati	3.31e+02
	146	13	7.6	28	40	V09456	CpG-containing methyl	3.31e+02
c	147	13	7.6	28	40	V09560	MSP amplification usi	3.31e+02
	148	13	7.6	29	39	V13308	Primer BF22 for B. ga	3.31e+02
c	149	13	7.6	29	22	T29044	Beta-1,3-glucanase 5'	3.31e+02
	150	13	7.6	30	40	V09458	CpG-containing unmeth	3.31e+02
c	151	13	7.6	30	40	V09562	MSP amplification usi	3.31e+02
c	152	13	7.6	34	10	Q57179	Enzymatic RNA molecul	3.31e+02
	153	13	7.6	36	20	Q90688	B. burgdorferi strain	3.31e+02
c	154	13	7.6	40	25	T38121	HLA-DRB1 gene PCR pri	3.31e+02
c	155	13	7.6	98	22	T24874	Human gene signature	3.31e+02
c	156	13	7.6	118	20	T23670	Human gene signature	3.31e+02
	157	13	7.6	126	2	Q12037	tPA-Factor VIII:C 80	3.31e+02
	158	13	7.6	126	32	T36279	pSVF9-80S in the regi	3.31e+02
	159	13	7.6	126	19	T20904	Human gene signature	3.31e+02
	160	13	7.6	130	21	T17892	DNA encoding amphiphi	3.31e+02
	161	13	7.6	135	2	Q12040	alpha-1-AT-Factor VII	3.31e+02
	162	13	7.6	135	32	T36282	pSVF8-80AT in the reg	3.31e+02
c	163	13	7.6	161	10	Q76945	Human genome fragment	3.31e+02
	164	13	7.6	225	8	Q60280	Human brain Expressed	3.31e+02
c	165	13	7.6	236	4	Q23866	Heat-stable toxin (ST	3.31e+02
	166	13	7.6	241	21	T23372	Human gene signature	3.31e+02
c	167	13	7.6	243	18	T19208	Human gene signature	3.31e+02
c	168	13	7.6	256	19	T20086	Human gene signature	3.31e+02
	169	13	7.6	276	22	T26052	Human gene signature	3.31e+02
	170	13	7.6	278	6	Q39773	Expressed Sequence Ta	3.31e+02
	171	13	7.6	278	8	Q59185	Human brain Expressed	3.31e+02
	172	13	7.6	290	19	T21329	Human gene signature	3.31e+02
	173	13	7.6	300	3	Q20064	P. carinii DNA from i	3.31e+02
c	174	13	7.6	305	10	Q55239	Grapevine ribosomal c	3.31e+02
	175	13	7.6	308	6	Q39902	Expressed Sequence Ta	3.31e+02
	176	13	7.6	308	8	Q59314	Human brain Expressed	3.31e+02
	177	13	7.6	309	8	Q60859	Human brain Expressed	3.31e+02
	178	13	7.6	311	8	Q60470	Human brain Expressed	3.31e+02
	179	13	7.6	315	10	Q57462	Osteo factor protein	3.31e+02
	180	13	7.6	318	34	T79547	Human collagen 1A1 ge	3.31e+02
	181	13	7.6	366	8	Q59399	Human brain Expressed	3.31e+02
c	182	13	7.6	368	6	Q39774	Expressed Sequence Ta	3.31e+02
c	183	13	7.6	368	8	Q59186	Human brain Expressed	3.31e+02
c	184	13	7.6	423	33	T85005	Type II topoisomerase	3.31e+02
c	185	13	7.6	433	14	Q89260	E. coli stable toxin	3.31e+02
c	186	13	7.6	433	3	Q20050	Sequence of enterotox	3.31e+02
c	187	13	7.6	446	8	Q59542	Human brain Expressed	3.31e+02
	188	13	7.6	452	19	T19905	Human gene signature	3.31e+02
c	189	13	7.6	462	22	T32444	Wasp venom BrhTX-1 su	3.31e+02
	190	13	7.6	466	20	T23279	Human gene signature	3.31e+02
c	191	13	7.6	477	9	Q65518	Human post-synaptic N	3.31e+02
c	192	13	7.6	494	28	T59835	Beta-casein gene frag	3.31e+02
	193	13	7.6	495	33	T74484	Pilosulin 2 precursor	3.31e+02
c	194	13	7.6	510	18	T19485	Human gene signature	3.31e+02
	195	13	7.6	519	12	Q73905	Borrelia KL11 antigen	3.31e+02

	196	13	7.6	519	12	Q73881	Borrelia PBI antigen	3.31e+02
	197	13	7.6	522	12	Q73898	Borrelia VSDA antigen	3.31e+02
	198	13	7.6	525	12	Q73896	Borrelia M57 antigen	3.31e+02
	199	13	7.6	525	12	Q73900	Borrelia 20047 antige	3.31e+02
	200	13	7.6	528	12	Q73897	Borrelia W antigen va	3.31e+02
	201	13	7.6	531	12	Q73903	Borrelia NBS18AB antig	3.31e+02
	202	13	7.6	531	12	Q73902	Borrelia IP90 antigen	3.31e+02
	203	13	7.6	531	12	Q73901	Borrelia KL10 antigen	3.31e+02
	204	13	7.6	531	12	Q73877	Borrelia IP90 antigen	3.31e+02
	205	13	7.6	531	12	Q73878	Borrelia NSB1AB antig	3.31e+02
	206	13	7.6	531	12	Q73899	Borrelia NBS23a antig	3.31e+02
	207	13	7.6	555	34	T69667	Helix modification re	3.31e+02
c	208	13	7.6	564	30	T77406	H. pylori transmembra	3.31e+02
c	209	13	7.6	564	30	T67726	H. pylori transmembra	3.31e+02
c	210	13	7.6	567	22	T35706	Diagnostic sequence f	3.31e+02
	211	13	7.6	570	12	Q73882	Borrelia HL13 antigen	3.31e+02
	212	13	7.6	570	12	Q73880	Borrelia KL11 antigen	3.31e+02
	213	13	7.6	573	12	Q73873	Borrelia VSDA antigen	3.31e+02
	214	13	7.6	576	12	Q73871	Borrelia M57 antigen	3.31e+02
	215	13	7.6	576	12	Q73875	Borrelia 20047 antige	3.31e+02
	216	13	7.6	579	12	Q73872	Borrelia W antigen va	3.31e+02
	217	13	7.6	582	12	Q73874	Borrelia NBS23a antig	3.31e+02
	218	13	7.6	582	12	Q73876	Borrelia KL10 antigen	3.31e+02
c	219	13	7.6	594	1	N81273	Phosphinothricin resi	3.31e+02
c	220	13	7.6	605	35	T39804	Mouse p53bp2 gene.	3.31e+02
	221	13	7.6	624	20	Q90717	B. burgdorferi strain	3.31e+02
	222	13	7.6	624	14	Q82827	Vibrio cholerae TcpC	3.31e+02
	223	13	7.6	630	20	Q90715	B. burgdorferi strain	3.31e+02
	224	13	7.6	630	39	V13306	B. garinii OspC gene.	3.31e+02
	225	13	7.6	669	3	N50105	DNA encoding truncate	3.31e+02
c	226	13	7.6	675	10	Q56792	Clone of receptor for	3.31e+02
c	227	13	7.6	675	30	T67026	Beta-amyloid precurso	3.31e+02
	228	13	7.6	678	13	Q72943	Rat Erythropoietin cD	3.31e+02
c	229	13	7.6	720	30	T67027	Beta-amyloid precurso	3.31e+02
c	230	13	7.6	720	10	Q56793	Clone of receptor for	3.31e+02
c	231	13	7.6	731	5	Q31693	RsaI restriction frag	3.31e+02
	232	13	7.6	785	15	Q04721	Human arginase cDNA c	3.31e+02
c	233	13	7.6	786	22	T32568	Rapeseed microspore-s	3.31e+02
	234	13	7.6	800	2	Q12138	cDNA encoding cold ha	3.31e+02
c	235	13	7.6	804	14	Q87438	P. falciparum Pfs28 g	3.31e+02
c	236	13	7.6	809	1	N80722	Canine SP5 cDNA clone	3.31e+02
c	237	13	7.6	809	2	N70716	Sequence of a canine	3.31e+02
c	238	13	7.6	923	38	V06094	Viral infection gene	3.31e+02
	239	13	7.6	966	1	Q04714	Human arginase coding	3.31e+02
c	240	13	7.6	970	5	Q29103	spsAg gene and H6 pro	3.31e+02
c	241	13	7.6	987	1	N91186	Kluyveromyces lactis	3.31e+02
	242	13	7.6	1033	40	V05281	Sequence of the MS8-1	3.31e+02
c	243	13	7.6	1053	30	T68192	H. pylori membrane pr	3.31e+02
c	244	13	7.6	1057	40	V17148	Insecticidal toxin su	3.31e+02
c	245	13	7.6	1058	22	T32432	Wasp venom BrhTX-1 su	3.31e+02
c	246	13	7.6	1065	6	Q37895	Beta-casein coding se	3.31e+02

c	247	13	7.6	1065	6	Q38097	Sequence of human bet	3.31e+02
c	248	13	7.6	1103	39	V05139	cDNA encoding murine	3.31e+02
c	249	13	7.6	1109	3	Q20728	Murine macrophage inf	3.31e+02
c	250	13	7.6	1109	1	Q03852	Sequence encoding mac	3.31e+02
c	251	13	7.6	1109	3	Q20612	Murine macrophage inf	3.31e+02
	252	13	7.6	1120	39	V15825	cDNA for interferon-g	3.31e+02
	253	13	7.6	1120	22	T32404	Human interferon-gamm	3.31e+02
c	254	13	7.6	1125	2	N60273	Sequence encoding hum	3.31e+02
c	255	13	7.6	1127	23	T37331	Wasp venom 30 KDa ins	3.31e+02
c	256	13	7.6	1159	7	Q40592	JSC1 precursor DNA.	3.31e+02
c	257	13	7.6	1191	27	T50947	CAMP factor gene cfu.	3.31e+02
	258	13	7.6	1198	32	T47647	Pineal gland specific	3.31e+02
	259	13	7.6	1207	40	V05279	Sequence of the MS8-1	3.31e+02
c	260	13	7.6	1218	3	Q20693	Human endothelin-2 ge	3.31e+02
	261	13	7.6	1249	5	Q29964	Sugar beet beta-1,3-g	3.31e+02
c	262	13	7.6	1251	24	T46122	Wild-type DNA encodin	3.31e+02
c	263	13	7.6	1251	13	Q74775	Staphylococcus epider	3.31e+02
c	264	13	7.6	1260	4	Q24473	Precursor ET-2 cDNA s	3.31e+02
c	265	13	7.6	1275	2	N60558	Sequence encoding mat	3.31e+02
c	266	13	7.6	1299	3	Q14625	Human Glioblastoma ce	3.31e+02
c	267	13	7.6	1299	21	T34368	Plasmid pJC310 (ATCC	3.31e+02
c	268	13	7.6	1299	3	Q14628	Human Glioblastoma ce	3.31e+02
c	269	13	7.6	1323	2	N60363	Sequence encoding hum	3.31e+02
c	270	13	7.6	1332	15	Q87443	Vaccinia virus semaph	3.31e+02
	271	13	7.6	1350	19	Q94961	E. coli LysC gene.	3.31e+02
	272	13	7.6	1350	8	Q49568	AKIII gene.	3.31e+02
	273	13	7.6	1350	20	T06479	Lysine insensitive va	3.31e+02
	274	13	7.6	1350	4	Q24787	LysC gene encoding as	3.31e+02
	275	13	7.6	1365	1	Q04715	Human arginase cDNA c	3.31e+02
	276	13	7.6	1380	39	V05368	cDNA encoding human i	3.31e+02
c	277	13	7.6	1383	30	T66164	Mouse interleukin-12	3.31e+02
c	278	13	7.6	1383	6	Q38811	Heparin-binding neuri	3.31e+02
c	279	13	7.6	1393	2	N60225	Sequence of autonomou	3.31e+02
c	280	13	7.6	1437	1	N90242	Alpha-factor profibro	3.31e+02
c	281	13	7.6	1437	1	Q02840	Profibrolase construc	3.31e+02
c	282	13	7.6	1459	22	T29043	Oerskovia beta-1,3-gl	3.31e+02
	283	13	7.6	1464	13	Q79877	Human adrenodoxin.	3.31e+02
c	284	13	7.6	1506	22	T04695	Plasmid pSK1L replic	3.31e+02
	285	13	7.6	1507	4	Q27078	Fla gene of Borrelia	3.31e+02
c	286	13	7.6	1525	7	Q43204	Sequence encoding a p	3.31e+02
c	287	13	7.6	1560	1	N80219	Sequence of pE4 encod	3.31e+02
c	288	13	7.6	1577	40	V21578	Equine interleukin-1	3.31e+02
c	289	13	7.6	1585	10	Q58037	Tumor necrosis factor	3.31e+02
c	290	13	7.6	1585	2	N60527	Sequence encoding tum	3.31e+02
c	291	13	7.6	1585	2	N60557	Sequence encoding mat	3.31e+02
	292	13	7.6	1589	24	T39516	Flea calreticulin cDN	3.31e+02
c	293	13	7.6	1589	24	T39518	Flea calreticulin cDN	3.31e+02
c	294	13	7.6	1606	2	N60445	Sequence encoding tum	3.31e+02
c	295	13	7.6	1606	17	T15424	Human tumour necrosis	3.31e+02
c	296	13	7.6	1611	6	Q39042	Fc-alpha-R coding seq	3.31e+02
c	297	13	7.6	1611	2	Q12528	Thymidylate phosphohy	3.31e+02

c	298	13	7.6	1611	11	Q69159	Human Fc-alpha-R cDNA	3.31e+02
	299	13	7.6	1622	24	T37309	Aromatic acyl transfe	3.31e+02
c	300	13	7.6	1627	8	Q47761	Antisense RNA complem	3.31e+02
c	301	13	7.6	1643	22	T31021	Human tumour necrosis	3.31e+02
c	302	13	7.6	1643	2	N71307	Sequence encoding hum	3.31e+02
c	303	13	7.6	1644	38	T90166	Promoter from class 1	3.31e+02
c	304	13	7.6	1645	12	Q71210	Linoleic-acid-desatur	3.31e+02
c	305	13	7.6	1660	36	T94661	Chrysanthemum flavono	3.31e+02
c	306	13	7.6	1674	18	T17846	Canola palmitoyl-ACP-	3.31e+02
c	307	13	7.6	1674	2	Q05605	Gene encoding protein	3.31e+02
c	308	13	7.6	1678	3	Q22829	Sequence encoding ovi	3.31e+02
	309	13	7.6	1679	2	N82172	cDNA corresponding to	3.31e+02
c	310	13	7.6	1695	1	N90240	Clone SVG-51 (FIB 51)	3.31e+02
c	311	13	7.6	1695	1	Q02838	Deduced sequence of s	3.31e+02
c	312	13	7.6	1722	15	Q94598	ShET2 enterotoxin loc	3.31e+02
c	313	13	7.6	1743	36	T85939	Shigella flexneri Ipa	3.31e+02
c	314	13	7.6	1770	1	Q06266	Promoter region and s	3.31e+02
c	315	13	7.6	1818	40	V17562	Coding sequence for t	3.31e+02
	316	13	7.6	1824	37	T91793	FIN13 serine/threonin	3.31e+02
c	317	13	7.6	1894	28	T60740	Alpha-1-acid glycopro	3.31e+02
c	318	13	7.6	1902	21	T13239	Choline acetyltransfe	3.31e+02
c	319	13	7.6	1910	40	V16114	cDNA encoding a trans	3.31e+02
c	320	13	7.6	1910	11	Q63994	Human tissue transglu	3.31e+02
	321	13	7.6	1924	19	Q97804	Clone pTbeta2 coding	3.31e+02
c	322	13	7.6	1944	37	T91276	Streptococcus pneumon	3.31e+02
c	323	13	7.6	1960	1	N90669	DNA sequence encoadin	3.31e+02
c	324	13	7.6	1996	34	T90830	cDNA encoding glucoam	3.31e+02
	325	13	7.6	1997	40	V05280	Sequence of the MS8-1	3.31e+02
c	326	13	7.6	2008	15	Q94599	EIET enterotoxin enco	3.31e+02
c	327	13	7.6	2039	22	T32569	Rapeseed microspore-s	3.31e+02
	328	13	7.6	2051	34	T86174	Mouse mD52 C1 nucleot	3.31e+02
c	329	13	7.6	2084	39	T96033	Rat kidney injury rel	3.31e+02
c	330	13	7.6	2095	1	N70692	Sequence of human atr	3.31e+02
c	331	13	7.6	2097	9	Q52316	Coriolus hirsutus ARS	3.31e+02
	332	13	7.6	2099	2	Q10418	Plasmodium knowlesi 6	3.31e+02
	333	13	7.6	2102	34	T80112	Coding sequence for c	3.31e+02
	334	13	7.6	2105	10	Q61559	cDNA sequence of huma	3.31e+02
c	335	13	7.6	2117	22	T27050	STR6 (suppressor of t	3.31e+02
	336	13	7.6	2147	16	Q94682	Wild-type E.coli aspa	3.31e+02
	337	13	7.6	2147	34	T59362	DNA encoding wild typ	3.31e+02
	338	13	7.6	2147	34	T70078	DNA encoding aspartok	3.31e+02
	339	13	7.6	2147	34	T70086	DNA encoding aspartok	3.31e+02
	340	13	7.6	2147	34	T70081	DNA encoding aspartok	3.31e+02
	341	13	7.6	2147	34	T70080	DNA encoding aspartok	3.31e+02
	342	13	7.6	2147	34	T70077	DNA encoding aspartok	3.31e+02
	343	13	7.6	2147	12	Q81833	E. coli aspartokinase	3.31e+02
	344	13	7.6	2147	34	T70085	DNA encoding aspartok	3.31e+02
	345	13	7.6	2147	34	T70087	DNA encoding aspartok	3.31e+02
	346	13	7.6	2147	34	T70076	DNA encoding aspartok	3.31e+02
	347	13	7.6	2147	34	T70079	DNA encoding aspartok	3.31e+02
	348	13	7.6	2147	34	T70083	DNA encoding aspartok	3.31e+02

	349	13	7.6	2147	34	T70082	DNA encoding aspartok 3.31e+02
	350	13	7.6	2147	34	T70084	DNA encoding aspartok 3.31e+02
c	351	13	7.6	2197	16	T05684	Lymphocyte activation 3.31e+02
c	352	13	7.6	2201	1	Q06005	Sequence encoding EP2 3.31e+02
c	353	13	7.6	2218	27	T50939	Canola cold responsiv 3.31e+02
	354	13	7.6	2264	22	T36723	Rat vasopressin V1b r 3.31e+02
	355	13	7.6	2276	15	Q87170	Natural resistance-as 3.31e+02
	356	13	7.6	2280	17	T01403	Canine herpesvirus gC 3.31e+02
	357	13	7.6	2290	11	Q64901	DRRG 49 genomic clone 3.31e+02
	358	13	7.6	2294	16	Q92940	Natural resistance-as 3.31e+02
c	359	13	7.6	2296	39	V12457	Human HP4 prostagland 3.31e+02
	360	13	7.6	2390	30	T67024	Rat beta-amyloid prec 3.31e+02
c	361	13	7.6	2390	30	T67024	Rat beta-amyloid prec 3.31e+02
	362	13	7.6	2406	23	T33677	Human transcription f 3.31e+02
c	363	13	7.6	2450	27	T43378	Human cytokine respon 3.31e+02
c	364	13	7.6	2476	31	T57234	Scytalidium thermophi 3.31e+02
	365	13	7.6	2485	15	Q87167	Natural resistance-as 3.31e+02
	366	13	7.6	2497	10	Q55941	Murine leukaemia inhi 3.31e+02
	367	13	7.6	2498	4	Q25800	mLIF-R. 3.31e+02
	368	13	7.6	2498	9	Q58426	Mouse LIF-R clone 3. 3.31e+02
	369	13	7.6	2498	16	Q92271	Murine leukaemia inhi 3.31e+02
c	370	13	7.6	2540	6	Q37687	Rat choline kinase ge 3.31e+02
c	371	13	7.6	2566	39	T96032	Rat kidney injury rel 3.31e+02
	372	13	7.6	2587	17	T15761	Gibberellin (GA1) clo 3.31e+02
c	373	13	7.6	2669	28	T59599	Rat-derived oxide squ 3.31e+02
	374	13	7.6	2702	18	Q95323	GalNAc-alpha-2,6-sial 3.31e+02
c	375	13	7.6	2727	1	N91188	BamHI/SalI insert of 3.31e+02
c	376	13	7.6	2734	1	N80666	Residues 4457 to 7190 3.31e+02
c	377	13	7.6	2745	34	T90831	Glucoamylase P gene. 3.31e+02
	378	13	7.6	2750	6	Q38808	G6PD coding sequence. 3.31e+02
	379	13	7.6	2774	39	T99535	Human DNA replication 3.31e+02
c	380	13	7.6	2809	1	Q02842	pKS311 Plasmid carryi 3.31e+02
c	381	13	7.6	2809	1	N90244	Alpha-factor profibro 3.31e+02
	382	13	7.6	2820	15	Q90654	Eph-related PTK CeK8 3.31e+02
c	383	13	7.6	2820	38	T86008	Mouse beta meltrin cD 3.31e+02
c	384	13	7.6	2828	37	T89888	Mouse frizzled gene 4 3.31e+02
c	385	13	7.6	2875	29	T47713	Mouse bone morphogene 3.31e+02
	386	13	7.6	3017	8	Q49956	Impatiens Necrotic Sp 3.31e+02
	387	13	7.6	3076	1	N80063	Equine gamma interfer 3.31e+02
	388	13	7.6	3100	2	Q12137	Genomic clone encodin 3.31e+02
	389	13	7.6	3101	1	Q02047	Sequence encoding a c 3.31e+02
c	390	13	7.6	3126	15	Q90112	TCI gene. 3.31e+02
	391	13	7.6	3131	36	T91708	Gene 036 differential 3.31e+02
	392	13	7.6	3182	24	T45659	Human myotonic-protei 3.31e+02
c	393	13	7.6	3215	8	Q48933	mOSP-2. 3.31e+02
	394	13	7.6	3218	29	T60328	Beta-casein promoter. 3.31e+02
c	395	13	7.6	3257	4	Q27101	Endothelial tissue tr 3.31e+02
	396	13	7.6	3366	14	Q84803	Spinocerebellar ataxi 3.31e+02
	397	13	7.6	3372	4	Q27254	Growth Factor Recepto 3.31e+02
	398	13	7.6	3372	19	T03198	Human GRB-1 genomic D 3.31e+02
c	399	13	7.6	3466	15	Q99883	KAR3 coding sequence. 3.31e+02

c	400	13	7.6	3486	4	Q27102	Macrophage tissue tra 3.31e+02
	401	13	7.6	3519	16	Q92937	Natural resistance-as 3.31e+02
c	402	13	7.6	3519	1	Q05328	Rat tumour necrosis f 3.31e+02
	403	13	7.6	3524	35	T72108	Human semaphorin Z ge 3.31e+02
	404	13	7.6	3565	14	Q84051	Sequence encoding mel 3.31e+02
	405	13	7.6	3622	1	N90567	Ovine Y-chromosomal D 3.31e+02
	406	13	7.6	3716	38	V05649	Rice SPK gene signal 3.31e+02
	407	13	7.6	3724	11	Q66256	Sequence encoding a h 3.31e+02
	408	13	7.6	3731	5	Q32631	Porcine PAM-8. 3.31e+02
	409	13	7.6	3756	3	N50525	Bacillus thuringiensi 3.31e+02
	410	13	7.6	3852	3	N50094	DNA encoding Factor-V 3.31e+02
	411	13	7.6	3898	3	Q13968	Bcg 4-4 ACP sequence. 3.31e+02
	412	13	7.6	3898	40	V15263	Brassica campestris B 3.31e+02
c	413	13	7.6	3925	21	T13237	HindIII-BamHI fragmen 3.31e+02
c	414	13	7.6	3975	3	Q22999	SERP gene. 3.31e+02
c	415	13	7.6	3975	1	N81157	Malaria-specific gene 3.31e+02
	416	13	7.6	4010	13	Q84303	Human NF-AT120, X sub 3.31e+02
	417	13	7.6	4090	13	Q85027	Entamoeba histolytica 3.31e+02
	418	13	7.6	4165	24	T18893	Rat RER7 cDNA. 3.31e+02
	419	13	7.6	4171	18	Q74082	Oncostatin M receptor 3.31e+02
c	420	13	7.6	4213	14	Q90019	Human betaglycan cont 3.31e+02
	421	13	7.6	4272	1	N80447	Modified factor VIII: 3.31e+02
	422	13	7.6	4275	1	N80446	Modified factor VIII: 3.31e+02
	423	13	7.6	4275	1	N90654	DNA encoding 740 Arg- 3.31e+02
	424	13	7.6	4423	5	Q29470	Extracellular protein 3.31e+02
c	425	13	7.6	4440	35	T89487	Human A20 protein cDN 3.31e+02
	426	13	7.6	4545	2	Q12224	EGRF-R erbB-3 gene cl 3.31e+02
	427	13	7.6	4545	1	N80444	Modified factor VIII: 3.31e+02
	428	13	7.6	4590	2	N60472	Sequence encoding the 3.31e+02
c	429	13	7.6	4608	13	Q80196	Deleted in colorectal 3.31e+02
	430	13	7.6	4616	1	N81545	Human Factor VIII-C a 3.31e+02
	431	13	7.6	4629	13	Q76016	B-domain deleted Fact 3.31e+02
	432	13	7.6	4663	1	N80085	BamHI fragment of equ 3.31e+02
	433	13	7.6	4706	17	T10157	Vector plasmid pOnMTG 3.31e+02
	434	13	7.6	4744	17	T10155	Sockeye salmon growth 3.31e+02
	435	13	7.6	4752	10	Q58032	Tyrosinase promoter. 3.31e+02
c	436	13	7.6	4762	37	T94052	Human multidrug resis 3.31e+02
c	437	13	7.6	4821	24	T38305	Tospovirus M-RNA sequ 3.31e+02
	438	13	7.6	4830	1	N81544	Human Factor VIII-C a 3.31e+02
c	439	13	7.6	4852	12	Q70448	Human nestin gene - i 3.31e+02
c	440	13	7.6	4854	5	Q28399	Human nestin gene. 3.31e+02
	441	13	7.6	4905	18	T10059	erbB-3 cDNA clone E3- 3.31e+02
	442	13	7.6	4944	35	T91059	Human gene fohy030 co 3.31e+02
	443	13	7.6	4944	24	T41470	Human 030 gene (fohy0 3.31e+02
	444	13	7.6	4978	2	Q13353	HER-3 epithelial grow 3.31e+02
	445	13	7.6	5035	30	T69811	Factor VIII-dB695-HCI 3.31e+02
	446	13	7.6	5055	24	T41471	Human 030 gene (fohy0 3.31e+02
	447	13	7.6	5055	35	T91060	Human gene fohy030 co 3.31e+02
	448	13	7.6	5166	39	V04024	Human dendritic cell 3.31e+02
c	449	13	7.6	5186	2	Q12752	Deleted in Colorectal 3.31e+02
c	450	13	7.6	5278	37	V04446	Rat programmed cell d 3.31e+02

	451	13	7.6	5477	23	T33514	Human DEC-205 cDNA. 3.31e+02
	452	13	7.6	5518	34	T73870	Cotton fibre promoter 3.31e+02
	453	13	7.6	5547	34	T73865	Cotton fibre promoter 3.31e+02
	454	13	7.6	5995	38	T96641	Human TUB locus 5' re 3.31e+02
	455	13	7.6	6124	1	Q03568	Sequence encoding the 3.31e+02
c	456	13	7.6	6124	1	Q03568	Sequence encoding the 3.31e+02
	457	13	7.6	6136	14	Q90478	Chinook salmon growth 3.31e+02
	458	13	7.6	6300	18	T03571	Factor-VIII cDNA. 3.31e+02
c	459	13	7.6	6412	19	Q79622	Flecked Morning Glory 3.31e+02
	460	13	7.6	6464	24	T40090	Human Serrate-1 (HJ1) 3.31e+02
	461	13	7.6	6744	5	Q29471	Extracellular factor 3.31e+02
	462	13	7.6	6869	30	T47442	Saccharomyces cerevis 3.31e+02
	463	13	7.6	6905	15	Q92779	Human thymopoietin co 3.31e+02
c	464	13	7.6	6959	40	V09152	Nucleotide sequence f 3.31e+02
	465	13	7.6	7053	33	T51357	Factor VIII:C coding 3.31e+02
	466	13	7.6	7056	10	Q66615	Sequence of human fac 3.31e+02
	467	13	7.6	7059	1	N81543	Human Factor VIII-C a 3.31e+02
	468	13	7.6	7241	3	Q15140	Genomic clone LE-ACC2 3.31e+02
	469	13	7.6	7244	40	V15705	Tomato ACC synthase L 3.31e+02
	470	13	7.6	7440	2	N60689	Sequence encoding hum 3.31e+02
	471	13	7.6	7440	3	N50375	DNA sequence encoding 3.31e+02
	472	13	7.6	7791	15	Q84660	Human neuronal calciu 3.31e+02
	473	13	7.6	7808	15	Q84659	Human neuronal calciu 3.31e+02
	474	13	7.6	7997	34	T79737	Rat cholesterol 7-alp 3.31e+02
	475	13	7.6	7997	15	Q89553	Rat cholesterol 7 alp 3.31e+02
	476	13	7.6	8241	1	N81096	cDNA sequence encodin 3.31e+02
	477	13	7.6	8241	1	N81439	Factor VIII cDNA inse 3.31e+02
	478	13	7.6	8585	28	T48867	Brassica napus micros 3.31e+02
	479	13	7.6	8585	1	Q05749	Microspore-specific c 3.31e+02
	480	13	7.6	8668	9	Q50433	Partial human genomic 3.31e+02
	481	13	7.6	8738	32	T72327	Lactobacillus bacteri 3.31e+02
c	482	13	7.6	8797	27	T50856	DNA cassette for amph 3.31e+02
c	483	13	7.6	8931	35	T62940	Regulatory gene from 3.31e+02
	484	13	7.6	8967	21	T31031	Factor-VIII full-leng 3.31e+02
	485	13	7.6	8975	3	N50054	Human factor VIII gen 3.31e+02
	486	13	7.6	9009	33	T61548	Factor VIII:C (Arg1689 3.31e+02
	487	13	7.6	9009	9	Q50185	Human Factor VIII cDN 3.31e+02
	488	13	7.6	9009	40	V12112	Homo sapiens factor V 3.31e+02
	489	13	7.6	9354	37	T73164	cDNA encoding human B 3.31e+02
	490	13	7.6	9520	2	N91779	Sequence encoding mou 3.31e+02
c	491	13	7.6	10475	1	N80616	Genomic clone encodin 3.31e+02
c	492	13	7.6	10475	2	N80643	Sequence of human pul 3.31e+02
c	493	13	7.6	10607	6	Q37894	Beta-casein coding se 3.31e+02
	494	13	7.6	10614	15	Q89555	Hamster cholesterol 7 3.31e+02
	495	13	7.6	10660	14	Q84793	Spinocerebellar ataxi 3.31e+02
c	496	13	7.6	11093	27	T50855	DNA cassette for amph 3.31e+02
c	497	13	7.6	11141	20	T15820	TBE virus strain Neud 3.31e+02
	498	13	7.6	11531	9	Q54222	BSSL/CEL Gene. 3.31e+02
	499	13	7.6	12364	2	Q13547	ACVS gene pcbAB. 3.31e+02
	500	13	7.6	12364	2	Q13607	ACV synthetase gene, 3.31e+02
	501	13	7.6	12752	8	Q47380	Sequence which corres 3.31e+02

c	502	13	7.6	13206	1	Q05243	Rat acyl peptide hydr 3.31e+02
	503	13	7.6	13747	8	Q47342	Myotonic dystrophy ge 3.31e+02
c	504	13	7.6	14176	34	T84564	Swinepox virus HindII 3.31e+02
	505	13	7.6	14507	32	T73568	Expression augmenting 3.31e+02
c	506	13	7.6	15144	29	T47715	Mouse bone morphogene 3.31e+02
	507	13	7.6	15155	3	N50107	DNA encoding Factor-V 3.31e+02
c	508	13	7.6	15155	3	N50107	DNA encoding Factor-V 3.31e+02
c	509	13	7.6	15377	4	Q25975	MH mutant porcine rya 3.31e+02
c	510	13	7.6	15464	2	Q11415	Ryanodine receptor ge 3.31e+02
	511	13	7.6	15581	39	V14355	Plasmid pNLnSG11 expr 3.31e+02
c	512	13	7.6	15672	2	Q10613	Rianodin receptor gen 3.31e+02
	513	13	7.6	17041	12	Q67057	PACPAP38 DNA. 3.31e+02
	514	13	7.6	17041	3	Q21065	Genomic DNA of human 3.31e+02
c	515	13	7.6	26698	20	T08126	Mouse syndecan-1 gene 3.31e+02
c	516	13	7.6	26700	40	V15946	Mouse syndecan gene s 3.31e+02
c	517	13	7.6	26700	11	Q67902	Syndecan gene. 3.31e+02
	518	13	7.6	28994	39	V15826	Genomic DNA for inter 3.31e+02
c	519	13	7.6	30967	23	T32454	Calpain large subunit 3.31e+02
c	520	13	7.6	133894	17	T13635	AcNPV genomic DNA clo 3.31e+02
	521	13	7.6	134525	1	Q04525	Total base sequence o 3.31e+02
	522	12	7.0	15	27	T54848	Mouse relA hammerhead 1.14e+03
c	523	12	7.0	15	26	T55817	Human TNF-alpha hamme 1.14e+03
c	524	12	7.0	15	26	T55801	Human TNF-alpha hamme 1.14e+03
c	525	12	7.0	15	26	T55811	Human TNF-alpha hamme 1.14e+03
c	526	12	7.0	20	25	T72539	Primer for exon 14 of 1.14e+03
	527	12	7.0	22	30	T72009	Primer detects marker 1.14e+03
c	528	12	7.0	25	32	T75976	DEN-2 cloning/sequenc 1.14e+03
c	529	12	7.0	29	39	V00068	Human G-protein coupl 1.14e+03
	530	12	7.0	30	30	T60499	GADPH specific primer 1.14e+03
	531	12	7.0	32	8	Q47885	SSP for flavonoid-3', 1.14e+03
c	532	12	7.0	40	12	Q72317	Adenomatous polyposis 1.14e+03
c	533	12	7.0	40	37	T96174	5' intron sequence of 1.14e+03
c	534	12	7.0	40	31	T71192	HPV type 35 probe dir 1.14e+03
c	535	12	7.0	41	7	Q44000	HIV-1 LTR region anti 1.14e+03
c	536	12	7.0	43	4	N30121	Sequence of the short 1.14e+03
c	537	12	7.0	57	18	T05186	Amplification primer 1.14e+03
c	538	12	7.0	59	11	Q65336	Prostaglandin I2 prod 1.14e+03
c	539	12	7.0	64	4	N30122	Sequence of the lengt 1.14e+03
c	540	12	7.0	82	13	Q74767	Human prepro-erythrop 1.14e+03
c	541	12	7.0	87	7	Q43995	HIV-1 LTR region trip 1.14e+03
	542	12	7.0	106	19	T20928	Human gene signature 1.14e+03
	543	12	7.0	124	19	T20911	Human gene signature 1.14e+03
	544	12	7.0	135	18	Q96011	Partial MAP kinase ph 1.14e+03
	545	12	7.0	159	19	T20251	Human gene signature 1.14e+03
	546	12	7.0	172	8	Q61227	Human brain Expressed 1.14e+03
c	547	12	7.0	177	26	T40821	Serine protease nfSP3 1.14e+03
c	548	12	7.0	179	19	T20211	Human gene signature 1.14e+03
	549	12	7.0	201	10	Q57420	Signal sequence recep 1.14e+03
	550	12	7.0	217	19	T20859	Human gene signature 1.14e+03
c	551	12	7.0	218	40	V09101	5' nucleotide sequenc 1.14e+03
c	552	12	7.0	220	38	V13667	Hepatitis G virus nuc 1.14e+03

c	553	12	7.0	220	38	V13671	Hepatitis G virus nuc	1.14e+03
c	554	12	7.0	220	38	V13702	Hepatitis G virus nuc	1.14e+03
c	555	12	7.0	220	38	V13661	Hepatitis G virus nuc	1.14e+03
c	556	12	7.0	220	38	V13701	Hepatitis G virus nuc	1.14e+03
c	557	12	7.0	220	38	V13726	Hepatitis G virus nuc	1.14e+03
c	558	12	7.0	220	38	V13728	Hepatitis G virus nuc	1.14e+03
c	559	12	7.0	220	38	V13565	Hepatitis G virus nuc	1.14e+03
c	560	12	7.0	220	38	V13572	Hepatitis G virus nuc	1.14e+03
c	561	12	7.0	220	38	V13555	Hepatitis G virus nuc	1.14e+03
c	562	12	7.0	220	38	V13673	Hepatitis G virus nuc	1.14e+03
c	563	12	7.0	220	38	V13676	Hepatitis G virus nuc	1.14e+03
c	564	12	7.0	220	38	V13800	Hepatitis G virus nuc	1.14e+03
c	565	12	7.0	220	38	V13704	Hepatitis G virus nuc	1.14e+03
c	566	12	7.0	220	38	V13796	Hepatitis G virus nuc	1.14e+03
c	567	12	7.0	220	38	V13683	Hepatitis G virus nuc	1.14e+03
c	568	12	7.0	220	38	V13698	Hepatitis G virus nuc	1.14e+03
c	569	12	7.0	220	38	V13562	Hepatitis G virus nuc	1.14e+03
c	570	12	7.0	220	38	V13727	Hepatitis G virus nuc	1.14e+03
c	571	12	7.0	220	38	V13560	Hepatitis G virus nuc	1.14e+03
c	572	12	7.0	220	38	V13549	Hepatitis G virus nuc	1.14e+03
c	573	12	7.0	220	38	V13626	Hepatitis G virus nuc	1.14e+03
c	574	12	7.0	220	38	V13588	Hepatitis G virus nuc	1.14e+03
	575	12	7.0	221	22	T22986	Human gene signature	1.14e+03
c	576	12	7.0	221	19	T21163	Human gene signature	1.14e+03
c	577	12	7.0	225	15	Q85549	DNA probe 55 detects	1.14e+03
	578	12	7.0	231	2	N71258	Sequence encoding a c	1.14e+03
	579	12	7.0	243	13	Q80705	AcMNPV 6.9K gene prom	1.14e+03
c	580	12	7.0	245	20	T22695	Human gene signature	1.14e+03
c	581	12	7.0	278	8	Q61106	Human brain Expressed	1.14e+03
c	582	12	7.0	279	19	T21657	Human gene signature	1.14e+03
	583	12	7.0	285	10	Q77447	Human genome fragment	1.14e+03
	584	12	7.0	299	1	Q05757	Clone cBp408 with hom	1.14e+03
	585	12	7.0	300	35	T69361	Murine metastatic nuc	1.14e+03
	586	12	7.0	300	35	T69368	Murine metastatic nuc	1.14e+03
c	587	12	7.0	305	8	Q60568	Human brain Expressed	1.14e+03
c	588	12	7.0	310	10	Q54242	P-1 cancer cell growt	1.14e+03
c	589	12	7.0	315	40	V24796	H. pylori ORF 11gel03	1.14e+03
	590	12	7.0	317	8	Q61187	Human brain Expressed	1.14e+03
	591	12	7.0	318	1	N90247	Amino acids 23-128 of	1.14e+03
	592	12	7.0	318	21	T26565	Human gene signature	1.14e+03
	593	12	7.0	322	20	T22160	Human gene signature	1.14e+03
	594	12	7.0	364	1	Q05756	Clone cBp405 with hom	1.14e+03
	595	12	7.0	367	8	Q60284	Human brain Expressed	1.14e+03
	596	12	7.0	370	2	Q06503	AcNPV basic protein p	1.14e+03
	597	12	7.0	377	29	T47807	Rabbit cDNA encoding	1.14e+03
	598	12	7.0	378	3	N50055	Sequence encoding hum	1.14e+03
c	599	12	7.0	381	30	T67697	H. pylori secreted or	1.14e+03
	600	12	7.0	384	1	Q05755	Clone cBp401 with hom	1.14e+03
	601	12	7.0	392	8	Q59811	Human brain Expressed	1.14e+03
c	602	12	7.0	396	38	V13591	Hepatitis G virus nuc	1.14e+03
	603	12	7.0	403	8	Q60573	Human brain Expressed	1.14e+03

c	604	12	7.0	406	6	Q39790	Expressed Sequence Ta	1.14e+03
c	605	12	7.0	406	8	Q59202	Human brain Expressed	1.14e+03
	606	12	7.0	427	36	T93338	EST locus ys82el2.r1	1.14e+03
	607	12	7.0	445	19	T19906	Human gene signature	1.14e+03
	608	12	7.0	449	24	T38481	Rat petrin cDNA clone	1.14e+03
	609	12	7.0	453	31	T43496	ATM gene exons 64-65.	1.14e+03
	610	12	7.0	491	35	T75665	Human preproamylin PC	1.14e+03
	611	12	7.0	492	38	T84952	Human prostate protei	1.14e+03
	612	12	7.0	497	9	Q61882	Human chromosome 6 cC	1.14e+03
	613	12	7.0	497	29	T65003	Human chromosome 6 re	1.14e+03
	614	12	7.0	515	32	T48198	ITS2 derived from P.	1.14e+03
	615	12	7.0	518	9	Q55749	Brassica root specifi	1.14e+03
c	616	12	7.0	527	5	Q31995	5' flanking region of	1.14e+03
	617	12	7.0	529	2	N60162	Sequence of enterotox	1.14e+03
	618	12	7.0	545	30	T43932	Sequence flanking mar	1.14e+03
c	619	12	7.0	557	11	Q65333	Prostaglandin I2 prod	1.14e+03
	620	12	7.0	564	30	T67571	H. pylori cytoplasmic	1.14e+03
c	621	12	7.0	576	14	Q83870	Hepatitis C virus env	1.14e+03
c	622	12	7.0	576	21	T16584	Hepatitis C virus iso	1.14e+03
c	623	12	7.0	585	11	Q68826	Partial sequence of h	1.14e+03
	624	12	7.0	600	19	T13269	TIMP-3-promoter.	1.14e+03
	625	12	7.0	636	9	Q54677	Promoter/Enhancer.	1.14e+03
	626	12	7.0	642	30	T68129	H. pylori cytoplasmic	1.14e+03
	627	12	7.0	663	18	T16152	Heptaprenyl diphospha	1.14e+03
c	628	12	7.0	669	40	V10151	Stealth virus nucleic	1.14e+03
	629	12	7.0	685	22	T32431	Wasp venom BrhTX-1 su	1.14e+03
c	630	12	7.0	693	1	Q06408	Sequence hybridising	1.14e+03
c	631	12	7.0	709	40	V16574	Polynucleotide sequen	1.14e+03
c	632	12	7.0	709	27	T45886	Human colon specific	1.14e+03
c	633	12	7.0	711	18	T16086	Mouse epimorphin frag	1.14e+03
	634	12	7.0	741	2	Q12866	Human Cytotoxic Cell	1.14e+03
	635	12	7.0	744	9	Q52274	Soybean seed storage	1.14e+03
c	636	12	7.0	768	38	T84941	Human prostate protei	1.14e+03
c	637	12	7.0	771	20	T12178	Partial pJG4-5-CDK-BP	1.14e+03
c	638	12	7.0	779	22	T36879	26-10 anti-digoxin tw	1.14e+03
	639	12	7.0	786	27	T46148	T. cruzi L19E homolog	1.14e+03
c	640	12	7.0	804	30	T67549	H. pylori transmembra	1.14e+03
c	641	12	7.0	809	9	Q52665	Chimeric Ig superfami	1.14e+03
c	642	12	7.0	823	35	T64847	Human erythropoietin	1.14e+03
c	643	12	7.0	830	31	T70299	Pea PsMTA gene promot	1.14e+03
	644	12	7.0	843	3	N50020	Sequence encoding the	1.14e+03
c	645	12	7.0	848	7	Q41069	26-10 sFv coding sequ	1.14e+03
c	646	12	7.0	864	38	T88157	H. pylori VacA autotr	1.14e+03
c	647	12	7.0	867	39	T88057	Partial cDNA clone en	1.14e+03
c	648	12	7.0	867	38	V02135	Human secreted protei	1.14e+03
	649	12	7.0	869	31	T49399	Banana bunchy top vir	1.14e+03
	650	12	7.0	885	37	T93256	Restriction enzyme Ac	1.14e+03
	651	12	7.0	895	39	V04180	Phytophthora parasiti	1.14e+03
	652	12	7.0	897	39	V04183	Phytophthora nicotian	1.14e+03
c	653	12	7.0	900	38	T91284	Bacillus subtilis mat	1.14e+03
	654	12	7.0	906	4	N30168	Sequence of bases 248	1.14e+03

c	655	12	7.0	911	38	V01548	Rat syntaxin 2 gene.	1.14e+03
	656	12	7.0	928	29	T63696	Brassica hsp80 gene p	1.14e+03
	657	12	7.0	941	17	T03745	Human gax gene.	1.14e+03
c	658	12	7.0	941	4	N30062	Sequence of a modifie	1.14e+03
	659	12	7.0	954	7	Q42321	EAEC coagulation sequ	1.14e+03
c	660	12	7.0	958	4	N20005	Hybrid human leukocyt	1.14e+03
	661	12	7.0	976	11	Q65404	Cuphea hookeriana thi	1.14e+03
c	662	12	7.0	999	2	Q12870	Promoter sequence for	1.14e+03
	663	12	7.0	1002	1	Q05037	Sequence encoding bac	1.14e+03
c	664	12	7.0	1005	39	T76918	Probe used to detect	1.14e+03
c	665	12	7.0	1020	1	N92873	Expression plasmid pU	1.14e+03
c	666	12	7.0	1033	40	V05281	Sequence of the MS8-1	1.14e+03
	667	12	7.0	1052	1	N90224	Malaria-specific DNA	1.14e+03
c	668	12	7.0	1062	2	Q11131	Histo-blood gp. A gly	1.14e+03
c	669	12	7.0	1065	9	Q50190	Human adipogenesis in	1.14e+03
c	670	12	7.0	1070	29	T59195	Clone pT218 containin	1.14e+03
c	671	12	7.0	1095	40	V18817	Human papillomavirus-	1.14e+03
c	672	12	7.0	1109	4	N20026	Human leukocyte inter	1.14e+03
c	673	12	7.0	1138	39	V09715	L. esculentum E8 prom	1.14e+03
c	674	12	7.0	1153	9	Q57495	Rat GAP-43.	1.14e+03
c	675	12	7.0	1153	5	Q30680	Sequence of rat GAP-4	1.14e+03
	676	12	7.0	1167	35	T88920	Scaffold attachment r	1.14e+03
	677	12	7.0	1183	11	Q63613	T-cell transmembrane	1.14e+03
	678	12	7.0	1187	17	T00739	Multiple tumour suppr	1.14e+03
	679	12	7.0	1187	11	Q64115	Thermal resistance ge	1.14e+03
c	680	12	7.0	1187	2	Q12402	Gene encoding dihydro	1.14e+03
	681	12	7.0	1197	35	T85615	Alpha subunit of huma	1.14e+03
c	682	12	7.0	1207	40	V05279	Sequence of the MS8-1	1.14e+03
c	683	12	7.0	1239	35	T93979	DNA for oligopeptide	1.14e+03
c	684	12	7.0	1246	40	V19643	Human CRF-binding pro	1.14e+03
c	685	12	7.0	1271	8	Q46869	Growth factor gene (B	1.14e+03
	686	12	7.0	1272	37	T98796	G protein coupled rec	1.14e+03
c	687	12	7.0	1299	4	Q27516	Alkali-protease Ya en	1.14e+03
	688	12	7.0	1317	31	T71697	Human deoxycytidylate	1.14e+03
c	689	12	7.0	1320	40	V17564	Coding sequence for t	1.14e+03
c	690	12	7.0	1322	2	N70148	Sequence of (2'-5') o	1.14e+03
c	691	12	7.0	1342	2	N60513	cDNA of clone lambda	1.14e+03
	692	12	7.0	1347	12	Q67347	VCAM 2D-IgG.	1.14e+03
	693	12	7.0	1357	1	Q04493	Sequence encoding hum	1.14e+03
	694	12	7.0	1359	18	Q98462	X.agilis phthalyl ami	1.14e+03
	695	12	7.0	1359	18	T16451	Xanthobacter agilis p	1.14e+03
	696	12	7.0	1362	18	Q98469	X.agilis phthalyl ami	1.14e+03
	697	12	7.0	1374	40	V24781	H. pylori ORF hp2e109	1.14e+03
c	698	12	7.0	1403	38	T93318	Tomato-infecting gemi	1.14e+03
c	699	12	7.0	1422	16	Q95109	Human papillomavirus	1.14e+03
	700	12	7.0	1423	14	Q85430	TOM41 gene.	1.14e+03
c	701	12	7.0	1454	40	V15439	Human gene fragment Q	1.14e+03
c	702	12	7.0	1457	40	V15409	Methicillin resistant	1.14e+03
c	703	12	7.0	1469	4	N30039	Sequence encoding bov	1.14e+03
	704	12	7.0	1476	19	T09293	SDS-resistant hyperth	1.14e+03
c	705	12	7.0	1483	13	Q74413	Spinach cysteine synt	1.14e+03

	706	12	7.0	1494	24	T42692	Glycerol kinase codin	1.14e+03
c	707	12	7.0	1507	29	T48511	Homology vector 443-8	1.14e+03
c	708	12	7.0	1518	1	N92872	Expression plasmid pU	1.14e+03
	709	12	7.0	1528	26	T45972	Mouse utrophin gene p	1.14e+03
	710	12	7.0	1588	37	T93278	Restriction enzyme Ac	1.14e+03
c	711	12	7.0	1590	16	T00622	Human (2'-5') oligo A	1.14e+03
	712	12	7.0	1601	11	Q64475	Glycogen synthase gen	1.14e+03
	713	12	7.0	1605	24	T37310	Aromatic acyl transfe	1.14e+03
c	714	12	7.0	1639	16	Q74059	Human interleukin-8.	1.14e+03
c	715	12	7.0	1656	39	V02978	Human interleukin-2 r	1.14e+03
c	716	12	7.0	1667	8	Q48589	Alcohol oxidase gene	1.14e+03
c	717	12	7.0	1678	17	T09967	E4 promoter/Adometase	1.14e+03
	718	12	7.0	1682	16	Q91763	Human Rad51 gene, enc	1.14e+03
c	719	12	7.0	1692	16	Q91099	Human herpesvirus-6 Z	1.14e+03
c	720	12	7.0	1697	1	Q06257	5' non-coding region	1.14e+03
	721	12	7.0	1700	19	T10582	Transcription factor	1.14e+03
	722	12	7.0	1700	10	Q65344	Transcription factor	1.14e+03
	723	12	7.0	1721	9	Q54257	APP-REP 751 amyloid p	1.14e+03
	724	12	7.0	1723	37	T91745	Tobacco calcium/calmo	1.14e+03
	725	12	7.0	1743	38	T96645	Human TULP4 cDNA.	1.14e+03
c	726	12	7.0	1744	40	V15229	Class II Cuphea hooke	1.14e+03
c	727	12	7.0	1744	11	Q63900	Acyl CoA synthetase g	1.14e+03
c	728	12	7.0	1744	15	Q92305	Cuphea classs II thioe	1.14e+03
c	729	12	7.0	1745	34	T69249	Flt-1 VEGF receptor g	1.14e+03
	730	12	7.0	1764	37	T90819	Bacillus stearothermo	1.14e+03
	731	12	7.0	1768	21	T07326	Mouse PH30 beta chain	1.14e+03
	732	12	7.0	1771	3	N50495	Human parotid alpha-a	1.14e+03
	733	12	7.0	1773	40	V00675	Hepatitis C virus opt	1.14e+03
c	734	12	7.0	1784	3	N40145	Sequence of protein A	1.14e+03
	735	12	7.0	1826	36	T86123	Cotton protoporphyrin	1.14e+03
	736	12	7.0	1826	38	V04313	Cotton protox-1 cDNA.	1.14e+03
	737	12	7.0	1881	35	T68838	Photorhabdus luminesc	1.14e+03
c	738	12	7.0	1894	16	T01009	Human tyrosinase acti	1.14e+03
c	739	12	7.0	1906	12	Q66548	Human tyrosinase DNA.	1.14e+03
	740	12	7.0	1922	22	T14712	Human CD19 antigen cD	1.14e+03
	741	12	7.0	1922	24	T38556	Human CD19 cDNA.	1.14e+03
	742	12	7.0	1929	35	T72642	Human testis-specific	1.14e+03
	743	12	7.0	1938	18	T08079	Fibrinogen binding su	1.14e+03
	744	12	7.0	1939	6	Q34622	Cytosolic GS3A glutam	1.14e+03
	745	12	7.0	1941	7	Q43586	VCAM-6D DNA.	1.14e+03
	746	12	7.0	1945	16	T02814	Spinach 2-oxoglutarat	1.14e+03
	747	12	7.0	1949	1	Q06018	Sequence encoding Pro	1.14e+03
	748	12	7.0	1950	2	Q10001	Sequecne encoding Str	1.14e+03
	749	12	7.0	1972	3	Q14098	B.thuringiensis cryII	1.14e+03
	750	12	7.0	1978	6	Q37189	Petunia EPSP synthase	1.14e+03
c	751	12	7.0	1997	40	V05280	Sequence of the MS8-1	1.14e+03
c	752	12	7.0	2000	3	Q21400	Vibrolysin gene seque	1.14e+03
c	753	12	7.0	2000	36	T77839	Human melanoma associ	1.14e+03
c	754	12	7.0	2024	1	Q03143	Nucleotide sequence o	1.14e+03
	755	12	7.0	2042	29	T63693	Brassica heat shock p	1.14e+03
	756	12	7.0	2080	1	Q03708	Gene encoding murine	1.14e+03

	757	12	7.0	2085 12	Q69003	Mouse amyloid precurs	1.14e+03
	758	12	7.0	2107 39	V15900	Uridine diphospho-glu	1.14e+03
c	759	12	7.0	2115 2	Q11120	Exon 4 of Human lutei	1.14e+03
	760	12	7.0	2148 2	N92751	DNA encoding human ca	1.14e+03
c	761	12	7.0	2160 15	Q93102	Human mu opiate recep	1.14e+03
c	762	12	7.0	2163 39	V04043	Nematode-inducible pl	1.14e+03
c	763	12	7.0	2173 1	Q06284	Rat Tumour Necrosis F	1.14e+03
c	764	12	7.0	2181 7	Q43207	Sequence encoding mic	1.14e+03
	765	12	7.0	2190 4	Q27945	Sequence of CryIIIC(b	1.14e+03
c	766	12	7.0	2191 33	T80581	Canine plasma platele	1.14e+03
	767	12	7.0	2196 3	Q20001	Encodes Human acyl am	1.14e+03
	768	12	7.0	2208 28	T51459	cDNA encoding p53 bin	1.14e+03
c	769	12	7.0	2215 12	Q70882	Tomato spotted wilt v	1.14e+03
	770	12	7.0	2220 3	Q21005	VCAM variant with 7 I	1.14e+03
	771	12	7.0	2237 37	T90287	Human glycophosphatid	1.14e+03
	772	12	7.0	2247 32	T61187	Human phospholipase A	1.14e+03
	773	12	7.0	2247 5	Q29430	DNA encoding cytosoli	1.14e+03
	774	12	7.0	2256 1	N80604	Lambda APCP16814,amin	1.14e+03
c	775	12	7.0	2293 35	T89279	TPRP-P1 gene promoter	1.14e+03
	776	12	7.0	2305 6	Q39149	Annexin XI gene in cl	1.14e+03
	777	12	7.0	2310 40	V10322	Human APP770 cDNA.	1.14e+03
	778	12	7.0	2320 13	Q85099	PagC gene.	1.14e+03
	779	12	7.0	2345 19	T07170	Human GRB-7 genomic D	1.14e+03
	780	12	7.0	2352 2	Q11490	gSPO-B1 insert encodi	1.14e+03
	781	12	7.0	2377 11	Q68538	Extracellular neutral	1.14e+03
c	782	12	7.0	2377 1	N90726	DNA sequence of extra	1.14e+03
c	783	12	7.0	2377 11	Q68538	Extracellular neutral	1.14e+03
c	784	12	7.0	2379 21	T26999	N. meningitidis serot	1.14e+03
	785	12	7.0	2381 1	Q05241	Sequence enoding acyl	1.14e+03
	786	12	7.0	2383 11	Q75036	Streptococcus Protein	1.14e+03
	787	12	7.0	2383 2	Q10002	Sequence encoding Str	1.14e+03
c	788	12	7.0	2384 22	T33316	Tyrosinase mRNA.	1.14e+03
	789	12	7.0	2408 14	Q81334	(1-3)-beta-D-glucan s	1.14e+03
c	790	12	7.0	2415 31	T75077	DNA encoding leucyl-t	1.14e+03
	791	12	7.0	2417 32	T77279	Deltal-pyrroline-5-ca	1.14e+03
	792	12	7.0	2427 1	N90786	Clone L4 of Brassica	1.14e+03
	793	12	7.0	2430 30	T61550	Murine interleukin-4	1.14e+03
	794	12	7.0	2483 30	T45982	Human cocaine and amp	1.14e+03
c	795	12	7.0	2511 21	T34685	Squash mosaic virus c	1.14e+03
c	796	12	7.0	2530 33	T59545	Coding sequence for O	1.14e+03
	797	12	7.0	2544 28	T47200	Mouse ETS2 repressor	1.14e+03
	798	12	7.0	2547 17	T15543	Borrelia garinii 66 k	1.14e+03
	799	12	7.0	2553 21	T07327	Mouse PH30 beta chain	1.14e+03
	800	12	7.0	2577 16	Q90941	Murine wild type MLH1	1.14e+03
	801	12	7.0	2581 2	N70974	Sequence of human atr	1.14e+03
	802	12	7.0	2589 7	Q41454	Human pre-proANVP gen	1.14e+03
	803	12	7.0	2590 4	N50695	Sequence encoding pre	1.14e+03
	804	12	7.0	2606 32	T72895	Plasmodium E31a gene.	1.14e+03
c	805	12	7.0	2606 36	T86134	Sugarbeet protoporphy	1.14e+03
	806	12	7.0	2608 25	T35909	Herpesvirus of turkey	1.14e+03
	807	12	7.0	2621 13	Q76109	Tomato spotted wilt v	1.14e+03

	808	12	7.0	2635 38	T95763	Arabidopsis SCARECROW	1.14e+03
	809	12	7.0	2636 37	V01871	Human brain specific	1.14e+03
	810	12	7.0	2676 13	Q76371	Mouse growth differen	1.14e+03
	811	12	7.0	2712 15	Q87836	Human neuronal calciu	1.14e+03
c	812	12	7.0	2720 12	Q70904	Cold acclimatization	1.14e+03
	813	12	7.0	2728 14	Q83797	S. avermitilis bkd ge	1.14e+03
	814	12	7.0	2734 6	Q39221	Encodes human epiderm	1.14e+03
	815	12	7.0	2744 5	Q31769	Gastrin-binding prote	1.14e+03
	816	12	7.0	2746 2	Q07044	Gene encoding AmF105L	1.14e+03
	817	12	7.0	2790 2	N60473	Sequence encoding the	1.14e+03
c	818	12	7.0	2796 17	T09966	Tomato E4 gene.	1.14e+03
	819	12	7.0	2811 1	Q06687	Vascular cell adhesio	1.14e+03
	820	12	7.0	2887 33	T61405	MACH isoform alphal c	1.14e+03
	821	12	7.0	2938 9	Q50644	Human Hum-Fli-1 gene	1.14e+03
	822	12	7.0	2960 2	N70224	Gene encoding human G	1.14e+03
c	823	12	7.0	2970 38	T99433	Yeast MPS1 gene encod	1.14e+03
c	824	12	7.0	3012 39	T94211	Thermotoga maritima e	1.14e+03
c	825	12	7.0	3013 17	Q94328	Rabbit telencephalin	1.14e+03
	826	12	7.0	3029 17	T15383	Phthalyl-amidase gene	1.14e+03
c	827	12	7.0	3036 40	V20474	Human PML/RARalpha on	1.14e+03
	828	12	7.0	3063 1	N91438	DNA encoding glycine	1.14e+03
	829	12	7.0	3069 2	Q11772	Sequence encoding hum	1.14e+03
	830	12	7.0	3080 1	Q06688	Vascular cell adhesio	1.14e+03
	831	12	7.0	3095 15	Q93911	Human mismatch repair	1.14e+03
	832	12	7.0	3131 11	Q66687	EHEC eae gene associa	1.14e+03
	833	12	7.0	3156 14	Q82976	Genomic clone encodin	1.14e+03
	834	12	7.0	3182 12	Q70730	TATA-binding protein-	1.14e+03
c	835	12	7.0	3194 10	Q58039	GM-CSF gene.	1.14e+03
c	836	12	7.0	3198 38	T92869	Candida CaRhol gene.	1.14e+03
	837	12	7.0	3230 1	N81381	Entire nucleotide seq	1.14e+03
	838	12	7.0	3230 4	Q25595	Human GnT I cDNA clon	1.14e+03
c	839	12	7.0	3245 7	Q40730	Mouse eps8.	1.14e+03
c	840	12	7.0	3252 31	T64784	Rat oxygen regulated	1.14e+03
c	841	12	7.0	3257 19	T14352	Nsk2 receptor gene.	1.14e+03
	842	12	7.0	3280 2	N82053	DNA encoding fibronec	1.14e+03
	843	12	7.0	3353 3	Q14097	Amyloid precursor pro	1.14e+03
	844	12	7.0	3353 1	Q04496	Sequence of gene enco	1.14e+03
	845	12	7.0	3404 17	T15595	Thermostable cellulas	1.14e+03
c	846	12	7.0	3415 22	T36481	Human integrin beta s	1.14e+03
	847	12	7.0	3435 14	Q79746	Mycoplasma pirum adhe	1.14e+03
c	848	12	7.0	3447 18	T08711	Sak-a serine-threonin	1.14e+03
c	849	12	7.0	3471 5	Q28940	B.thuringiensis PS50C	1.14e+03
c	850	12	7.0	3511 22	T33259	PML-RAR-alpha DNA seq	1.14e+03
c	851	12	7.0	3521 7	Q41226	Clone GP3 encoding ma	1.14e+03
c	852	12	7.0	3522 8	Q47291	Delta endotoxin gene.	1.14e+03
	853	12	7.0	3588 6	Q40057	Bacterial delta-6-des	1.14e+03
	854	12	7.0	3588 21	T30396	Synechocystis delta-6	1.14e+03
	855	12	7.0	3695 6	Q32844	Encodes E. coli heat	1.14e+03
c	856	12	7.0	3695 6	Q32844	Encodes E. coli heat	1.14e+03
c	857	12	7.0	3746 1	N90693	Bovine peptidyl-glyci	1.14e+03
	858	12	7.0	3748 33	T84330	Maize DnaJ-related ge	1.14e+03

	859	12	7.0	3771 16	Q94055	Strain 33f2 DNA.	1.14e+03
	860	12	7.0	3771 15	Q51689	Bacillus thuringiensi	1.14e+03
c	861	12	7.0	3826 10	Q45288	Sequence of plasmid p	1.14e+03
c	862	12	7.0	3856 1	Q03097	Enod2b genomic clone.	1.14e+03
c	863	12	7.0	3892 13	Q79354	Sequence of the exten	1.14e+03
c	864	12	7.0	3934 20	T30309	Human SOX-9 cDNA.	1.14e+03
c	865	12	7.0	3960 8	Q48732	CT gene.	1.14e+03
	866	12	7.0	3974 26	T43360	Cotton FbLate2-82A ge	1.14e+03
c	867	12	7.0	4000 38	T91902	Mannose-1-phosphate t	1.14e+03
	868	12	7.0	4040 29	T59731	Human interleukin-12	1.14e+03
	869	12	7.0	4060 15	Q90768	BAV3 genome (left 11%	1.14e+03
c	870	12	7.0	4093 9	Q49264	ced-4.	1.14e+03
	871	12	7.0	4146 37	T93281	Restriction enzyme Ac	1.14e+03
c	872	12	7.0	4200 17	Q98503	Rabbit 97 kDa dystrog	1.14e+03
	873	12	7.0	4235 12	Q70180	Sequence encoding pro	1.14e+03
	874	12	7.0	4253 34	T86172	Human MLN 51 nucleoti	1.14e+03
	875	12	7.0	4286 19	T15158	CDK4I cyclin-dependen	1.14e+03
c	876	12	7.0	4320 35	T89278	TPRP-P1 gene, includi	1.14e+03
	877	12	7.0	4374 13	Q72914	Murine receptor-type	1.14e+03
c	878	12	7.0	4394 3	Q21604	Alpha galactosidase g	1.14e+03
c	879	12	7.0	4425 32	T72781	Human interleukin-1 r	1.14e+03
	880	12	7.0	4428 40	V15700	Human IDS gene.	1.14e+03
c	881	12	7.0	4440 30	T68322	Human multidrug resis	1.14e+03
c	882	12	7.0	4467 2	N92428	Sequence encoding amy	1.14e+03
c	883	12	7.0	4515 24	T38486	Rat petrin cDNA.	1.14e+03
c	884	12	7.0	4544 22	T34552	Platelet-derived grow	1.14e+03
	885	12	7.0	4565 6	Q36024	LeEF-1 genomic clone	1.14e+03
c	886	12	7.0	4617 31	T72653	Rat tripeptidylpeptid	1.14e+03
c	887	12	7.0	4621 31	T74034	Mouse male-enhanced a	1.14e+03
c	888	12	7.0	4626 40	V20444	Human c-met oncogene.	1.14e+03
	889	12	7.0	4698 35	T68849	Photorhabdus luminesc	1.14e+03
c	890	12	7.0	4708 3	N50415	Human acetyl choline	1.14e+03
c	891	12	7.0	4746 27	T47874	K. pneumoniae diol de	1.14e+03
c	892	12	7.0	4752 38	V04868	Nucleotide sequence o	1.14e+03
c	893	12	7.0	4799 38	V04887	Nucleotide sequence o	1.14e+03
c	894	12	7.0	4815 38	V04886	Nucleotide sequence o	1.14e+03
	895	12	7.0	4823 17	T03943	Human thrombopoietin	1.14e+03
	896	12	7.0	4856 34	T70388	Starch binding protei	1.14e+03
	897	12	7.0	4864 36	V01455	A. thaliana inorganic	1.14e+03
	898	12	7.0	4910 22	T15287	prEP78/52 hybrid aden	1.14e+03
	899	12	7.0	4918 10	Q56696	Sequence of the L. la	1.14e+03
c	900	12	7.0	4951 17	T05848	Polycistronic cps gen	1.14e+03
c	901	12	7.0	4954 14	Q55753	Escherichia coli geno	1.14e+03
c	902	12	7.0	4972 31	T69203	Lawsonia intracellula	1.14e+03
c	903	12	7.0	4998 1	N81114	Non-A, non-B hepatiti	1.14e+03
	904	12	7.0	5023 2	Q13821	Hamster GAP b3 gene.	1.14e+03
c	905	12	7.0	5113 26	T48479	Human tyrosine dephos	1.14e+03
	906	12	7.0	5211 13	Q77789	Pre-pro-cobra C3 codi	1.14e+03
c	907	12	7.0	5248 14	Q88185	Bacillus circulans CF	1.14e+03
	908	12	7.0	5283 35	T91972	DNA encoding rat thro	1.14e+03
c	909	12	7.0	5291 1	Q05768	GPAL2 promoter.	1.14e+03

c	910	12	7.0	5359 17	T12251	Cytochrome P450 isoen	1.14e+03
c	911	12	7.0	5362 39	T97296	Mouse additional sex	1.14e+03
c	912	12	7.0	5362 39	T97294	Mouse additional sex	1.14e+03
	913	12	7.0	5467 24	T43685	Medium chain-specific	1.14e+03
	914	12	7.0	5721 1	Q06613	Adenyl cyclase gene f	1.14e+03
c	915	12	7.0	5737 16	Q74053	Human interleukin-2.	1.14e+03
c	916	12	7.0	5737 10	Q58034	Interleukin-2 gene.	1.14e+03
c	917	12	7.0	5775 40	V20437	Human c-K-ras oncogen	1.14e+03
c	918	12	7.0	5925 8	Q48733	CAI gene.	1.14e+03
	919	12	7.0	6005 35	T68837	Photorhabdus luminesc	1.14e+03
	920	12	7.0	6042 20	T09018	Arabidopsis thaliana	1.14e+03
	921	12	7.0	6057 10	Q54012	Alpha-agglutinin codi	1.14e+03
c	922	12	7.0	6075 4	Q27451	Type A human platelet	1.14e+03
	923	12	7.0	6169 24	T39284	Mobilisable cloning v	1.14e+03
c	924	12	7.0	6175 26	T34291	Coding sequence for s	1.14e+03
	925	12	7.0	6180 24	T36387	Periplasmic chitodext	1.14e+03
c	926	12	7.0	6402 40	V12113	Sus scrofa factor VII	1.14e+03
	927	12	7.0	6464 8	Q48768	Arylamine N-acetyl-tr	1.14e+03
c	928	12	7.0	6479 39	V05172	Sequence of the spind	1.14e+03
	929	12	7.0	6519 12	Q74990	P. vulgaris chondroit	1.14e+03
	930	12	7.0	6525 40	V16345	cDNA encoding full le	1.14e+03
	931	12	7.0	6766 39	V14507	AmEPV entomopoxvirus	1.14e+03
	932	12	7.0	6768 39	V14517	AmEPV entomopoxvirus	1.14e+03
c	933	12	7.0	6914 14	Q55757	Enterobacter cloacae	1.14e+03
	934	12	7.0	7277 35	T85178	Equine rhinovirus 1 (	1.14e+03
	935	12	7.0	7350 17	Q98955	Complete B-G gene enc	1.14e+03
c	936	12	7.0	7451 40	V05856	Complete sequence of	1.14e+03
c	937	12	7.0	7587 40	V15703	Zucchini ACC synthase	1.14e+03
c	938	12	7.0	7653 9	Q64743	ced-3 (C6485T) gene.	1.14e+03
c	939	12	7.0	7653 9	Q64741	ced-3 (G6372A) gene.	1.14e+03
c	940	12	7.0	7653 24	T38196	Nematode Ced-3 gene.	1.14e+03
	941	12	7.0	7722 2	Q12023	Norwalk virus genome	1.14e+03
c	942	12	7.0	7904 35	T94723	Human papillomavirus	1.14e+03
c	943	12	7.0	7904 35	T94724	Human papillomavirus	1.14e+03
	944	12	7.0	8022 29	T61982	Yeast rad3 gene.	1.14e+03
	945	12	7.0	8045 17	T12327	Retroviral vector MFG	1.14e+03
	946	12	7.0	8119 24	T33477	Subtilisin N62D/G166D	1.14e+03
c	947	12	7.0	8201 20	T12170	pJG4-5-CDK-BP cDNA cl	1.14e+03
	948	12	7.0	8220 14	Q83529	P. falciparum Proj3 g	1.14e+03
c	949	12	7.0	8247 2	N70996	Sequence encoding N-t	1.14e+03
	950	12	7.0	8450 36	T74994	Plasmid pFU1001 conta	1.14e+03
c	951	12	7.0	8585 28	T48867	Brassica napus micros	1.14e+03
	952	12	7.0	8591 34	T87084	Plasmid pCLL621 encod	1.14e+03
	953	12	7.0	8591 34	T87083	Plasmid pCLL602 encod	1.14e+03
	954	12	7.0	8591 38	V04865	cDNA encoding amyloid	1.14e+03
	955	12	7.0	8601 4	Q24828	AFP-1.	1.14e+03
	956	12	7.0	8767 31	T68735	ATM mutant 8269del403	1.14e+03
c	957	12	7.0	8808 22	T32227	Plasmid pBE146.	1.14e+03
c	958	12	7.0	8906 36	T85392	Human TRIO phosphopro	1.14e+03
	959	12	7.0	8996 31	T68783	ATM mutant 4443del175	1.14e+03
	960	12	7.0	8997 31	T68756	ATM mutant 3403del174	1.14e+03

	961	12	7.0	9046	31	T68761	ATM mutant 2125del125	1.14e+03
c	962	12	7.0	9103	21	T15657	HGV-JC variant polypr	1.14e+03
	963	12	7.0	9162	31	T68777	ATM mutant 7636del9 c	1.14e+03
	964	12	7.0	9165	31	T68778	ATM mutant 7279del6 c	1.14e+03
	965	12	7.0	9166	31	T68738	ATM mutant 7883del5 c	1.14e+03
	966	12	7.0	9167	31	T68769	ATM mutant 7240del4 c	1.14e+03
	967	12	7.0	9167	31	T68742	ATM mutant 7517del4 c	1.14e+03
	968	12	7.0	9168	31	T68776	ATM mutant 8578del3 c	1.14e+03
	969	12	7.0	9169	31	T68734	ATM mutant 8283delTC	1.14e+03
	970	12	7.0	9169	31	T68731	ATM mutant 9001delAG	1.14e+03
	971	12	7.0	9169	31	T68763	ATM mutant 1563delAG	1.14e+03
	972	12	7.0	9171	31	T68773	ATM mutant C9139T cod	1.14e+03
	973	12	7.0	9171	37	T43487	ATM open reading fram	1.14e+03
	974	12	7.0	9171	31	T68781	ATM mutant A8711G cod	1.14e+03
	975	12	7.0	9171	37	T73810	ATM gene variant G830	1.14e+03
	976	12	7.0	9171	31	T68737	ATM mutant C8140T cod	1.14e+03
	977	12	7.0	9172	31	T68732	ATM mutant 8946insA c	1.14e+03
	978	12	7.0	9202	2	N70608	VISNA sheep lentiviru	1.14e+03
	979	12	7.0	9326	4	Q25051	DNA complementary to	1.14e+03
	980	12	7.0	9370	13	Q75165	AF-4 transcript (cDNA	1.14e+03
	981	12	7.0	9432	17	Q99312	LKP1 operon.	1.14e+03
c	982	12	7.0	10140	22	T32232	Plasmid pBE92.	1.14e+03
	983	12	7.0	10216	24	T39279	Transposon delivery v	1.14e+03
c	984	12	7.0	10597	36	T60554	Plasmid pAdMLVneo-int	1.14e+03
	985	12	7.0	10723	31	T49303	cDNA sequence encodin	1.14e+03
	986	12	7.0	10723	31	T49304	cDNA encoding polypro	1.14e+03
c	987	12	7.0	11800	11	Q67676	Acetyl CoA carboxylas	1.14e+03
c	988	12	7.0	12284	12	Q71308	Sequence encoding 55	1.14e+03
c	989	12	7.0	12284	20	T29591	Hog cholera virus.	1.14e+03
	990	12	7.0	13633	7	Q40862	VAC-beta (genomic - p	1.14e+03
c	991	12	7.0	14985	40	V15693	Human Binl gene.	1.14e+03
	992	12	7.0	18177	1	N90490	DNA of human retinobl	1.14e+03
	993	12	7.0	19012	6	Q36952	HSA gene.	1.14e+03
	994	12	7.0	26764	31	T71696	Human deoxycytidylate	1.14e+03
	995	12	7.0	30967	23	T32454	Calpain large subunit	1.14e+03
	996	12	7.0	32026	36	T60559	Recombinant adenoviru	1.14e+03
	997	12	7.0	35000	36	T60557	Recombinant cis-actin	1.14e+03
	998	12	7.0	36538	36	T60558	Recombinant trans-act	1.14e+03
c	999	12	7.0	46899	9	Q54386	T. niveum Cyclosporin	1.14e+03
	1000	12	7.0	53526	40	T94101	Human PKD1 gene.	1.14e+03

ALIGNMENTS

```
RESULT    1
ID   T21965 standard; cDNA to mRNA; 172 BP.
AC   T21965;
DT   05-AUG-1996  (first entry)
DE   Human gene signature HUMGS03508.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
```



```
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K, Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 1012; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   172 BP;    46 A;    28 C;    48 G;    49 T;

  Query Match          100.0%;  Score 171;  DB 20;  Length 172;
  Best Local Similarity 100.0%;  Pred. No. 2.85e-138;
  Matches   172;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db       1 gatctgaacgatgccgtgtccagtttgcgaagtcctattcccctcctcctgtcgtgtgcc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy       1 GATCTGAACGATGCCGTGTCCAGTTTGCGAAGTCCTATTCCCCTCCTCCTGTCGTGTGCC 60

Db      61 tttgttcaggtggggatgtatttcatgtagaaggtggaagaaggctgctatnactctttg 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      61 TTTGTTCAGGTGGGGATGTATTTCATGTAGAAGGTGGAAGAAGGCTGCTATNACTCTTTG 120

Db     121 gatgggagtctggcaagaggaaattggaagataaaataaataataaagtgaaa 172
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     121 GATGGGAGTCTGGCAAGAGGAAATTGGAAGATAAAATAAATAATAAGTGAAA 172


RESULT    2
ID   N91248 standard; DNA; 3331 BP.
```



```
AC   N91248;
DT   22-MAR-1991  (first entry)
DE   Sequence of pLH1 or pLAB10000 encoding an enzyme that breaks down
KW   Alpha-amylase; lactic acid bacteria; carbohydrate; ss.
FH   Key             Location/Qualifiers
FT   cds             1..78
FT                   /*tag= a
FT                   /label=ORF 4
FT                   /note="P95652"
FT   cds             complement (1..639)
FT                   /*tag= b
FT                   /label=ORF 1
FT                   /note="P94759"
FT   cds             887..1831
FT                   /*tag= c
FT                   /label=ORF 2
FT                   /note="P93143"
FT   cds             complement (2107..2382)
FT                   /*tag= d
FT                   /label=ORF 3
FT                   /note="P94741"
FT   cds             complement (2885..0)
FT                   /*tag= e
FT                   /label=ORF 1
FT                   /note="P94742"
FT   cds             3146..0
FT                   /*tag= f
FT                   /label=ORF 4
FT                   /note="P92262"
PN   WO8901970-A.
PD   09-MAR-1989.
PF   02-SEP-1988; E00813.
PR   02-SEP-1987; EP-401972.
PA   (PLAN-) Plant Genetic Syst (UYLO-).
PI   Michiels F, Delcour J, Mahillon J, Joos H, Platteeuw C, Josson K;
DR   WPI; 89-085544/11.
DR   N-PSDB; P95652, P92262, P94759, P93143, P94741, P94742.
PT   Transformed lactic acid bacteria -
PT   using exogenous DNA hich codes enzyme that breaks down
PT   oligosaccharide and/or polysaccharide, used in silage or as a
PT   probiotic
PS   Claim 7; Fig 3a; 112pp; English.
CC   an oligosaccharide and/or a polysaccharide into a fermentable
CC   carbohydrate.
CC   The inventors claim an inoculum for silage which comprises lactic
CC   acid bacteria, esp. Lactobacillus plantarum and E. faecalis, transformed
CC   with an exogenous DNA encoding an enzyme which breaks down an oligosacc-
CC   -haride and/or a polysaccharide into a fermentable carbohydrate. The
CC   exogenous enzyme is especially an alpha-amylase encoded by the DNA
CC   sequence in N91246, particularly from nucleotides 184 to 2131 (Claims 2,
CC   6 and 10). The transformed lactic acid bacteria can break down poly-
```



```
CC   saccharides and oligosaccharides in a silage crop to provide a complete
CC   source of the fermentable carbohydrates that the lactic acid bacteria
CC   need for producing lactic acid. Also claimed is a probiotic characterised
CC   by the transformed lactic acid bacteria which can be used for establish-
CC   -ing and maintaining optimal intestinal flora in animals, including
CC   humans.  A plasmid contg. the DNA sequence in N91248 is claimed.  The
CC   plasmid is pLH1 or pLAB1000.
SQ   Sequence   3331 BP;    921 A;    735 C;    561 G;    1109 T;

  Query Match           9.9%;  Score 17;  DB 2;  Length 3331;
  Best Local Similarity 100.0%;  Pred. No. 1.26e+00;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2715 cttattatttattttat 2731
           |||||||||||||||||
Cp    167 CTTATTATTTATTTTAT 151


RESULT    3
ID   Q98818 standard; DNA; 6295 BP.
AC   Q98818;
DT   19-JAN-1996  (first entry)
DE   Arabidopsis CTR1-2 mutant genomic DNA.
KW   CTR1; constitutive triple response; transgenic plant; ethylene;
KW   stress tolerance; crop improvement; mutagenesis; ss.
OS   Arabidopsis thaliana ecotype Columbia.
PN   US5444166-A.
PD   22-AUG-1995.
PF   10-AUG-1992; 928464.
PR   10-AUG-1992; US-928464.
PR   12-JAN-1993; US-003311.
PA   (UYPE-) UNIV PENNSYLVANIA.
PI   Ecker JR,  Kieber JJ;
DR   WPI; 95-302162/39.
PT   New constitutive triple response genes and mutants - isolated from
PT   Arabidopsis thaliana, used to produce transgenic plants with
PT   improved properties
PS   Claim 3; Column 37-44; 41pp; English.
CC   Seeds of A. thaliana ecotype Columbia were subjected to X-ray
CC   mutagenesis and screened for mutants that constitutively displayed
CC   the triple response by plating on agar in the absence of added
CC   ethylene in the dark.  Sequence analysis of isolated mutant ctr1-2
CC   genomic DNA (given in Q98818) indicated a 17 bp deletion beginning at
CC   the position corresp. to position 2770 of the wild-type genomic
CC   sequence (Q98817).  Transgenic plants expressing recessive ctr1
CC   mutations resemble ethylene-treated wild-type plants.
SQ   Sequence   6295 BP;    1815 A;    1104 C;    1295 G;    2081 T;

  Query Match           9.9%;  Score 17;  DB 15;  Length 6295;
  Best Local Similarity 100.0%;  Pred. No. 1.26e+00;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```

```
Db    272 ttttatcttccaatttc 288
          ||||||||||||||||
Cp    156 TTTTATCTTCCAATTTC 140


RESULT    4
ID   T67128 standard; DNA; 6295 BP.
AC   T67128;
DT   17-JUL-1997  (first entry)
DE   Constitutive triple response ctr1-2 mutation genomic DNA.
KW   CTR; Arabidopsis thaliana; mutant; agricultural; ethylene; ss.
OS   Arabidopsis thaliana.
PN   US5602322-A.
PD   11-FEB-1997.
PF   10-AUG-1992; 928464.
PR   10-AUG-1992; US-928464.
PR   12-JAN-1993; US-003311.
PR   17-JUN-1994; US-261432.
PA   (UYPE-) UNIV PENNSYLVANIA.
PI   Ecker JR,  Kieber JJ;
DR   WPI; 97-131867/12.
PT   Plants over-expressing constitutive triple response gene - resulting
PT   in insensitivity to ethylene which can improve the quality and
PT   longevity of crops e.g. by affecting fruit ripening
PS   Claim 3; Column 37-44; 41pp; English.
CC   The present sequence represents the ctr 1-2 mutation of the constitutive
CC   triple response (CTR) gene. The mutation is a 17bp deletion beginning at
CC   position 1995 of the wild type genomic DNA which is predicted to result
CC   in a frame shift in the coding region, and was produced by X-rays. This
CC   sequence is used to transform plant cells. The plants overexpress the
CC   Arabidopsis thaliana CTR1 gene, which results in insensitivity to
CC   ethylene. This benefits plants in relation to stress such as infection
CC   due to pathogens (i.e. bacteria, viruses and fungi), wound healing and
CC   soil penetration. Since ethylene affects agriculturally important
CC   plant processes, e.g. fruit ripening, flower and leaf senescence and
CC   leaf abscission, the ability to control ethylene sensitivity could
CC   improve the quality and longevity of crops.
SQ   Sequence   6295 BP;    1815 A;    1104 C;    1295 G;    2081 T;

  Query Match             9.9%;  Score 17;  DB 30;  Length 6295;
  Best Local Similarity 100.0%;  Pred. No. 1.26e+00;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    272 ttttatcttccaatttc 288
          ||||||||||||||||
Cp    156 TTTTATCTTCCAATTTC 140


RESULT    5
ID   T67129 standard; DNA; 6312 BP.
```



```
AC   T67129;
DT   17-JUL-1997  (first entry)
DE   Constitutive triple response ctr 1-3 mutation genomic DNA.
KW   CTR; Arabidopsis thaliana; mutant; agricultural; ethylene; ss.
OS   Arabidopsis thaliana.
PN   US5602322-A.
PD   11-FEB-1997.
PF   10-AUG-1992; 928464.
PR   10-AUG-1992; US-928464.
PR   12-JAN-1993; US-003311.
PR   17-JUN-1994; US-261432.
PA   (UYPE-) UNIV PENNSYLVANIA.
PI   Ecker JR,  Kieber JJ;
DR   WPI; 97-131867/12.
PT   Plants over-expressing constitutive triple response gene - resulting
PT   in insensitivity to ethylene which can improve the quality and
PT   longevity of crops e.g. by affecting fruit ripening
PS   Claim 4; Column 43-50; 41pp; English.
CC   The present sequence represents the ctr 1-3 mutation of the constitutive
CC   triple response (CTR) gene. Specifically, a C to T transition results
CC   in a stop codon at position 1927 of the genomic sequence, produced by
CC   ethyl methanesulphonate (EMS). This sequence is used to transform plant
CC   cells. The plants overexpress the Arabidopsis thaliana CTR1 gene, which
CC   results in insensitivity to ethylene. This benefits plants in relation to
CC   stress such as infection due to pathogens (i.e. bacteria, viruses and
CC   fungi), wound healing and soil penetration. Since ethylene affects
CC   various agriculturally important plant processes, e.g. fruit ripening,
CC   flower and leaf senescence and leaf abscission, the ability to control
CC   ethylene sensitivity could improve the quality and longevity of crops.
SQ   Sequence   6312 BP;    1817 A;    1106 C;    1297 G;    2092 T;

  Query Match             9.9%;  Score 17;  DB 30;  Length 6312;
  Best Local Similarity 100.0%;  Pred. No. 1.26e+00;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    272 ttttatcttccaatttc 288
          ||||||||||||||||
Cp    156 TTTTATCTTCCAATTTC 140


RESULT    6
ID   Q98821 standard; DNA; 6312 BP.
AC   Q98821;
DT   19-JAN-1996  (first entry)
DE   Arabidopsis CTR1-4 mutant genomic DNA.
KW   CTR1; constitutive triple response; transgenic plant; ethylene;
KW   stress tolerance; crop improvement; mutagenesis; ss.
OS   Arabidopsis thaliana ecotype Columbia.
PN   US5444166-A.
PD   22-AUG-1995.
PF   10-AUG-1992; 928464.
```



```
PR   10-AUG-1992; US-928464.
PR   12-JAN-1993; US-003311.
PA   (UYPE-) UNIV PENNSYLVANIA.
PI   Ecker JR,  Kieber JJ;
DR   WPI; 95-302162/39.
PT   New constitutive triple response genes and mutants - isolated from
PT   Arabidopsis thaliana, used to produce transgenic plants with
PT   improved properties
PS   Claim 5; Column 57-64; 41pp; English.
CC   Seeds of A. thaliana ecotype Columbia were subjected to
CC   ethylmethanesulfonate mutagenesis and screened for mutants that
CC   constitutively displayed the triple response by plating on agar in
CC   the absence of added ethylene in the dark.  Sequence analysis of
CC   isolated mutant ctr1-4 genomic DNA (given in Q98818) indicated a
CC   G to A transition at a position corresp. to position 4008 of the
CC   wild-type genomic sequence (Q98817), resulting in a E596K mutation
CC   at an invariant residue in kinase catalytic domains.  Transgenic
CC   plants carrying ctr1 mutations resemble ethylene-treated
CC   wild-type plants.
SQ   Sequence   6312 BP;    1818 A;    1107 C;    1296 G;    2091 T;

  Query Match             9.9%;  Score 17;  DB 15;  Length 6312;
  Best Local Similarity 100.0%;  Pred. No. 1.26e+00;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    272 ttttatcttccaatttc 288
          ||||||||||||||||
Cp    156 TTTTATCTTCCAATTTC 140


RESULT    7
ID   Q98819 standard; DNA; 6312 BP.
AC   Q98819;
DT   19-JAN-1996  (first entry)
DE   Arabidopsis CTR1-3 mutant genomic DNA.
KW   CTR1; constitutive triple response; transgenic plant; ethylene;
KW   stress tolerance; crop improvement; mutagenesis; ss.
OS   Arabidopsis thaliana ecotype Columbia.
PN   US5444166-A.
PD   22-AUG-1995.
PF   10-AUG-1992; 928464.
PR   10-AUG-1992; US-928464.
PR   12-JAN-1993; US-003311.
PA   (UYPE-) UNIV PENNSYLVANIA.
PI   Ecker JR,  Kieber JJ;
DR   WPI; 95-302162/39.
PT   New constitutive triple response genes and mutants - isolated from
PT   Arabidopsis thaliana, used to produce transgenic plants with
PT   improved properties
PS   Claim 4; Column 43-50; 41pp; English.
CC   Seeds of A. thaliana ecotype Columbia were subjected to
```



```
CC   ethylmethane sulfonate mutagenesis and screened for mutants that
CC   constitutively displayed the triple response by plating on agar in the
CC   absence of added ethylene in the dark.  Sequence analysis of isolated
CC   mutant ctr1-3 genomic DNA (given in Q98819) indicated a C to T
CC   point mutation resulting in a stop codon at position 2701-2703 of
CC   the wild-type genomic sequence (Q98817).  Transgenic plants
CC   expressing recessive ctr1 mutations resemble ethylene-treated
CC   wild-type plants.
SQ   Sequence   6312 BP;    1818 A;    1106 C;    1297 G;    2091 T;

  Query Match             9.9%;  Score 17;  DB 15;  Length 6312;
  Best Local Similarity 100.0%;  Pred. No. 1.26e+00;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    272 ttttatcttccaatttc 288
          ||||||||||||||||
Cp    156 TTTTATCTTCCAATTTC 140


RESULT    8
ID   T67127 standard; DNA; 6312 BP.
AC   T67127;
DT   17-JUL-1997  (first entry)
DE   Constitutive triple response genomic DNA.
KW   CTR; Arabidopsis thaliana; mutant; agricultural; ethylene; ss.
OS   Arabidopsis thaliana.
FH   Key             Location/Qualifiers
FT   mutation        3041
FT                   /*tag= a
FT                   /note= "Site of T-DNA insertion in mutant ctr1-5"
PN   US5602322-A.
PD   11-FEB-1997.
PF   10-AUG-1992; 928464.
PR   10-AUG-1992; US-928464.
PR   12-JAN-1993; US-003311.
PR   17-JUN-1994; US-261432.
PA   (UYPE-) UNIV PENNSYLVANIA.
PI   Ecker JR,  Kieber JJ;
DR   WPI; 97-131867/12.
DR   P-PSDB; W17938.
PT   Plants over-expressing constitutive triple response gene - resulting
PT   in insensitivity to ethylene which can improve the quality and
PT   longevity of crops e.g. by affecting fruit ripening
PS   Claim 8; Column 29-36; 41pp; English.
CC   The present sequence represents the genomic DNA for constitutive triple
CC   response (CTR) gene. The genomic DNA contains 14 introns ranging in size
CC   from 77 bp to 357 bp, and the intron/exon boundaries all fit the
CC   consensus for splice donor and acceptor sites fairly well. This sequence
CC   is used to transform plant cells. The ctr1-5 mutation was produced by
CC   a T-DNA insertion at position 3041 in intron 7 and deletion of 25 bp
CC   from the left border of the T-DNA at the junction with plant DNA. The
```

```
CC   plants overexpress the Arabidopsis thaliana CTR1 gene, which results
CC   in insensitivity to ethylene. This benefits plants in relation to stress
CC   such as infection due to pathogens (i.e. bacteria, viruses and fungi),
CC   wound healing and soil penetration. Since ethylene affects various
CC   agriculturally important plant processes, e.g. fruit ripening, flower
CC   and leaf senescence and leaf abscission, the ability to control ethylene
CC   sensitivity could improve the quality and longevity of crops.
SQ   Sequence   6312 BP;    1817 A;    1107 C;    1297 G;    2091 T;

  Query Match             9.9%;  Score 17;  DB 30;  Length 6312;
  Best Local Similarity 100.0%;  Pred. No. 1.26e+00;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    272 ttttatcttccaatttc 288
          |||||||||||||||||
Cp    156 TTTTATCTTCCAATTTC 140


RESULT   9
ID   T67131 standard; DNA; 6312 BP.
AC   T67131;
DT   17-JUL-1997  (first entry)
DE   Constitutive triple response ctr 1-4 mutation genomic DNA.
KW   CTR; Arabidopsis thaliana; mutant; agricultural; ethylene; ss.
OS   Arabidopsis thaliana.
PN   US5602322-A.
PD   11-FEB-1997.
PF   10-AUG-1992; 928464.
PR   10-AUG-1992; US-928464.
PR   12-JAN-1993; US-003311.
PR   17-JUN-1994; US-261432.
PA   (UYPE-) UNIV PENNSYLVANIA.
PI   Ecker JR,  Kieber JJ;
DR   WPI; 97-131867/12.
PT   Plants over-expressing constitutive triple response gene - resulting
PT   in insensitivity to ethylene which can improve the quality and
PT   longevity of crops e.g. by affecting fruit ripening
PS   Claim 5; Column 57-64; 41pp; English.
CC   The present sequence represents the ctr 1-4 mutation of the constitutive
CC   triple response (CTR) gene. The mutation is a point mutation from a G to
CC   A transition at position 3233 which is predicted to result in a Glu to
CC   Lys change in the amino acid 596; this was produced by ethyl
CC   methanesulphonate (EMS). This sequence is used to transform plant cells.
CC   The plants overexpress the Arabidopsis thaliana CTR1 gene, which
CC   results in insensitivity to ethylene. This benefits plants in relation to
CC   stress such as infection due to pathogens (i.e. bacteria, viruses and
CC   fungi), wound healing and soil penetration. Since ethylene affects
CC   various agriculturally important plant processes, e.g. fruit ripening,
CC   flower and leaf senescence and leaf abscission, the ability to control
CC   ethylene sensitivity could improve the quality and longevity of crops.
SQ   Sequence   6312 BP;    1818 A;    1107 C;    1296 G;    2091 T;
```



```
  Query Match             9.9%;  Score 17;  DB 30;  Length 6312;
  Best Local Similarity 100.0%;  Pred. No. 1.26e+00;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    272 ttttatcttccaatttc 288
          |||||||||||||||||
Cp    156 TTTTATCTTCCAATTTC 140


RESULT   10
ID   T67130 standard; DNA; 6312 BP.
AC   T67130;
DT   17-JUL-1997  (first entry)
DE   Constitutive triple response ctr 1-1 mutation genomic DNA.
KW   CTR; Arabidopsis thaliana; mutant; agricultural; ethylene; ss.
OS   Arabidopsis thaliana.
PN   US5602322-A.
PD   11-FEB-1997.
PF   10-AUG-1992; 928464.
PR   10-AUG-1992; US-928464.
PR   12-JAN-1993; US-003311.
PR   17-JUN-1994; US-261432.
PA   (UYPE-) UNIV PENNSYLVANIA.
PI   Ecker JR,  Kieber JJ;
DR   WPI; 97-131867/12.
PT   Plants over-expressing constitutive triple response gene - resulting
PT   in insensitivity to ethylene which can improve the quality and
PT   longevity of crops e.g. by affecting fruit ripening
PS   Claim 2; Column 49-56; 41pp; English.
CC   The present sequence represents the ctr 1-1 mutation of the constitutive
CC   triple response (CTR) gene. The mutation is a T to A transition at
CC   position 3295 of the wild-type genomic DNA predicted to result in a
CC   highly conservative substitution (Asp to Glu) at amino acid position
CC   694, this was produced by diepoxybutane (DEB). This sequence is used to
CC   transform plant cells. The plants overexpress the Arabidopsis thaliana
CC   CTR1 gene, which results in insensitivity to ethylene. This benefits
CC   plants in relation to stress such as infection due to pathogens (i.e.
CC   bacteria, viruses and fungi), wound healing and soil penetration. Since
CC   ethylene affects various agriculturally important plant processes, e.g.
CC   fruit ripening, flower and leaf senescence and leaf abscission, the
CC   ability to control ethylene sensitivity could improve the quality and
CC   longevity of crops.
SQ   Sequence   6312 BP;    1818 A;    1107 C;    1297 G;    2090 T;

  Query Match             9.9%;  Score 17;  DB 30;  Length 6312;
  Best Local Similarity 100.0%;  Pred. No. 1.26e+00;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    272 ttttatcttccaatttc 288
          |||||||||||||||||
```



```
Cp    156 TTTTATCTTCCAATTTC 140


RESULT   11
ID   Q98820 standard; DNA; 6312 BP.
AC   Q98820;
DT   19-JAN-1996  (first entry)
DE   Arabidopsis CTR1-1 mutant genomic DNA.
KW   CTR1; constitutive triple response; transgenic plant; ethylene;
KW   stress tolerance; crop improvement; mutagenesis; ss.
OS   Arabidopsis thaliana ecotype Columbia.
PN   US5444166-A.
PD   22-AUG-1995.
PF   10-AUG-1992; 928464.
PR   10-AUG-1992; US-928464.
PR   12-JAN-1993; US-003311.
PA   (UYPE-) UNIV PENNSYLVANIA.
PI   Ecker JR,  Kieber JJ;
DR   WPI; 95-302162/39.
PT   New constitutive triple response genes and mutants - isolated from
PT   Arabidopsis thaliana, used to produce transgenic plants with
PT   improved properties
PS   Claim 2; Column 51-58; 41pp; English.
CC   Seeds of A. thaliana ecotype Columbia were subjected to
CC   diepoxybutane mutagenesis and screened for mutants that
CC   constitutively displayed the triple response by plating on agar in the
CC   absence of added ethylene in the dark.  Sequence analysis of isolated
CC   mutant ctr1-1 genomic DNA (given in Q98820) indicated a T to A point
CC   mutation at a position corresp. to position 4800 of the wild-type
CC   genomic sequence (Q98817).  Transgenic plants expressing recessive ctr1
CC   mutations resemble ethylene-treated wild-type plants.
SQ   Sequence   6312 BP;    1818 A;    1107 C;    1296 G;    2091 T;

  Query Match             9.9%;  Score 17;  DB 15;  Length 6312;
  Best Local Similarity 100.0%;  Pred. No. 1.26e+00;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    272 ttttatcttccaatttc 288
          |||||||||||||||||
Cp    156 TTTTATCTTCCAATTTC 140


RESULT   12
ID   Q98817 standard; DNA; 6312 BP.
AC   Q98817;
DT   19-JAN-1996  (first entry)
DE   Arabidopsis CTR1 genomic DNA.
KW   CTR1; constitutive triple response; transgenic plant; ethylene;
KW   stress tolerance; crop improvement; ss.
OS   Arabidopsis thaliana ecotype Columbia.
PN   US5444166-A.
```



```
PD   22-AUG-1995.
PF   10-AUG-1992; 928464.
PR   10-AUG-1992; US-928464.
PR   12-JAN-1993; US-003311.
PA   (UYPE-) UNIV PENNSYLVANIA.
PI   Ecker JR,  Kieber JJ;
DR   WPI; 95-302162/39.
PT   New constitutive triple response genes and mutants - isolated from
PT   Arabidopsis thaliana, used to produce transgenic plants with
PT   improved properties
PS   Claim 9; Column 31-38; 41pp; English.
CC   The CTR1 gene was mapped to an interval between 2 RFLPs in
CC   a chromosome 5 of A. thaliana and a chromosome walk in this area was
CC   initiated using a YAC library.  A sequence for the CTR1 gene was
CC   obtd.  Expression of the CTR1 gene in transgenic plants results in a
CC   dominant ethylene-insensitive phenotype.
SQ   Sequence   6312 BP;    1818 A;    1107 C;    1297 G;    2090 T;

  Query Match             9.9%;  Score 17;  DB 15;  Length 6312;
  Best Local Similarity 100.0%;  Pred. No. 1.26e+00;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    272 ttttatcttccaatttc 288
          |||||||||||||||||
Cp    156 TTTTATCTTCCAATTTC 140


RESULT   13
ID   Q70415 standard; DNA; 1234 BP.
AC   Q70415;
DT   07-APR-1995  (first entry)
DE   Full length pre-pro-vasoactive intestinal contractor DNA.
KW   Mouse; prepro; vasoactive intestinal contractor; VIC;
KW   transformation; ss.
OS   Mus musculus.
FH   Key             Location/Qualifiers
FT   cds             78..560
FT                   /*tag= a
FT                   /product= Pre-pro-vasoactive intestinal contractor
PN   J06169774-A.
PD   21-JUN-1994.
PF   28-JUN-1990; 168535.
PR   25-MAY-1990; JP-136838.
PA   (ISEH-) ISEHAN KK.
DR   WPI; 94-237591/29.
DR   P-PSDB; R60320.
PT   New DNA encoding pre-pro-vasoactive intestinal contractor (VIC) -
PT   useful for producing pre-pro-vasoactive easily and effectively
PS   Disclosure; Fig 5; 8pp; Japanese.
CC   This sequence represents the full length mouse prepro vasoactive
CC   intestinal contractor (VIC).  The VIC gene can be used to transform
```

```
CC   microorganisms for the production of VIC.
SQ   Sequence   1234 BP;    269 A;    358 C;    319 G;    288 T;

  Query Match           9.4%;  Score 16;  DB 12;  Length 1234;
  Best Local Similarity 100.0%;  Pred. No. 5.49e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1152 cttattatttatttta 1167
          ||||||||||||||||
Cp    167 CTTATTATTTATTTA 152


RESULT   14
ID   N90025 standard; DNA; 1542 BP.
AC   N90025;
DT   1-NOV-1989  (first entry)
DE   DNA encoding antigenic circumsporozoite protein of Plasmodium
DE   malariae
KW   Antigenic; vaccine; Plasmodium malariae;
KW   circumsporozoite protein.
OS   Plasmodium malariae
FH   Key             Location/Qualifiers
FT   cds             98..1383
FT                   /*tag=  a
FT   misc_feature    424..1036
FT                   /*tag=  b
PN   US7238746-A.
PD   16-MAY-1989.
PF   31-AUG-1988; 238746.
PR   31-AUG-1988; US-238746.
PA   (USSH) Nat Inst of Health.
DR   WPI; 89-172958/23.
PT   Antigenic protein for vaccination against malaria
PT   - encoded by cloned gene coding for Plasmodium
PT   malariae circumsporozoite protein.
PS   Disclosure; fig 1; 11pp; English.
CC   DNA encoding antigenic circumsporozoite protein of
CC   Plasmodium malariae. Used as a vaccine against malaria.
CC   Nucleotides 424 - 1036 make up the immunodominant region.
SQ   Sequence   1542 BP;    618 A;    228 C;    344 G;    352 T;

  Query Match           9.4%;  Score 16;  DB 1;  Length 1542;
  Best Local Similarity 100.0%;  Pred. No. 5.49e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     33 aaaataaataataagt 48
          ||||||||||||||||
Qy    153 AAAATAAATAATAAGT 168


RESULT   15
```



```
ID   T91638 standard; DNA; 2344 BP.
AC   T91638;
DT   27-MAR-1998  (first entry)
DE   Castorbean ricin gene.
KW   Ricin; cytotoxin; hybrid protein; cell delivery;
KW   cell binding ligand; translocation domain; diphtheria toxin B';
KW   interleukin-2; T-cell lymphoma; organ rejection; therapy; ss.
OS   Ricinus communis.
FH   Key             Location/Qualifiers
FT   CDS             291..2021
FT                   /*tag=  a
FT   sig_peptide     291..395
FT                   /*tag=  b
FT   mat_peptide     396..2018
FT                   /*tag=  c
PN   US5668255-A.
PD   16-SEP-1997.
PF   07-JUN-1984; 618199.
PR   27-JUN-1991; US-722484.
PR   07-JUN-1984; US-618199.
PR   25-APR-1985; US-726808.
PR   07-JUN-1985; US-742554.
PR   22-DEC-1989; US-456095.
PR   14-JUN-1990; US-538276.
PR   04-AUG-1993; US-102387.
PA   (SERA-) SERAGEN INC.
PI   Murphy JR;
DR   WPI; 97-470103/43.
DR   P-PSDB; W25787.
PT   New hybrid molecules for delivery of agents to cells - comprise a
PT   binding domain of a cell binding ligand and a portion of a
PT   translocation domain of a protein
PS   Example 4; Fig 11A-B; 30pp; English.
CC   This sequence comprises the castorbean gene that codes for ricin
CC   (see W25787).  DNA encoding the ricin A domain and a portion
CC   of the ricin A-to-B linker peptide was used to construct a
CC   ricin-diphtheria toxin B'-interleukin-2 gene that was expressed in
CC   E. coli.  The hybrid protein can be isolated and used to treat
CC   conditions involving over-production of cells bearing IL2 receptors,
CC   such as certain T-cell lymphomas and organ transplant rejection
CC   crises.  The hybrid inactivates ribosomes in cells bearing IL2
CC   receptors, resulting in cessation of protein synthesis and death of
CC   target cells.  Claimed hybrid proteins comprise a translocation
CC   domain and a cell binding domain from e.g. a hormone, growth factor
CC   or polypeptide toxin.  The hybrid molecules can be used for the
CC   delivery of agents (e.g. therapeutic genes, toxins, detectable
CC   labels) into cells.  The use of a translocation mechanism ensures
CC   that the hybrid will be effective in relatively low doses, since a
CC   high proportion of the substance of interest will be taken into the
CC   targeted cells.  The hybrid molecules can be manufactured as a
CC   single hybrid recombinant protein, permitting reproducibility,
```



```
CC   consistency, and the precise control of composition.
SQ   Sequence   2344 BP;    718 A;    429 C;    467 G;    730 T;

  Query Match           9.4%;  Score 16;  DB 36;  Length 2344;
  Best Local Similarity 100.0%;  Pred. No. 5.49e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2237 ataaaataaataataa 2252
          ||||||||||||||||
Qy    151 ATAAAATAAATAATAA 166


RESULT   16
ID   Q04307 standard; DNA; 2475 BP.
AC   Q04307;
DT   17-SEP-1990  (first entry)
DE   Coding region of human IL-4 receptor
KW   mammalian interleukin-4 receptor; cytokine; antibody production; ss.
OS   synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    698..779
FT                   /*tag= a
FT                   /label= putative transmembrane region
PN   EP-367566-A.
PD   09-MAY-1990.
PF   31-OCT-1989; 311244.
PR   31-OCT-1988; US-265047.
PR   23-JUN-1989; US-370924.
PA   (IMMU-) Immunex Corp.
PI   Cosman DJ, Park L, Mosley B, Beckmann P, March CJ, Idzerda R;
DR   WPI; 90-141470/19.
DR   P-PSDB; R04575.
PT   Recombinant mammalian interleukin-4 receptor used in diagnosis,
PT   assays and therapy and for prodn. of antibodies for diagnosis,therapy
PT   and for prodn. of antibodies
PS   Claim 4; Fig 4; 37pp; English.
CC   The interleukin-4 receptor can be used to regulate immune responses or to
CC   treat IgE-induced hypersensitivity.
CC   See also Q04305.
SQ   Sequence   2475 BP;    527 A;    755 C;    713 G;    480 T;

  Query Match           9.4%;  Score 16;  DB 1;  Length 2475;
  Best Local Similarity 100.0%;  Pred. No. 5.49e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1292 gccttcttccaccttc 1307
          ||||||||||||||||
Cp    105 GCCTTCTTCCACCTTC 90


RESULT   17
```



```
ID   T61551 standard; cDNA; 2475 BP.
AC   T61551;
DT   07-JUL-1997  (first entry)
DE   Human interleukin-4 receptor cDNA.
KW   Interleukin-4 receptor; IL-4; immunosuppressive; antiinflammatory;
KW   inflammation; allergy; arthritis; diabetes; transplant rejection;
KW   therapy; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   signal_peptide  1..75
FT                   /*tag=  a
FT   mat_peptide     76..2475
FT                   /*tag=  b
PN   US5599905-A.
PD   04-FEB-1997.
PF   31-OCT-1988; 265047.
PR   31-OCT-1988; US-265047.
PR   02-MAR-1989; US-319438.
PR   20-MAR-1989; US-326156.
PR   23-JUN-1989; US-370924.
PR   14-FEB-1990; US-480694.
PR   20-JUL-1993; US-094669.
PA   (IMMV ) IMMUNEX CORP.
PI   Beckmann MP,  Cosman DJ,  Idzerda R,  March CJ,  Mosley B;
PI   Park L;
DR   WPI; 97-118336/11.
DR   P-PSDB; W13499.
PT   Interleukin-4 receptor proteins - useful for suppressing immune and
PT   inflammatory responses
PS   Claim 16; Fig4A-C; 47pp; English.
CC   A cDNA clone (T61551), designated T22-8, codes for human
CC   interleukin-4 (IL-4) receptor (W13499).  It was obtd. from a cDNA
CC   library derived from the T cell line T22 using a probe derived from
CC   murine Il-4 receptor cDNA (see also T61550).  IL-4 receptor cDNA
CC   can be used for recombinant prodn. of IL-4 receptor polypeptides
CC   (esp. soluble polypeptides) useful for suppressing IL-4-dependent
CC   immune or inflammatory responses, e.g. for treating allergies,
CC   arthritis or diabetes or for preventing transplant rejection.
SQ   Sequence   2475 BP;    527 A;    757 C;    711 G;    480 T;

  Query Match           9.4%;  Score 16;  DB 30;  Length 2475;
  Best Local Similarity 100.0%;  Pred. No. 5.49e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1292 gccttcttccaccttc 1307
          ||||||||||||||||
Cp    105 GCCTTCTTCCACCTTC 90


RESULT   18
ID   N90844 standard;  DNA; 4698 BP.
```

```
AC   N90844;
DT   24-JAN-1990  (first entry)
DE   Tobacco basic beta-1,3 glucanase gene.
KW   Tobacco basic beta-1,3 glucanase; pBSGluc39.3; herbicide resistance;
KW   transgenic plants.
OS   Nicotiana tabaccum cv. Xanthi.
PN   EP-332104-A.
PD   13-SEP-1989.
PF   06-MAR-1989; 103888.
PR   08-MAR-1988; US-165667.
PA   (CIBA) Ciba Geigy AG.
PI   Ryals J, Montoya A, Harms C, Duesing J, Sperisen C, Meins F,
PI   Payne G;
DR   WPI; 89-265342/37.
PT   Chemical induction of cloned genes in plants - DNA sequence and vectors
PT   and processes for inducible expression leading to herbicide and insect
PT   resistance.
PS   Claim 80; page 75-78; 117pp; English.
CC   Sequence is the tobacco basic beta-1,3-glucanase gene which is contained
CC   in clone pBSGluc39.3. It is a chemically inducible gene used to regulate
CC   transcription of an associated DNA sequence in plant tissue.
CC   See also N90367, N90368,  N90841-N90846, and P91384-P91386.
SQ   Sequence   4698 BP;   1508 A;   857 C;   785 G;   1546 T;   2 Others;

  Query Match           9.4%;  Score 16;  DB 1;  Length 4698;
  Best Local Similarity 100.0%;  Pred. No. 5.49e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    716 catgtagaaggtggaa 731
          ||||||||||||||||
Qy     84 CATGTAGAAGGTGGAA 99


RESULT   19
ID   Q03254 standard; DNA; 4699 BP.
AC   Q03254;
DT   25-JUL-1990  (first entry)
DE   Sequence of clone BSGluc39.3 encoding beta-1,3-glucanase from genomic
DE   sequence.
KW   Beta-1,3-glucanase; papermaking; polysaccharide hydrolase;
KW   ds.
OS   Nicotiana sp.
PN   EP-353191-A.
PD   31-JAN-1990.
PF   20-JUL-1989; 810555.
PR   29-JUL-1988; US-353312.
PA   (CIBA) Ciba Geigy AG.
PI   Shinshi H, Wenzler H, Hofsteenge J, Ryals J, Sperisen C;
DR   WPI; 90-031717/05.
PT   Recombinant DNA encoding beta-1,3-glucanase polypeptide(s) -
PT   useful in plant protection and biomass conversion pref. isolated
```



```
PT   from tobacco.
PS   Disclosure; Fig 9; 41pp; English.
CC   Beta-1,3-glucanase is involved in plant defence against pathogens and as
CC   it hydrolyses polyssaccharides, useful in biomass conversion or paper
CC   making. The clones allow for transfer of the gene to other species and
CC   control of gene expression.
SQ   Sequence   4699 BP;   1508 A;   857 C;   785 G;   1547 T;   2 Others;

  Query Match           9.4%;  Score 16;  DB 1;  Length 4699;
  Best Local Similarity 100.0%;  Pred. No. 5.49e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    716 catgtagaaggtggaa 731
          ||||||||||||||||
Qy     84 CATGTAGAAGGTGGAA 99


RESULT   20
ID   Q26727 standard; DNA; 7753 BP.
AC   Q26727;
DT   20-JAN-1993  (first entry)
DE   Encodes exons XV to XVIII of human hepatocyte growth factor.
KW   HGF; enhance growth; preparing transgenic animals; hepatic disease;
KW   clinical diagnostic reagent; drug.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             101..241
FT                   /*tag=  a
FT                   /note= "exon XV"
FT   cds             773..879
FT                   /*tag=  b
FT                   /note= "exon XVI"
FT   cds             991..1136
FT                   /*tag=  c
FT                   /note= "exon XVII"
FT   cds             3764..3937
FT                   /*tag=  d
FT                   /note= "exon XVIII"
FT   polya_signal    7492..7497
FT                   /*tag=  e
PN   J04183394-A.
PD   30-JUN-1992.
PF   19-NOV-1990; 314548.
PR   19-NOV-1990; JP-314548.
PA   (NAKA/) NAKAMURA T.
PA   (TOYM ) TOYOBO KK.
DR   WPI; 92-265591/32.
DR   P-PSDB; R25692.
PT   Recombinant human hepatocyte growth factor and DNA encoding it -
PT   useful for diagnosis and treatment of hepatic disease and
PT   transgenic animal prepn.
```



```
PS   Disclosure; Page 22; 28pp; Japanese.
CC   This sequence contains exons XV to XVIII of human hepatocyte growth
CC   factor. See also R25676-92, Q26713-27.
SQ   Sequence   7753 BP;   2567 A;   1177 C;   1243 G;   2766 T;

  Query Match           9.4%;  Score 16;  DB 4;  Length 7753;
  Best Local Similarity 100.0%;  Pred. No. 5.49e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1467 atttattttatcttcc 1482
          ||||||||||||||||
Cp    161 ATTTATTTTATCTTCC 146


RESULT   21
ID   T56308 standard; RNA; 15 BP.
AC   T56308;
DT   14-MAY-1997  (first entry)
DE   Mouse TNF-a hammerhead ribozyme target sequence (nt position 1303).
KW   Enzymatic nucleic acid; ribozyme; trans cleavage; inhibition;
KW   gene expression; downregulation; interleukin-5; IL-5; ICAM-1;
KW   intercellular adhesion molecule; rel A; tumour necrosis factor;
KW   TNF-alpha; respiratory syncytial virus; RSV; bcr-abl; oncogene;
KW   translocation; chronic myelogenous leukaemia; CML; cancer;
KW   Philadelphia chromosome; inflammation; autoimmune disease;
KW   atherosclerosis; myocardial infarction; stroke; restenosis;
KW   transplant rejection; rheumatoid arthritis; psoriasis;
KW   myocardial ischaemia; Kawasaki disease; septic shock; HIV;
KW   human immunodeficiency virus; acquired immune deficiency syndrome;
KW   AIDS; ss.
OS   Mus musculus.
PN   WO9523225-A2.
PD   31-AUG-1995.
PF   23-FEB-1995; IB0156.
PR   23-FEB-1994; US-201109.
PR   29-MAR-1994; US-218934.
PR   04-APR-1994; US-222795.
PR   07-APR-1994; US-224483.
PR   15-APR-1994; US-228041.
PR   15-APR-1994; US-227958.
PR   18-MAY-1994; US-245736.
PR   06-JUL-1994; US-271280.
PR   15-AUG-1994; US-291932.
PR   16-AUG-1994; US-291433.
PR   17-AUG-1994; US-292620.
PR   19-AUG-1994; US-293520.
PR   02-SEP-1994; US-300000.
PR   08-SEP-1994; US-303039.
PR   23-SEP-1994; US-311749.
PR   23-SEP-1994; US-311486.
PR   28-SEP-1994; US-314397.
```



```
PR   03-OCT-1994; US-316771.
PR   07-OCT-1994; US-319492.
PR   11-OCT-1994; US-321993.
PR   04-NOV-1994; US-334847.
PR   10-NOV-1994; US-337608.
PR   28-NOV-1994; US-345516.
PR   16-DEC-1994; US-357577.
PR   23-DEC-1994; US-363233.
PR   30-JAN-1995; US-380734.
PA   (RIBO-) RIBOZYME PHARM INC.
PI   Stinchcomb DT,  Chowrira B,  Direnzo A,  Draper KG,  Dudycz LW;
PI   Grimm S,  Karpeisky A,  Kisich K,  Matulic-Adamic J;
PI   McSwiggen JA,  Modak A,  Pavco P,  Beigelman L,  Sullivan SM;
PI   Sweedler D,  Thompson JD,  Tracz D,  Usman N,  Wincott FE;
PI   Woolf T;
DR   WPI; 95-351090/45.
PT   Ribozymes having modified bases and methods for producing them  -
PT   for use in inhibiting disease related genes
PS   Claim 2; Page 252; 407pp; English.
CC   The present sequence represents a preferred target sequence for an
CC   enzymatic nucleic acid (i.e. a ribozyme) which cleaves TNF-alpha
CC   mRNA at the nucleotide base position indicated in the DE line.
CC   Regions of the mRNA that do not form secondary folding
CC   structures and that contain potential hammerhead and hairpin
CC   ribozyme cleavage sites were identified by computer analysis.
CC   Ribozymes directed against these mRNA sequences were designed and
CC   synthesised with modifications that improve their nuclease
CC   resistance. The ribozymes are designed to cleave the target
CC   sequences and thereby inhibit TNF-alpha expression, making them
CC   potentially useful for treating rheumatoid arthritis, septic shock
CC   and other inflammatory disorders including psoriasis, as well as
CC   for treatment of AIDS.
SQ   Sequence   15 BP;   4 A;   2 C;   0 G;   9 U;

  Query Match           8.8%;  Score 15;  DB 29;  Length 15;
  Best Local Similarity 40.0%;  Pred. No. 2.28e+01;
  Matches    6;  Conservative    9;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 cacuuauuauuuauu 15
          |||:|::|:::|::
Cp    169 CACTTATTATTTATT 155


RESULT   22
ID   Q88983 standard; RNA; 50 BP.
AC   Q88983;
DT   28-SEP-1995  (first entry)
DE   VEGF 2'-NH2-RNA nucleic acid ligand family 1, oligo 16B.
KW   Nucleic acid; ligand; thrombin; elastase; theophylline; caffeine;
KW   pharmaceutical; diagnosis; vascular endothelial growth factor;
KW   gene therapy; RNA; DNA; ss.
```

```
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   modified_base   1
FT                   /*tag= a
FT                   /note= "2'-NH2-Uracil"
PN   WO9507364-A.
PD   16-MAR-1995.
PF   08-SEP-1994; U10306.
PR   08-SEP-1993; US-117991.
PR   07-OCT-1993; US-134028.
PR   22-FEB-1994; US-199507.
PR   25-APR-1994; US-233012.
PR   28-APR-1994; US-234997.
PA   (NEXA-) NEXAGEN INC.
PI   Biesecker G,  Gold L,  Janjic N,  Jayasena S,  Jenison RD;
PI   Kirschenheuter GP,  Pieken W,  Polisky B,  Smith D,  Tasset D;
DR   WPI; 95-123436/16.
PT   Identifying nucleic acid ligands for target molecules - by
PT   partitioning increased affinity nucleic acids from a candidate
PT   mixt. and amplifying
PS   Claim 37; Fig 34; 251pp; English.
CC   The sequences given in Q88982-91 represent nucleic acid ligands to
CC   vascular endothelial growth factor (VEGF).  These ligands all
CC   contain the consensus sequence given in Q88992.  These ligands were
CC   identified using the method of the invention.  The method comprises
CC   contacting a candidate mixture with the target molecule (i.e. VEGF)
CC   where the nucleic acids which have an increased affinity to the target
CC   relative to the candidate mixture can be partitioned from the remainder
CC   of the candidate mixture.  The increased affinity nucleic acids are
CC   partitioned from the remainder of the candidate mixture and the isolated
CC   nucleic acids are amplified to yield a ligand-enriched mixture of
CC   nucleic acids, in which the nucleic acid ligands can be identified.
CC   The isolated ligands may be used as pharmaceuticals, diagnostic agents
CC   and in gene therapy.  The ligands may be RNA or DNA molecules.
SQ   Sequence   50 BP;    8 A;    7 C;    22 G;    13 U;

  Query Match           8.8%;  Score 15;  DB 14;  Length 50;
  Best Local Similarity  73.3%;  Pred. No. 2.28e+01;
  Matches   11;  Conservative    4;  Mismatches   0;  Indels   0;  Gaps   0;

Db     25 uggaugggagucugg 39
          :|||:|||||:|:||
Qy    119 TGGATGGGAGTCTGG 133


RESULT   23
ID   T23517 standard; cDNA to mRNA; 247 BP.
AC   T23517;
DT   02-SEP-1996  (first entry)
DE   Human gene signature HUMGS05362.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
```



```
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 1388; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   247 BP;    97 A;    24 C;    64 G;    49 T;

  Query Match           8.8%;  Score 15;  DB 20;  Length 247;
  Best Local Similarity  88.2%;  Pred. No. 2.28e+01;
  Matches   15;  Conservative    0;  Mismatches   2;  Indels   0;  Gaps   0;

Db    230 ggaaaattngaagntaaa 246
          ||||||| |||| ||||
Qy    139 GGAAATTGGAAGATAAA 155


RESULT   24
ID   T20096 standard; cDNA to mRNA; 315 BP.
AC   T20096;
DT   26-JUL-1996  (first entry)
DE   Human gene signature HUMGS01239.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
```



```
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 559; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   315 BP;    63 A;    93 C;    79 G;    69 T;

  Query Match           8.8%;  Score 15;  DB 19;  Length 315;
  Best Local Similarity 100.0%;  Pred. No. 2.28e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    103 cctcctcctgtcgtg 117
          |||||||||||||||
Qy     42 CCTCCTCCTGTCGTG 56


RESULT   25
ID   Q91404 standard; DNA; 648 BP.
AC   Q91404;
DT   11-MAR-1996  (first entry)
DE   Human olfactory receptor 17-228 gene.
KW   Olfactory receptor; human; chromosome library; probe; rat; amplification;
KW   primer; PCR; transmembrane region; cosmid mapping; gene cluster; signal;
KW   polymorphism; ligand binding; fragrance; aroma; flavour; ss.
OS   Homo sapiens.
PN   WO9518140-A1.
PD   06-JUL-1995.
PF   28-DEC-1994; U14554.
```



```
PR   28-DEC-1993; IL-108205.
PA   (RYCU/) RYCUS A.
PA   (YEDA ) YEDA RES & DEV CO LTD.
PI   Ben-Arie N,  Lancet D;
DR   WPI; 95-246331/32.
DR   P-PSDB; R75902.
PT   Novel DNA mols. encoding olfactory receptor (OR) genes - for use in
PT   studying olfactory polymorphism in a population or for detecting
PT   inter-individual differences in OR's or olfactory clusters in humans
PS   Claim 1; Fig 4e; 46pp; English.
CC   Q91400-415 are sequences encoding olfactory receptors from humans.
CC   This sequence is from the gene designated 17-228. The genes were isolated
CC   from a flow-sorted human chromosome library using, as probes, regions of
CC   the human and rat genome amplified by primers Q91416-8, which were
CC   designed on the second and seventh transmembrane region of the rat
CC   olfactory receptor gene sequence. By cosmid mapping the sequences were
CC   shown to be present at chromosome location 17q13, with several as gene
CC   clusters.  The nucleic acid sequences of the receptors can be used as
CC   probes for the study of olfactory receptor polymorphisms in a population
CC   or for detecting inter-individual differences in olfactory receptors or
CC   olfactory clusters. The proteins can be used for probing for the presence
CC   of a certain fragrance, aroma or flavour in a medium by detecting a
CC   signal upon ligand binding on an immobilised protein.
SQ   Sequence   648 BP;    121 A;    198 C;    154 G;    175 T;

  Query Match           8.8%;  Score 15;  DB 16;  Length 648;
  Best Local Similarity 100.0%;  Pred. No. 2.28e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    105 cttcttccaccttct 119
          |||||||||||||||
Cp    103 CTTCTTCCACCTTCT 89


RESULT   26
ID   Q91405 standard; DNA; 648 BP.
AC   Q91405;
DT   11-MAR-1996  (first entry)
DE   Human olfactory receptor 17-207 gene.
KW   Olfactory receptor; human; chromosome library; probe; rat; amplification;
KW   primer; PCR; transmembrane region; cosmid mapping; gene cluster; signal;
KW   polymorphism; ligand binding; fragrance; aroma; flavour; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   misc_difference 337..339
FT                   /*tag= a
FT                   /transl_except= seq: GAC,  a.a.: Glu
FT   misc_difference 625..627
FT                   /*tag= b
FT                   /transl_except= seq: ATT,  a.a.: Val
PN   WO9518140-A1.
```

```
PD   06-JUL-1995.
PF   28-DEC-1994; U14554.
PR   28-DEC-1993; IL-108205.
PA   (RYCU/) RYCUS A.
PA   (YEDA ) YEDA RES & DEV CO LTD.
PI   Ben-Arie N,  Lancet D;
DR   WPI; 95-246331/32.
DR   P-PSDB; R75903.
PT   Novel DNA mols. encoding olfactory receptor (OR) genes - for use in
PT   studying olfactory polymorphism in a population or for detecting
PT   inter-individual differences in OR's or olfactory clusters in humans
PS   Claim 1; Fig 4f; 46pp; English.
CC   Q91400-415 are sequences encoding olfactory receptors from humans.
CC   This sequence is from the gene designated 17-207. The genes were isolated
CC   from a flow-sorted human chromosome library using, as probes, regions of
CC   the human and rat genome amplified by primers Q91416-8, which were
CC   designed on the second and seventh transmembrane region of the rat
CC   olfactory receptor gene sequence. By cosmid mapping the sequences were
CC   shown to be present at chromosome location 17q13, with several as gene
CC   clusters.  The nucleic acid sequences of the receptors can be used as
CC   probes for the study of olfactory receptor polymorphisms in a population
CC   or for detecting inter-individual differences in olfactory receptors or
CC   olfactory clusters. The proteins can be used for probing for the presence
CC   of a certain fragrance, aroma or flavour in a medium by detecting a
CC   signal upon ligand binding on an immobilised protein.
SQ   Sequence   648 BP;    126 A;    196 C;    144 G;    182 T;

  Query Match           8.8%;  Score 15;  DB 16;  Length 648;
  Best Local Similarity 100.0%;  Pred. No. 2.28e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    105 cttcttccaccttct 119
          |||||||||||||||
Cp    103 CTTCTTCCACCTTCT 89


RESULT   27
ID   Q91406 standard; DNA; 648 BP.
AC   Q91406;
DT   11-MAR-1996  (first entry)
DE   Human olfactory receptor 17-201 gene.
KW   Olfactory receptor; human; chromosome library; probe; rat; amplification;
KW   primer; PCR; transmembrane region; cosmid mapping; gene cluster; signal;
KW   polymorphism; ligand binding; fragrance; aroma; flavour; ss.
OS   Homo sapiens.
PN   WO9518140-A1.
PD   06-JUL-1995.
PF   28-DEC-1994; U14554.
PR   28-DEC-1993; IL-108205.
PA   (RYCU/) RYCUS A.
PA   (YEDA ) YEDA RES & DEV CO LTD.
```



```
PI   Ben-Arie N,  Lancet D;
DR   WPI; 95-246331/32.
DR   P-PSDB; R75904.
PT   Novel DNA mols. encoding olfactory receptor (OR) genes - for use in
PT   studying olfactory polymorphism in a population or for detecting
PT   inter-individual differences in OR's or olfactory clusters in humans
PS   Claim 1; Fig 4g; 46pp; English.
CC   Q91400-415 are sequences encoding olfactory receptors from humans.
CC   This sequence is from the gene designated 17-201. The genes were isolated
CC   from a flow-sorted human chromosome library using, as probes, regions of
CC   the human and rat genome amplified by primers Q91416-8, which were
CC   designed on the second and seventh transmembrane region of the rat
CC   olfactory receptor gene sequence. By cosmid mapping the sequences were
CC   shown to be present at chromosome location 17q13, with several as gene
CC   clusters.  The nucleic acid sequences of the receptors can be used as
CC   probes for the study of olfactory receptor polymorphisms in a population
CC   or for detecting inter-individual differences in olfactory receptors or
CC   olfactory clusters. The proteins can be used for probing for the presence
CC   of a certain fragrance, aroma or flavour in a medium by detecting a
CC   signal upon ligand binding on an immobilised protein.
SQ   Sequence   648 BP;    115 A;    199 C;    161 G;    173 T;

  Query Match           8.8%;  Score 15;  DB 16;  Length 648;
  Best Local Similarity 100.0%;  Pred. No. 2.28e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    105 cttcttccaccttct 119
          |||||||||||||||
Cp    103 CTTCTTCCACCTTCT 89


RESULT   28
ID   T29191 standard; cDNA; 1400 BP.
AC   T29191;
DT   20-NOV-1996  (first entry)
DE   Human urokinase plasminogen activator receptor cDNA.
KW   Antibody; urokinase; plasminogen activator; receptor; u-PAR; human;
KW   detection; quantification; targeted drug delivery; cancer cells;
KW   therapeutic inhibition; receptor binding; local protease activity;
KW   drug screening; purification; affinity chromatography; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             47..1054
FT                   /*tag=  a
FT   signal_peptide  47..112
FT                   /*tag=  b
FT   mat_peptide     113..1054
FT                   /*tag=  c
PN   US5519120-A.
PD   21-MAY-1996.
PF   07-APR-1989; 334613.
```



```
PR   07-APR-1989; US-334613.
PR   03-JUL-1989; US-374854.
PR   09-APR-1990; WO-DK0090.
PR   06-DEC-1991; US-824189.
PR   17-JUN-1993; US-085122.
PA   (CANC-) CANCERFORSKNINGSFONDET AF 1989.
PI   Behrendt N,  Bruenner N,  Dano K,  Ellis V,  Hoyer-Hansen G;
PI   Pyke C,  Ronne E;
DR   WPI; 96-259120/26.
DR   P-PSDB; R97612.
PT   Antibodies specific for urokinase plasminogen activator receptor -
PT   useful for e.g. receptor quantification, drug delivery; inhibition
PT   of receptor-activator binding e.g. in cancer etc.
PS   Disclosure; Columns 99-104; 90pp; English.
CC   Antibodies (Ab) against urokinase plasminogen activator receptors
CC   (u-PAR), e.g. the human u-PAR encoded by the present sequence, are
CC   claimed. The Ab pref. bind the u-PA binding domain of u-PAR, but
CC   not u-PA/u-PAR complexes. Four monoclonal Ab produced by the
CC   hybridomas I 4E-8, I 5A-7, I 10A-4 and II 1D-6, designated ECACC
CC   90101007-10 respectively, are specified.
CC   The Ab can be used to detect/quantify u-PAR, target diagnostic or
CC   therapeutic agents to cells having u-PAR on their surface (e.g.
CC   cancer cells), therapeutically inhibit u-PA/receptor binding,
CC   block local protease activity, screen drugs for their ability to
CC   inhibit u-PA/u-PAR interaction and purify u-PAR by affinity
CC   chromatography.
SQ   Sequence   1400 BP;    346 A;    396 C;    366 G;    292 T;

  Query Match           8.8%;  Score 15;  DB 23;  Length 1400;
  Best Local Similarity 100.0%;  Pred. No. 2.28e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1322 tattatttattttat 1336
          |||||||||||||||
Cp    165 TATTATTTATTTTAT 151


RESULT   29
ID   Q06354 standard; DNA; 1400 BP.
AC   Q06354;
DT   06-FEB-1991  (first entry)
DE   Recombinant urokinase-type plasminogen activator receptor.
KW   Recombinant urokinase-type plasminogen activator receptor; antibody;
KW   u-PAR; cancer therapy; metastasis; p-uPAR-1; woundhealing; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             47..1051
FT                   /*tag=  a
FT                   /product=u-PAR
FT   signal_peptide  47..112
FT                   /*tag=  b
```



```
FT   misc_rna        956..1006
FT                   /*tag=  c
FT                   /label=transmembrane-domain
PN   WO9012091-A.
PD   18-OCT-1990.
PF   09-APR-1990; DK0090.
PR   07-APR-1989; US-334613.
PR   03-JUL-1989; US-374854.
PA   (CANC-) CANCERFORSKNINGSFON.
PI   Dano K, Blasi F, Roldan A, Cubellis M, Masucci M, Appella E;
PI   Schleuning W, Behrendt N;
DR   WPI; 90-334842/44.
DR   P-PSDB; R07561.
PT   Recombinant urokinase-type plasminogen activator receptor -
PT   useful for antibody generation e.g. for inhibiting cell surface
PT   plasminogen activation
PS   Disclosure; Page 34-37; 278pp; English.
CC   On the basis of the amino-terminal amino acid sequence of pure
CC   u-PAR, a 24-mer nucleotide probe (Q06355) was synthesized and used to
CC   screen a library to identify and isolate recombinant clones carrying
CC   the cDNA for u-PAR.  The identity of the cDNA clones was confirmed by
CC   comparing the nucleotide sequence with the amino terminal sequence
CC   of the purified u-PAR, and by expressing the cDNA in mouse L cells
CC   and assaying their u-PA-binding activity.  The sequence
CC   represented here is from clone p-uPAR-1.
CC   The inhibition of receptor binding of u-PA as a means of inhibiting
CC   some of its physiological functions has a variety of therapeutic
CC   applicns, e.g.  prevention of invasion and metastasis of cancer cells,
CC   and treatment of inflammatory bowel disease, premalignant colonic
CC   adenomas, septic arthritis, osteoarthritis, rheumatoid arthritis,
CC   osteoporosis, cholesteatoma, corneal ulcers, keratitis,
CC   epidermolysis bullosa, psoriasis and pemphigus.  Since u-PAR are
CC   also present in blood and endothelial cells, their regulation might
CC   also affect intravascular fibrinolytic activity.  A contraceptive
CC   effect is also expected of measures that inhibit receptor binding.
CC   The protein may also be used to treat disorders related to a
CC   reduced amt. or an impaired function of u-PAR, e.g. impaired wound
CC   healing and some cases of thromboembolic disorders.
CC   See also Q06898.
SQ   Sequence   1400 BP;    346 A;    396 C;    366 G;    292 T;

  Query Match           8.8%;  Score 15;  DB 1;  Length 1400;
  Best Local Similarity 100.0%;  Pred. No. 2.28e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1322 tattatttattttat 1336
          |||||||||||||||
Cp    165 TATTATTTATTTTAT 151


RESULT   30
```

```
ID   T89675 standard; cDNA; 1407 BP.
AC   T89675;
DT   02-MAR-1998  (first entry)
DE   Human phospholipase inhibitory protein (hPIP) encoding cDNA.
KW   phospholipase inhibitory protein; hPIP; human; phospholipase A2; PA2;
KW   antibody; inflammation; apolipoprotein IV; apoAIV; diagnosis; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             46..1053
FT                   /*tag= a
FT                   /product= phospholipase_inhibitory_protein
FT   mat_peptide     112..1051
FT                   /*tag= b
FT   misc_feature    490..852
FT                   /*tag= c
FT                   /note= "active PIP fragment"
PN   US5646254-A.
PD   08-JUL-1997.
PF   26-APR-1994; 233464.
PR   09-JUL-1986; US-883598.
PR   15-APR-1985; US-723046.
PR   14-JUN-1990; US-538692.
PR   20-APR-1992; US-871577.
PR   01-SEP-1993; US-115316.
PR   26-APR-1994; US-233464.
PA   (SCIO-) SCIOS NOVA INC.
PI   Johnson LK,  Longenecker JP;
DR   WPI; 97-362999/33.
DR   P-PSDB; W31165.
PT   Antibodies to human phospholipase inhibitory protein - useful as
PT   immunoassay reagents
PS   Claim 1; Fig 13A-B; 34pp; English.
CC   This DNA encodes for human phospholipase inhibitory protein (hPIP) having
CC   a defined sequence of 335 amino acids. This protein has a phospholipase
CC   A2 (PA2) inhibiting activity. Pure hPIP can be derived by recombinant
CC   techniques or from the human peritoneal dialysis fluid by various
CC   purification procedures. Antibodies that can react with a hPIP/
CC   apolipoprotein IV (apoAIV) complex are used in immunoassays for
CC   diagnostic and therapy monitoring purposes, in the diagnosis and
CC   treatment of inflammatory conditions.
SQ   Sequence   1407 BP;    348 A;    397 C;    366 G;    296 T;

  Query Match           8.8%;  Score 15;  DB 36;  Length 1407;
  Best Local Similarity 100.0%;  Pred. No. 2.28e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1321 tattatttattttat 1335
         ||||||||||||||||
Cp    165 TATTATTTATTTTAT 151
```



```
RESULT   31
ID   Q71647 standard; DNA; 1407 BP.
AC   Q71647;
DT   28-MAR-1995  (first entry)
DE   Human phospholipase inhibitory protein.
KW   Phospholipase inhibitory protein; PIP; cloning; expression; probe;
KW   anti-inflammatory; inflammation; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             46..1053
FT                   /*tag= a
PN   US5342764-A.
PD   30-AUG-1994.
PF   15-APR-1985; 723046.
PR   15-APR-1985; US-723046.
PR   09-JUL-1986; US-883598.
PR   14-JUN-1990; US-538692.
PR   20-APR-1992; US-871577.
PR   01-SEP-1993; US-115316.
PA   (SCIO-) SCIOS NOVA INC.
PI   Johnson LK,  Longenecker JP;
DR   WPI; 94-278932/34.
DR   P-PSDB; R58707.
PT   Recombinant prodn. of human phospholipase inhibitory protein -
PT   useful for treating inflammation in e.g. infections, wounds,
PT   arthritis, asthma, dermatitis, conjunctivitis or allergies.
PS   Claim 1; Fig 13; 36pp; English.
CC   Cloning and expression of the human PIP gene permits prodn. of an
CC   anti-inflammatory protein in practical quantities using recombinant
CC   techniques.  The probes given in Q71651-52 were used in the
CC   isolation of the gene.
SQ   Sequence   1407 BP;    348 A;    397 C;    366 G;    296 T;

  Query Match           8.8%;  Score 15;  DB 12;  Length 1407;
  Best Local Similarity 100.0%;  Pred. No. 2.28e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1321 tattatttattttat 1335
         ||||||||||||||||
Cp    165 TATTATTTATTTTAT 151


RESULT   32
ID   N60374 standard; cDNA; 1407 BP.
AC   N60374;
DT   25-JUN-1991  (first entry)
DE   Sequence encoding human phospholipase inhibitory
DE   protein (hPIP) in pLE-1.
KW   Anti-inflammatory; phospholipase inhibitor; inflammatory disorder;
KW   therapy; ss.
OS   Homo sapiens.
```



```
FH   Key             Location/Qualifiers
FT   cds             46..1053
FT                   /*tag= a
PN   WO8606100-A.
PD   23-OCT-1986.
PF   14-APR-1986; U00772.
PR   15-APR-1985; US-723046.
PA   (BIOT-) BIOTECHN RES PARTN.
PI   Johnson LF, Longenecker J;
DR   WPI; 86-291667/44.
DR   P-PSDB; P60436.
PT   Human antiinflammatory phospholipase inhibitor protein - obtd. by
PT   purificn. of human inter:peritoneal dialysis fluid
PS   Claim 15; Fig 13; 61pp; English.
CC   Lambda U200 was used to probe cDNA libraries for hPIP. Lambda U200
CC   and lambda U500 share approximately 50 bp at the 5' ends of each and
CC   then diverge. This divergence was due to intron splicing. PHN600 not
CC   only shares the 50 bp segment common to lambda U200 and lambda U500,
CC   but its homology with lambda U500 extends for another approximately
CC   200 bp. These sequences then also diverge. This divergence was shown
CC   to be due to the presence of optional 3' terminal exons in the gene.
CC   pLE-1 contains the full-length coding sequence for hPIP. The hPIP
CC   can be used for producing antibodies (claimed). The PIP proteins
CC   show in vitro activity as phospholipase inihibitors. The proteins are
CC   useful both to control inflammatory responses and to treat
CC   inflammatory disorders.
SQ   Sequence   1407 BP;    348 A;    397 C;    366 G;    296 T;

  Query Match           8.8%;  Score 15;  DB 2;  Length 1407;
  Best Local Similarity 100.0%;  Pred. No. 2.28e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1321 tattatttattttat 1335
         ||||||||||||||||
Cp    165 TATTATTTATTTTAT 151


RESULT   33
ID   Q52515 standard; DNA; 1407 BP.
AC   Q52515;
DT   02-JUN-1994  (first entry)
DE   Sequence encoding human phospholipase inhibitor protein.
KW   Human phospholipase inhibitor protein; hPIP; inflammation;
KW   rheumatoid arthritis; conjunctivitis; allergies;
KW   lupus erythomatosis; immunoassay; diagnosis; monitoring; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             46..1053
FT                   /*tag= a
FT                   /product= Human phospholipase inhibitory protein.
PN   US5264550-A.
```



```
PD   23-NOV-1993.
PF   15-APR-1985; 723046.
PR   15-APR-1985; US-723046.
PR   09-JUL-1986; US-883598.
PR   14-JUN-1990; US-538692.
PR   20-APR-1992; US-871577.
PA   (SCIO-) SCIOS NOVA INC.
PI   Johnson LK,  Longenecker JP;
DR   WPI; 93-385719/48.
DR   P-PSDB; R44424.
PT   Purified human phospholipase inhibitor protein - obtd. from
PT   dialysis fluid or by recombinant methods, useful for treating
PT   inflammation, in infection, rheumatoid arthritis, allergy,
PT   injury, etc.
PS   Claim 6; Figure 13; 36pp; English.
CC   The human phospholipase inhibitor protein (hPIP) is capable of
CC   suppressing the production of PGE2 in cell culture and of combating
CC   inflammation in in vivo models.  Applications include treatment of
CC   inflammation due to infection or wounding and treatment of
CC   inflammatory disorders.  Purified hPIP is also useful to prepare
CC   antisera or monoclonal antibodies for immunoassays useful in
CC   diagnosis and in monitoring of therapy.
SQ   Sequence   1407 BP;    348 A;    397 C;    366 G;    296 T;

  Query Match           8.8%;  Score 15;  DB 9;  Length 1407;
  Best Local Similarity 100.0%;  Pred. No. 2.28e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1321 tattatttattttat 1335
         ||||||||||||||||
Cp    165 TATTATTTATTTTAT 151


RESULT   34
ID   T15423 standard; cDNA; 1644 BP.
AC   T15423;
DT   23-APR-1996  (first entry)
DE   Mouse tumour necrosis factor cDNA.
KW   Tumour necrosis factor; TNF; phage T4; phage lambda; pL promoter;
KW   antitumour; anticancer; antimalarial; ss.
OS   Mus musculus.
FH   Key             Location/Qualifiers
FT   misc_difference 156
FT                   /*tag= a
FT                   /note= "base n at position 156 is given as 'o' in
FT                   the specification"
FT   cds             157..864
FT                   /*tag= b
FT   signal_peptide  157..393
FT                   /*tag= c
FT   mat_peptide     394..861
```

```
PT                      /*tag= d
FT   polya_signal       1614..1619
FT                      /*tag= e
FT   polya_signal       1630..1635
FT                      /*tag= f
PN   US5487984-A.
PD   30-JAN-1996.
PF   21-DEC-1984; 684595.
PR   21-DEC-1984; US-684595.
PR   09-OCT-1985; US-785847.
PR   20-DEC-1985; US-811654.
PA   (BIOJ ) BIOGEN INC.
PI   Allet B, Kawashima EH;
DR   WPI; 96-105230/11.
DR   P-PSDB; R88589.
PT   Prodn. of tumour necrosis factor - using recombinant DNA encoding
PT   TNF under the control of T4 or lambda pL-T4 expression control
PT   sequences.
PS   Example 9; Fig 7; 43pp; English.
CC   The complete cDNA sequence (T15423) coding for mouse tumour
CC   necrosis factor (mTNF) precursor (R88589) was obtd. by sequencing
CC   2 cDNA clones, p-mTNF-1 and p-mTNF-3 (DSM 3159), isolated using
CC   probes based on rabbit TNF sequences.  The cDNA may be linked
CC   to expression control sequences from phage T4 and phage lambda
CC   (see T15402-05) for expression in host cells, esp. Escherichia
CC   coli, for prodn. of recombinant mTNF.
SQ   Sequence   1644 BP;    399 A;    476 C;    375 G;    393 T;

  Query Match           8.8%;  Score 15;  DB 17;  Length 1644;
  Best Local Similarity 100.0%;  Pred. No. 2.28e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1296 cacttattatttatt 1310
         |||||||||||||||
Cp    169 CACTTATTATTTATT 155


RESULT   35
ID   N60445 standard; cDNA; 1644 BP.
AC   N60445;
DT   07-AUG-1991  (first entry)
DE   Sequence encoding tumour necrosis factor (TNF).
KW   Anticancer agent; antitumour; antimalarial; tumour necrosis factor;
KW   ss.
OS   Mus musculus
FH   Key             Location/Qualifiers
FT   cds             157..864
FT                   /*tag= a
FT   polya_signal    1614..1619
FT                   /*tag= b
FT   polya_signal    1630..1635
```



```
PT                      /*tag= c
PN   WO8603751-A.
PD   03-JUL-1986.
PF   19-DEC-1985; E00721.
PR   21-DEC-1984; US-684595.
PR   09-OCT-1985; US-785847.
PR   09-OCT-1986; WO-U02133.
PA   (BIOJ ) BIOGEN NV.
PA   (FIER/) FIERS W C.
PA   (ALLE/) ALLET B.
PA   (BIOG-) BIOGEN INC.
PI   Fiers WC, Fransen LM, Tavernier JHL, Marmenout ALM,
PI   VanDerHeyden J, Allet B, Washima EH;
DR   WPI; 86-182891/28.
DR   P-PSDB; P60530.
PT   Mammalian tumour necrosis factors - produced by culturing
PT   pro-karyotic hosts transformed with recombinant DNA
PS   Example; Fig 7; 93pp; English.
CC   TNF-like polypeptides and compsns. are produced by the fermentation
CC   of host cells transformed with at least one DNA sequence which codes
CC   for a mammalian TNF-like polypeptide operatively linked to an
CC   expression control sequence in the transformed host.
SQ   Sequence   1644 BP;    399 A;    476 C;    375 G;    393 T;

  Query Match           8.8%;  Score 15;  DB 2;  Length 1644;
  Best Local Similarity 100.0%;  Pred. No. 2.28e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1296 cacttattatttatt 1310
         |||||||||||||||
Cp    169 CACTTATTATTTATT 155


RESULT   36
ID   Q55058 standard; cDNA; 1685 BP.
AC   Q55058;
DT   20-JUL-1994  (first entry)
DE   Sequence encoding amino acid transporter (AAT).
KW   Amino acid transporter; AAt; plant transformation; ss.
OS   Arabidopsis thaliana.
FH   Key             Location/Qualifiers
FT   cds             57..1511
FT                   /*tag= a
PN   DE4222315-A.
PD   13-JAN-1994.
PF   05-JUL-1992; 222315.
PR   05-JUL-1992; DE-222315.
PA   (GENB-) INST GENBIOLOGISCHE FORSCHUNG.
PI   Frommer W;
DR   WPI; 94-017036/03.
DR   P-PSDB; R47812.
```



```
PT   DNA coding for aminoacid transporter proteins - used for
PT   modifying transporter activity in plants
PS   Claim 2; Page 12-14; 15pp; German.
CC   cDNA from 2-leaf seedlings of A. thaliana was inserted into pFL61
CC   and used to transform the Pro-transport deficient yeast strain 22574d
CC   Plasmid pFL61-pppp1-20 was isolated from a transformant capable of
CC   growing with Pro as the sole N source. Q55058 is the coding sequence
CC   in pFL61. A notI fragment from this was inserted into p-Bluescript
CC   SK to obtain pPPPl-20  (DSM 7130).
SQ   Sequence   1685 BP;    433 A;    382 C;    371 G;    499 T;

  Query Match           8.8%;  Score 15;  DB 9;  Length 1685;
  Best Local Similarity 100.0%;  Pred. No. 2.28e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      9 atttatttttatcttc 23
          |||||||||||||||
Cp    161 ATTTATTTTATCTTC 147


RESULT   37
ID   T77840 standard; mRNA; 1735 BP.
AC   T77840;
DT   17-MAR-1998  (first entry)
DE   Human melanoma associated delayed early response variant mRNA sequence.
KW   Melanoma associated delayed early response gene; MADER gene;
KW   MADER protein; growth alteration; malignant melanoma; breast carcinoma;
KW   cancerous condition; MADER translocation event; MADER immunogen;
KW   MADER antigen; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             158..1735
FT                   /*tag= a
PN   WO9728193-A1.
PD   07-AUG-1997.
PF   30-JAN-1997; U01586.
PR   30-JAN-1997; US-593563.
PR   30-JAN-1996; US-593563.
PA   (MELC-) MELCORP DIAGNOSTICS INC.
PI   Johnson JP;
DR   WPI; 97-402557/37.
DR   P-PSDB; W24230.
PT   Monoclonal antibody which binds to the MADER protein - used for
PT   detecting cancerous conditions, such as breast carcinoma
PS   Disclosure; Fig 3; 69pp; English.
CC   The present sequence represents a melanoma associated delayed early
CC   response (MADER) gene, which encodes a novel 55 kDa nuclear protein. This
CC   protein is associated with growth alterations in malignant melanomas and
CC   other cancerous conditions, and is over-expressed in human malignant
CC   melanomas. Several variants of the protein have been identified
CC   (W24228-31), the present seqeunce encoding a variant which binds
```



```
CC   erg-1 and inhibits its activity. Chromosomal rearrangement of MADER can
CC   be detected by hybridising immobilised chromosomal target DNA, that has
CC   been rendered single stranded and is obtained from a cell suspected of
CC   having undergone a MADER translocation event with a single stranded
CC   oligonucleotide probe complementary to a MADER nucleotide sequence. The
CC   probe contains a moiety capable of direct or indirect visualisation.
CC   Antibodies raised against the MADER protein can be used for detecting a
CC   cancerous condition, particularly melanoma malignancies, and especially
CC   a breast carcinoma. A composition comprising a MADER immunogen and a
CC   pharmaceutically acceptable vehicle can be used to elicit an immune
CC   response against a cell which over-expresses a MADER antigen.
SQ   Sequence   1735 BP;    357 A;    516 C;    594 G;    268 T;

  Query Match           8.8%;  Score 15;  DB 36;  Length 1735;
  Best Local Similarity 100.0%;  Pred. No. 2.28e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     63 tggacacggcatcgt 77
          |||||||||||||||
Cp     22 TGGACACGGCATCGT 8


RESULT   38
ID   V15514 standard; DNA; 1827 BP.
AC   V15514;
DT   22-MAY-1998  (first entry)
DE   Hamster oral papilloma virus E1 DNA.
KW   E1 DNA; E1 protein; diagnosis; papilloma virus infectious disease;
KW   cancer; screening; antitumour agent; antisense treatment; ss.
OS   Hamster oral papilloma virus.
PN   J10042875-A.
PD   17-FEB-1998.
PF   06-AUG-1996; 207143.
PR   06-AUG-1996; JP-207143.
PA   (TORA ) TORAY IND INC.
DR   WPI; 98-186869/17.
DR   P-PSDB; W47220.
PT   DNA encoding hamster oral papilloma virus protein - useful for, e.g.
PT   diagnosing cancer related papilloma virus infections
PS   Claim 2; Page 12; 17pp; Japanese.
CC   The present sequence is the hamster oral papilloma virus E1 DNA.
CC   The E1 protein and DNA can be used for the diagnosis of papilloma
CC   virus infectious diseases related to cancer, screening for
CC   antitumour agents and antisense treatment.
SQ   Sequence   1827 BP;    549 A;    361 C;    453 G;    464 T;

  Query Match           8.8%;  Score 15;  DB 39;  Length 1827;
  Best Local Similarity 93.8%;  Pred. No. 2.28e+01;
  Matches   15;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db   1209 agtaatagcagccttc 1224
```

```
          ||| ||||||||||||
Cp    115 AGTNATAGCAGCCTTC 100


RESULT   39
ID   Q98564 standard; DNA; 1849 BP.
AC   Q98564;
DT   21-DEC-1995  (first entry)
DE   N. crassa cellobiohydrolase-1 gene.
KW   Cellobiohydrolase-1; CBH-1; cellulase complex; promoter;
KW   cellulose binding domain; ds.
OS   Neurospora crassa.
FH   Key             Location/Qualifiers
FT   misc_feature    12..17
FT                   /*tag= a
FT                   /note= "consensus sequence"
FT   misc_feature    63..70
FT                   /*tag= b
FT                   /note= "consensus sequence"
FT   misc_feature    100..124
FT                   /*tag= c
FT                   /note= "consensus sequence"
FT   cds             152..1761
FT                   /*tag= d
FT   exon            154..832
FT                   /*tag= e
FT   misc_feature    833..837
FT                   /*tag= f
FT                   /note= "consensus sequence"
FT   misc_feature    874..880
FT                   /*tag= g
FT                   /note= "consensus sequence"
FT   misc_feature    889..891
FT                   /*tag= h
FT   intron          843..891
FT                   /*tag= i
FT   exon            892..1761
FT                   /*tag= j
PN   WO9519441-A1.
PD   20-JUL-1995.
PF   11-JAN-1995; G00049.
PR   14-JAN-1994; GB-000623.
PA   (UYLE-) UNIV LEEDS.
PI   Parish JH,  Radford A;
DR   WPI; 95-263871/34.
DR   P-PSDB; R80279.
PT   New regulated promoter from Neurospora crassa cello:bio:hydrolase-1
PT   gene - and constructs etc. contg. it, and transformed filamentous
PT   fungi, provides increased prodn. of cellulolytic enzymes or of
PT   heterologous peptide(s), e.g. hormones
PS   Claim 1; Fig.1; 32pp; English.
```



```
CC   The N. crassa cbh-1 gene was isolated by probing a genomic library
CC   with a PCR product generated using primers based on known CBH
CC   sequences.  The regulated promoter of the gene can be used to
CC   achieve high yields of glycosylated heterologous proteins in N.
CC   crassa, and the gene can also be used for CBH-1 manufacture.
SQ   Sequence   1849 BP;    385 A;    601 C;    469 G;    394 T;

  Query Match             8.8%;  Score 15;  DB 15;  Length 1849;
  Best Local Similarity 100.0%;  Pred. No. 2.28e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1610 agccttcttccacct 1624
          ||||||||||||||
Cp    106 AGCCTTCTTCCACCT 92


RESULT   40
ID   N81254 standard; cDNA; 2444 BP.
AC   N81254;
DT   05-OCT-1990  (first entry)
DE   cDNA sequence in plasmid pT3BP-3.
KW   Plasmid pT3BP-3; protein disulphide isomerase.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             50..1576
FT                   /*tag= a
PN   EP-293793-A.
PD   07-DEC-1988.
PF   28-MAY-1988; 108570.
PR   01-JUN-1987; JP-138628.
PR   18-SEP-1987; JP-235492.
PR   05-OCT-1987; JP-251144.
PA   (TAKE) Takeda Chemical Industries Ltd.
PI   Toyoshima K, Horiushi R, Yamauchi K, Yamamoto T, Igarashi K;
DR   WPI; 88-347496/49.
DR   P-PSDB; P80664.
PT   New polypeptide with protein disulphide isomerase activity -
PT   ensuring natural disulphide bridge arrangement in recombinant
PT   proteins, and new DNA encoding sequences.
PS   Disclosure; -(2); 37pp; English.
CC   The cDNA encodes a polypeptide with protein disulphide isomerase
CC   activity.  The cDNA is cloned in plasmid pT3BP-3.
SQ   Sequence   2444 BP;    559 A;    680 C;    705 G;    500 T;

  Query Match             8.8%;  Score 15;  DB 1;  Length 2444;
  Best Local Similarity 100.0%;  Pred. No. 2.28e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2220 cctcctcctgtcgtg 2234
          ||||||||||||||
Qy     42 CCTCCTCCTGTCGTG 56
```



```
RESULT   41
ID   Q58987 standard; cDNA to mRNA; 2454 BP.
AC   Q58987;
DT   26-OCT-1994  (first entry)
DE   Human PDI.
KW   Protein disulphide isomerase; PDI; human serum albumin; HSA;
KW   expression; fusion; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             49..1575
FT                   /*tag= a
FT   signal_peptide  49..99
FT                   /*tag= b
PN   J05038771-A.
PD   15-FEB-1994.
PF   18-APR-1991; 114074.
PR   31-OCT-1990; JP-295017.
PA   (TOFU ) TONEN CORP.
DR   WPI; 94-111888/14.
DR   P-PSDB; R51696.
PT   Expression of human protein di:sulphide isomerase gene - used to
PT   prepare polypeptide in high yield
PS   Disclosure; Page 13-15; 30pp; Japanese.
CC   Human PDI can be produced in large amts. by expression
CC   of a vector contg. human PDI DNA linked to HSA DNA, in a
CC   host cell.
SQ   Sequence   2454 BP;    562 A;    683 C;    707 G;    502 T;

  Query Match             8.8%;  Score 15;  DB 10;  Length 2454;
  Best Local Similarity 100.0%;  Pred. No. 2.28e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2219 cctcctcctgtcgtg 2233
          ||||||||||||||
Qy     42 CCTCCTCCTGTCGTG 56


RESULT   42
ID   Q29618 standard; DNA; 2454 BP.
AC   Q29618;
DT   10-MAR-1993  (first entry)
DE   PDI.
KW   Protein disulfide isomerase; proline-4-hydroxylase; ER;
KW   endoplasmic reticulum; HSA; human serum albumin;
KW   disulphide bond; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             49..1575
FT                   /*tag= a
```



```
FT                   /label= PDI
FT   signal_peptide  49..99
FT                   /*tag= b
FT   mat_peptide     100..1575
FT                   /*tag= c
PN   EP-509841-A.
PD   21-OCT-1992.
PF   16-APR-1992; 303492.
PR   18-APR-1991; JP-114074.
PR   30-OCT-1991; JP-311601.
PA   (TOFU ) TONEN CORP.
PI   Hayano T, Honma K, Katoh S, Suzuki M, Takahashi N;
DR   WPI; 92-351464/43.
DR   P-PSDB; R25297.
PT   Increased secretion of human serum albumin from yeast cells -
PT   using a co-expression system of protein di-sulphide isomerase
PT   gene and protein gene
PS   Example 3; Page 18-23; 51pp; English.
CC   PDI enhances formation of normal disulphide bonds in HSA, and so
CC   assists the formation of the higher order structure of HSA more
CC   efficiently.  Its co-expression with HSA reduces the chance of
CC   association of HSA mols. with unstable structure and their
CC   decomposition by host cell proteases.  The secretion level of HSA
CC   is thus increased, and the PDI-retention efficiency in the
CC   endoplasmic reticulum (ER) is improved in conjuction with a large
CC   scale expression system of a protein localising in the ER of a
CC   eukaryote.
SQ   Sequence   2454 BP;    562 A;    681 C;    709 G;    502 T;

  Query Match             8.8%;  Score 15;  DB 5;  Length 2454;
  Best Local Similarity 100.0%;  Pred. No. 2.28e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2219 cctcctcctgtcgtg 2233
          ||||||||||||||
Qy     42 CCTCCTCCTGTCGTG 56


RESULT   43
ID   Q32665 standard; cDNA; 2808 BP.
AC   Q32665;
DT   04-MAY-1993  (first entry)
DE   Nearly complete p107 cDNA.
KW   p107; probe; cDNA; amplify; primer; clone; tumour; suppressor; gene;
KW   cell cycle; regulation; RB; protein; retinoblastoma susceptibilty;
KW   viral; cellular; binding domain; pocket; T-p107; binding function;
KW   E1A-p107; homology; map; 20q 11.2; ss.
OS   Homo sapiens.
PN   WO9221781-A.
PD   10-DEC-1992.
PF   29-MAY-1992; U04504.
```

```
PR   31-MAY-1991; US-708962.
PA   (DAND ) DANA FARBER CANCER INST INC.
PI   Even ME, Livingston DM;
DR   WPI; 92-433679/52.
DR   P-PSDB; R30418.
PT   cDNA encoding P107 and its probe - is used for early diagnosis of
PT   tumours resulting from point mutation(s) or specific deletions
PS   Claim 1; Fig 1; 30pp; English.
CC   This represents the p107 cDNA sequence.  This cDNA was isolated using
CC   probes designed using fragments of the p107 amino acid sequence
CC   (see also R30415-17).  The protein p107 is a tumour suppressor gene
CC   product which is involved in cell cycle regulatory activity.  p107
CC   has been shown to be structurally and functionally similar to the RB
CC   protein of the retinoblastoma susceptibilty gene.  p107 binds viral
CC   and cellular proteins at a binding domain designated the "p107 pocket".
CC   The p107 pocket can perform a T-p107 binding function and a E1A-p107
CC   binding function.  RB and p107 proteins show a high degree of homology
CC   around their respective "pockets" but very little homology outside this
CC   region.  The p107 coding sequence has been found to map to 20q 11.2.
SQ   Sequence   2808 BP;    921 A;    546 C;    577 G;    764 T;

  Query Match            8.8%;  Score 15;  DB 5;  Length 2808;
  Best Local Similarity 100.0%;  Pred. No. 2.28e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    581 aagaggaaattggaa 595
          |||||||||||||||
Qy    135 AAGAGGAAATTGGAA 149


RESULT   44
ID   Q15171 standard; DNA; 2828 BP.
AC   Q15171;
DT   11-MAR-1992  (first entry)
DE   Bacillus thuringiensis ori43 replication origin.
KW   origin of replication; 43mDa plasmid; AT-rich region; ss.
OS   Bacillus thuringiensis var. kurstaki strain HD-263.
FH   Key             Location/Qualifiers
FT   cds             829..2362
FT                   /*tag= a
FT   repeat_unit     28..35
FT                   /*tag= b
FT                   /rpt_type= inverted
FT                   /note= "see /*tag= c for other member of pair"
FT   repeat_unit     54..61
FT                   /*tag= c
FT                   /rpt_type= inverted
FT   repeat_unit     435..444
FT                   /*tag= d
FT                   /rpt_type= direct
FT                   /note= "see /*tag= e for other member of pair"
```



```
FT   repeat_unit     571..580
FT                   /*tag= e
FT                   /rpt_type= direct
FT   repeat_unit     485..491
FT                   /*tag= f
FT                   /rpt_type= direct
FT                   /note= "see /*tag= g for other member of pair"
FT   repeat_unit     578..584
FT                   /*tag= g
FT                   /rpt_type= direct
FT   repeat_unit     758..773
FT                   /*tag= h
FT                   /rpt_type= inverted
FT                   /note= "see /*tag= i for other member of pair"
FT   repeat_unit     782..796
FT                   /*tag= i
FT                   /rpt_type= inverted
FT   repeat_unit     815..825
FT                   /*tag= j
FT                   /rpt_type= direct
FT                   /note= "see /*tag= k for other member of pair"
FT   repeat_unit     11174..1184
FT                   /*tag= k
FT                   /rpt_type= direct
FT   repeat_unit     1173..1197
FT                   /*tag= l
FT                   /rpt_type= inverted
FT                   /note= "see /*tag= m for other member of pair"
FT   repeat_unit     1280..1306
FT                   /*tag= m
FT                   /rpt_type= inverted
FT   repeat_unit     1452..1465
FT                   /*tag= n
FT                   /rpt_type= inverted
FT                   /note= "see /*tag= o for other member of pair"
FT   repeat_unit     1561..1572
FT                   /*tag= o
FT                   /rpt_type= inverted
FT   repeat_unit     1518..1532
FT                   /*tag= p
FT                   /rpt_type= inverted
FT                   /note= "see /*tag= q for other member of pair"
FT   repeat_unit     1537..1551
FT                   /*tag= q
FT                   /rpt_type= inverted
FT   repeat_unit     1990..1999
FT                   /*tag= r
FT                   /rpt_type= inverted
FT                   /note= "see /*tag= s for other member of pair"
FT   repeat_unit     2030..2038
FT                   /*tag= s
```



```
FT                   /rpt_type= inverted
FT   misc_feature    492..569
FT                   /*tag= t
FT                   /note= "AT-rich region"
PN   WO9118102-A.
PD   28-NOV-1991.
PF   14-MAY-1991; U03360.
PR   15-MAY-1990; US-523671.
PA   (ECOG-) ECOGEN INC.
PI   Baum JA;
DR   WPI; 91-369254/50.
DR   P-PSDB; R15354.
PT   Shuttle vector for recombinant Bacillus strain development -
PT   useful for inserting cloned insecticidal Bt toxin genes and
PT   contg. no DNA from non-Bt biological sources
PS   Claim 28; Fig 7; 94pp; English.
CC   The ori43 sequence was isolated from plasmid pEG599, a 43mDa
CC   plasmid containing a B.thuringiensis ori. The over all AT content of
CC   ori60 is 67.9 per cent although the sequence contains a segment of
CC   AT-rich DNA (i.e. greater than 85 per cent AT). One of the
CC   direct repeat sequences which precedes the ORF (i.e. /*tag= j)
CC   lies over the putative Shine-Dalgarno sequence.
CC   See also Q15169 and Q15170.
SQ   Sequence   2828 BP;    1032 A;    361 C;    546 G;    889 T;

  Query Match            8.8%;  Score 15;  DB 3;  Length 2828;
  Best Local Similarity 100.0%;  Pred. No. 2.28e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1700 tatttattttatctt 1714
          |||||||||||||||
Cp    162 TATTTATTTTATCTT 148


RESULT   45
ID   Q63261 standard; cDNA; 2924 BP.
AC   Q63261;
DT   07-DEC-1994  (first entry)
DE   CD26 cDNA clone.
KW   Human; T cell activation antigen; CD26; analogues; deletion; soluble;
KW   signal peptidase; immune-stimulating; response-stimulating; AIDS;
KW   immunosuppression; AIDS-related complex; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             11..2311
FT                   /*tag= a
FT                   /product= CD26
PN   WO9409132-A.
PD   28-APR-1994.
PF   19-AUG-1993; U07923.
PR   21-AUG-1992; US-934162.
```



```
PA   (DAND ) DANA FARBER CANCER INST INC.
PI   Morimoto C,  Schlossman S,  Tanaka T;
DR   WPI; 94-151317/18.
DR   P-PSDB; R54611.
PT   Polypeptide fragments and analogues of CD26 and encoding nucleic
PT   acid - useful for stimulating immune response, e.g. for treatment
PT   of AIDS to counteract immunosuppressive drug, and as vaccine
PT   adjuvant.
PS   Disclosure; Page 46-49; 85pp; English.
CC   This sequence encodes the human T cell activation antigen CD26.  The
CC   invention is concerned with polypeptide fragments and analogues of
CC   CD26 which have internal deletions (see also R54612-14).  The
CC   analogues pref. lack residues 3-9 or 24-34.  These analogues are
CC   soluble under physiological conditions and lack enough amino acid
CC   residues to render them susceptible to cleavage by signal peptidase.
CC   The peptide fragments and analogues are useful as immune or response-
CC   stimulating therapeutics, eg. they may be used for treatment of
CC   disease conditions characterised by immunosuppression, eg. AIDS or
CC   AIDS-related complex, other virally or environmentally-induced
CC   conditions, and certain congenital immune deficiencies.  The peptides
CC   can be employed to increase immune function which has been impaired by
CC   use of immunosupressive drugs, such as certain chemotherapeutic drugs.
SQ   Sequence   2924 BP;    912 A;    586 C;    623 G;    803 T;

  Query Match            8.8%;  Score 15;  DB 11;  Length 2924;
  Best Local Similarity 100.0%;  Pred. No. 2.28e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     92 ctgaacaaaggcaca 106
          |||||||||||||||
Cp     69 CTGAACAAAGGCACA 55


RESULT   46
ID   Q46089 standard; cDNA; 2924 BP.
AC   Q46089;
DT   05-FEB-1994  (first entry)
DE   Sequence of human CD26 cDNA.
KW   Human T cell activation antigen; monoclonal antibody Tal; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             11..2311
FT                   /*tag= a
PN   WO9316102-A.
PD   19-AUG-1993.
PF   09-APR-1992; U02892.
PR   06-FEB-1992; US-832211.
PA   (DAND ) DANA FARBER CANCER INST INC.
PI   Morimoto C,  Schlossman SF,  Tanaka T;
DR   WPI; 93-272827/34.
DR   P-PSDB; R40909.
```

```
PT   Polypeptide fragments of CD26 - are capable of disrupting binding
PT   of CD45 and CD26 and thus interfering with T-cell activation
PS   Disclosure; pages 39-43; 73pp; English.
CC   C26 is a human T cell antivation antigen originally identified by
CC   its reactivity with the MAb Ta1. C26 cDNA library was constructed
CC   from human PHA-activated T cells using the CDM7vector.The hydrophobic
CC   N-terminal of the predicted CD26 polypeptide has the characteristics
CC   of a signal sequence of the type II membrane protein, which is
CC   reinforced by the observation that potential N-glycosylation sites
CC   are located in the carboxy side of the hydrophobic core.Therefore
CC   the N-terminal 6 AAs are predicted to be cytoplasmic, the next 22
CC   AAs are predicted to transverse the cytoplasmic membrane, and the
CC   738 C-terminal AAs constitute the predicted extracellular domain.
SQ   Sequence   2924 BP;    912 A;    587 C;    622 G;    803 T;

  Query Match           8.8%;  Score 15;  DB 8;  Length 2924;
  Best Local Similarity 100.0%;  Pred. No. 2.28e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     92 ctgaacaaaggcaca 106
          |||||||||||||||
Cp     69 CTGAACAAAGGCACA 55


RESULT   47
ID   N90068 standard; DNA; 3077 BP.
AC   N90068;
DT   1-NOV-1989  (first entry)
DE   Ricin D gene
KW   Ricin D gene; Ricinus communis; caster beans; Zanibariensis variety;
KW   modified; lectin binding removed; reduced cell binding.
OS   Ricinus communis (caster beans)
PN   WO8904839-A.
PD   01-NOV-1989.
PF   23-NOV-1988; U04238
PR   24-NOV-1987; US-124735
PA   (GENE) Genetics Inst. Inc.
PI   Brown E, Jones S;
DR   89-178366/24.
DR   P-PSDB; P90079.
PT   Modified ricin molecules and toxin conjugates
PT   - in which the lectin binding function of the B chain
PT   is removed or diminished to reduce cell binding.
PS   Disclosure; Fig. 1; 66pp; English.
CC   DNA (see corresp. P90079) encoding ricin D gene found on the
CC   EcoRI-HindIII fragment of DNA from Ricinus communis,
CC   Zanibariensis variety. Patent discloses many modifications of ricin
CC   in which the lectin binding function of the B chain is diminished or
CC   removed, and conjugation to toxins to eliminate cell binding.
SQ   Sequence   3077 BP;    985 A;    514 C;    517 G;   1059 T;   2 Others;
```



```
  Query Match           8.8%;  Score 15;  DB 1;  Length 3077;
  Best Local Similarity 100.0%;  Pred. No. 2.28e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2973 ataaaataaataata 2987
          |||||||||||||||
Qy    151 ATAAAATAAATAATA 165


RESULT   48
ID   N90229 standard; cDNA; 3500 BP.
AC   N90229;
DT   1-NOV-1989  (first entry)
DE   Mouse tumour necrosis factor DNA.
KW   Tumour necrosis factor; mouse; DNA; exons; introns;
KW   carcinostatic mechanism; expression vector.
OS   Murine (Mouse)
FH   Key             Location/Qualifiers
FT   intron          1..814
FT                   /*tag= a
FT   exon            815..999
FT                   /*tag= b
FT   intron          1000..1519
FT                   /*tag= c
FT   exon            1520..1573
FT                   /*tag= d
FT   intron          1574..1751
FT                   /*tag= e
FT   exon            1752..1799
FT                   /*tag= f
FT   intron          1800..2092
FT                   /*tag= g
FT   exon            2093..2508
FT                   /*tag= h
FT   intron          2509..3501
FT                   /*tag= i
FT   repeat_region   354..383
FT                   /*tag= j
FT   tata_signal     618..623
FT                   /*tag= k
FT   polya_site      3261..3266
FT                   /*tag= l
PN   J01137976-A.
PD   24-MAY-1989.
PF   20-NOV-1987; 291841.
PR   20-NOV-1987; JP-291841.
PA   (ASAH) Asahi Chemical Ind KK.
DR   WPI; 89-198221/27.
DR   P-PSDB; P90424.
PT   DNA fragment contg. mouse TNF-coding chromosomal gene
PT   - contains exon and intron and about 800 bases in the
```



```
PT   5'-terminal adjacent to the exon.
PS   Claim; fig 2; 8pp; Japanese.
CC   DNA sequence of tumour necrosis factor of mouse. The peptide obtd. after
CC   intron splicing is shown in P90424. Used to study carcinostatic
CC   mechanism of TNF, and the promoter can be used in an expression vector.
SQ   Sequence   3500 BP;    893 A;    938 C;    901 G;    768 T;

  Query Match           8.8%;  Score 15;  DB 1;  Length 3500;
  Best Local Similarity 100.0%;  Pred. No. 2.28e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2943 cacttattatttatt 2957
          |||||||||||||||
Cp    169 CACTTATTATTTATT 155


RESULT   49
ID   Q99852 standard; DNA; 6854 BP.
AC   Q99852;
DT   23-DEC-1995  (first entry)
DE   Yeast TIH2 gene.
KW   TIH2; targets interacting with HHR25-like protein-kinase isoforms;
KW   casein-kinase I; CKI; cell growth; ss.
OS   Saccharomyces cerevisiae.
FH   Key             Location/Qualifiers
FT   cds             2050..4056
FT                   /*tag= a
PN   WO9519988-A1.
PD   27-JUL-1995.
PF   23-JAN-1995; U00912.
PR   21-JAN-1994; US-184605.
PA   (ICOS-) ICOS CORP.
PI   DeMaggio AJ, Hoekstra MF;
DR   WPI; 95-269414/35.
DR   P-PSDB; R81315.
PT   Isolation and detection of novel DNA encoding proteins which bind to
PT   casein kinase I isoforms - useful in understanding the control of
PT   normal and malignant cell growth.
PS   Claim 24; Page 36-40; 84pp; English.
CC   To identify genes encoding proteins that interact with S. cerevisiae
CC   HHR25 CK1 protein-kinase, a dihybrid screening method was used
CC   in which 'bait' components (CK1 isoforms) were fused to the DNA
CC   binding domain of a transcription factor (lexA protein) and 'prey'
CC   components (putative CK1 interactive proteins) were fused to the
CC   transactivation domain of the transcription factor (GAL4). Fusions
CC   were expressed in hosts containing a reporter gene fused to a LexA DNA
CC   binding site. Prey-bait binding led to transcriptional activation
CC   producing beta-galactosidase activity. Clones TIH1, TIH2 and TIH3
CC   were identified, and full-length genomic clones encoding TIH1-3
CC   (Q99851-53) were isolated from genomic libraries.
SQ   Sequence   6854 BP;    2329 A;    1358 C;    1160 G;    2007 T;
```



```
  Query Match           8.8%;  Score 15;  DB 15;  Length 6854;
  Best Local Similarity 100.0%;  Pred. No. 2.28e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

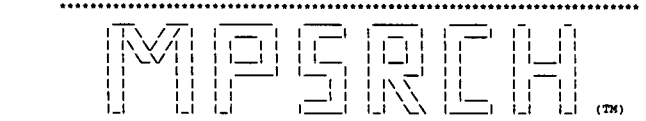
Db   5579 ggaagataaaataaa 5593
          |||||||||||||||
Qy    146 GGAAGATAAAATAAA 160


RESULT   50
ID   V15519 standard; DNA; 7647 BP.
AC   V15519;
DT   22-MAY-1998  (first entry)
DE   Hamster oral papilloma virus genomic DNA.
KW   Diagnosis; papilloma virus infectious disease;
KW   cancer; screening; antitumour agent; antisense treatment; ss.
OS   Hamster oral papilloma virus.
PN   J10042875-A.
PD   17-FEB-1998.
PF   06-AUG-1996; 207143.
PR   06-AUG-1996; JP-207143.
PA   (TORA ) TORAY IND INC.
DR   WPI; 98-186869/17.
PT   DNA encoding hamster oral papilloma virus protein - useful for, e.g.
PT   diagnosing cancer related papilloma virus infections
PS   Claim 11; Pages 14-16; 17pp; Japanese.
CC   The present sequence is a hamster oral papilloma virus genomic
CC   DNA. The DNA can be used for the diagnosis of papilloma virus
CC   infectious diseases related to cancer, screening for antitumour
CC   agents and antisense treatment.
SQ   Sequence   7647 BP;    2102 A;    1703 C;    1894 G;    1948 T;

  Query Match           8.8%;  Score 15;  DB 39;  Length 7647;
  Best Local Similarity 93.8%;  Pred. No. 2.28e+01;
  Matches   15;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db   2176 agtaatagcagccttc 2191
          ||| ||||||||||||
Cp    115 AGTNATAGCAGCCTTC 100


Search completed: Tue Dec 22 21:57:04 1998
Job time : 151 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Tue Dec 22 21:38:27 1998; MasPar time 284.76 Seconds
1331.141 Million cell updates/sec

Tabular output not generated.

Title: >US-08-530-112A-2965
Description: (1-172) from US08530112A.seq
Perfect Score: 171
N.A. Sequence: 1 GATCTGAACGATGCCGTGTC.....AAAATAAATAATAAGTGAAA 172
Comp: CTAGACTTGCTACGGCACAG.....TTTTATTTATTATTCACTTT

Scoring table: TABLE jmetric
Gap 60

Nmatch STD : Dbase 0; Query 0

Searched: 567134 seqs, 1101898692 bases x 2

Post-processing: Minimum Match 0%
Listing first 1000 summaries

Database: embl55
1:em_ba 2:em_fun 3:em_htg 4:em_hum1 5:em_hum2 6:em_in
7:em_om 8:em_or 9:em_ov 10:em_pat 11:em_ph 12:em_pl
13:em_ro 14:em_vi
Database: genbank107
15:gb_ba1 16:gb_ba2 17:gb_htg 18:gb_in 19:gb_om 20:gb_ov
21:gb_pat 22:gb_ph 23:gb_pl1 24:gb_pl2 25:gb_pr1
26:gb_pr2 27:gb_pr3 28:gb_ro 29:gb_st 30:gb_sts 31:gb_sy
32:gb_un 33:gb_vi

Statistics: Mean 7.930; Variance 2.979; scale 2.662

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
c	1	19	11.1	1316	18	AF012048	Drosophila yakuba retr	1.09e+00
c	2	19	11.1	1317	18	AF012049	Drosophila yakuba retr	1.09e+00
c	3	19	11.1	1317	18	AF012050	Drosophila yakuba retr	1.09e+00
	4	19	11.1	27694	23	HASMT	Hansenula wingei mitoc	1.09e+00
	5	19	11.1	188205	27	AC003986	Homo sapiens BAC clone	1.09e+00
c	6	18	10.5	1196	18	AF012051	Drosophila teissieri r	5.43e+00
c	7	18	10.5	1316	18	AF012047	Drosophila teissieri r	5.43e+00
	8	18	10.5	1396	15	VIBTCPJA	Vibrio cholerae toxin	5.43e+00
	9	18	10.5	2041	15	VIBACFB	Vibrio cholerae access	5.43e+00
	10	18	10.5	4785	23	YSCMTCG19	Saccharomyces cerevisi	5.43e+00
	11	18	10.5	5692	23	MISC19	Yeast mitochondrial ge	5.43e+00
	12	18	10.5	5692	23	YSCMTRG2P	yeast (s.cerevisiae) m	5.43e+00
	13	18	10.5	13065	15	AE001074	Archaeoglobus fulgidus	5.43e+00
c	14	18	10.5	26833	25	GORAFPA	Gorilla alpha-fetoprot	5.43e+00
c	15	18	10.5	26874	25	PTU21916	Pan troglodytes alpha-	5.43e+00
c	16	18	10.5	27553	26	HUMAFP	Homo sapiens alpha-fet	5.43e+00
c	17	18	10.5	71803	17	AC005504	*** SEQUENCING IN PROG	5.43e+00
c	18	18	10.5	114827	26	AC004413	Homo sapiens PAC clone	5.43e+00
c	19	18	10.5	133687	17	AC002093	*** SEQUENCING IN PROG	5.43e+00
	20	18	10.5	170000	17	AC004394	*** SEQUENCING IN PROG	5.43e+00
c	21	18	10.5	170000	17	AC004579	*** SEQUENCING IN PROG	5.43e+00
	22	18	10.5	171952	28	AC002324	Mus musculus chromosom	5.43e+00
	23	18	10.5	190000	17	AC004580	*** SEQUENCING IN PROG	5.43e+00
	24	18	10.5	207957	26	AC004470	Homo sapiens Xp22 BAC	5.43e+00
	25	17	9.9	38	25	HUMTCVD1ES	Human (clone: 4coj22)	2.56e+01
c	26	17	9.9	569	28	MMCKIT19	M.musculus DNA for c-k	2.56e+01
c	27	17	9.9	1549	33	HIVU43136	Human immunodeficiency	2.56e+01
	28	17	9.9	1739	26	HSNCPIIO1	Human neutrophil oxida	2.56e+01
	29	17	9.9	2467	33	AF002732	Cydia pomonella granul	2.56e+01
c	30	17	9.9	2574	23	SCU17018	Saccharomyces cerevisi	2.56e+01
	31	17	9.9	3131	15	ECOLEP	E.coli lep operon: pro	2.56e+01
	32	17	9.9	3331	15	LABORGRPL	Plasmid pLAB1000 (from	2.56e+01
c	33	17	9.9	3331	21	A14660	plasmid pLAB1000 or pH	2.56e+01
	34	17	9.9	4053	15	ECPUTA	E.coli (K12-CSH4) PUTA	2.56e+01
	35	17	9.9	4082	18	DMBOSS	D. melanogaster gene f	2.56e+01
	36	17	9.9	4102	15	STPUTA	S.typhimurium putA gen	2.56e+01
c	37	17	9.9	6295	21	I35766	Sequence 4 from patent	2.56e+01
c	38	17	9.9	6295	21	I14048	Sequence 4 from patent	2.56e+01
c	39	17	9.9	6312	21	I35768	Sequence 6 from patent	2.56e+01
c	40	17	9.9	6312	21	I35769	Sequence 7 from patent	2.56e+01
c	41	17	9.9	6312	23	ATHCTR1B	Arabidopsis thaliana n	2.56e+01
c	42	17	9.9	6312	21	I14047	Sequence 3 from patent	2.56e+01

c	43	17	9.9	6312	21	I14049	Sequence 5 from patent	2.56e+01
c	44	17	9.9	6312	21	I14051	Sequence 7 from patent	2.56e+01
c	45	17	9.9	6312	21	I14050	Sequence 6 from patent	2.56e+01
c	46	17	9.9	6312	21	I35765	Sequence 3 from patent	2.56e+01
c	47	17	9.9	6312	21	I35767	Sequence 5 from patent	2.56e+01
	48	17	9.9	7553	18	DROBRSEVM	Drosophila melanogaste	2.56e+01
	49	17	9.9	7603	16	AF036961	Mycoplasma hominis glu	2.56e+01
	50	17	9.9	7634	19	CHBLGLPS	C.hircus beta-lactoglo	2.56e+01
	51	17	9.9	8540	25	HUMLIMGP	Human lens membrane pr	2.56e+01
c	52	17	9.9	9800	15	ECAE000368	Escherichia coli K-12	2.56e+01
c	53	17	9.9	10004	15	ECAE000343	Escherichia coli K-12	2.56e+01
	54	17	9.9	10519	15	U67524	Methanococcus jannasch	2.56e+01
	55	17	9.9	18301	18	PFSC03014	Plasmodium falciparum	2.56e+01
c	56	17	9.9	29254	15	ECOK12RIII	Escherichia coli ribon	2.56e+01
	57	17	9.9	31184	23	YSCD9476	Saccharomyces cerevisi	2.56e+01
c	58	17	9.9	33847	17	AC000033	*** SEQUENCING IN PROG	2.56e+01
	59	17	9.9	41781	18	CELZK250	Caenorhabditis elegans	2.56e+01
c	60	17	9.9	42497	26	AC004798	Homo sapiens chromosom	2.56e+01
c	61	17	9.9	44089	26	AC002101	Genomic sequence from	2.56e+01
	62	17	9.9	55136	25	HUMVITDBP	Human vitamin D-bindin	2.56e+01
	63	17	9.9	59748	23	HVCH4H	Hordeum vulgare DNA fo	2.56e+01
c	64	17	9.9	71128	15	ECU29581	Escherichia coli K-12	2.56e+01
	65	17	9.9	75475	23	AB005235	Arabidopsis thaliana g	2.56e+01
	66	17	9.9	81004	23	ATAC002521	Arabidopsis thaliana c	2.56e+01
c	67	17	9.9	86759	26	HSAF003626	Homo sapiens cosmids I	2.56e+01
c	68	17	9.9	90832	27	AC003065	Homo sapiens Chromosom	2.56e+01
	69	17	9.9	93250	26	HS447B16	Human DNA sequence fro	2.56e+01
	70	17	9.9	97979	26	AC004739	Homo sapiens PAC clone	2.56e+01
c	71	17	9.9	100000	26	AP000040	Homo sapiens genomic D	2.56e+01
c	72	17	9.9	100368	26	HS514K20	Human DNA sequence fro	2.56e+01
	73	17	9.9	101877	27	HS37M3	Homo sapiens DNA seque	2.56e+01
c	74	17	9.9	102717	26	AC004079	Homo sapiens PAC clone	2.56e+01
	75	17	9.9	102857	6	AC004781	Drosophila melanogaste	2.56e+01
c	76	17	9.9	104278	18	AC005269	Drosophila melanogaste	2.56e+01
c	77	17	9.9	110000	17	CEY57G11	Caenorhabditis elegans	2.56e+01
	78	17	9.9	114736	18	PFMAL3P3	Plasmodium falciparum	2.56e+01
	79	17	9.9	127039	18	AC005732	Drosophila melanogaste	2.56e+01
	80	17	9.9	138313	17	AC004848	*** SEQUENCING IN PROG	2.56e+01
	81	17	9.9	142498	17	AC004967	*** SEQUENCING IN PROG	2.56e+01
	82	17	9.9	145247	17	AC004886	*** SEQUENCING IN PROG	2.56e+01
	83	17	9.9	157970	27	AC005609	Homo sapiens chromosom	2.56e+01
c	84	17	9.9	162472	17	AC005730	*** SEQUENCING IN PROG	2.56e+01
	85	17	9.9	163023	17	AC004815	*** SEQUENCING IN PROG	2.56e+01
	86	17	9.9	163974	27	AB011399	Homo sapiens gene for	2.56e+01
	87	17	9.9	164511	27	AC004894	Homo sapiens PAC clone	2.56e+01
c	88	17	9.9	164675	26	AC004384	Human Chromosome X clo	2.56e+01
c	89	17	9.9	165507	17	AC004903	*** SEQUENCING IN PROG	2.56e+01
c	90	17	9.9	167468	17	AC005033	*** SEQUENCING IN PROG	2.56e+01
c	91	17	9.9	173275	17	AC005377	*** SEQUENCING IN PROG	2.56e+01
	92	17	9.9	177311	26	AC004385	Human Chromosome X clo	2.56e+01
c	93	17	9.9	177756	17	AC005020	*** SEQUENCING IN PROG	2.56e+01

c	94	17	9.9	180175	17	AC005105	*** SEQUENCING IN PROG	2.56e+01
c	95	17	9.9	187051	17	AC004688	*** SEQUENCING IN PROG	2.56e+01
c	96	17	9.9	206056	17	AC004387	*** SEQUENCING IN PROG	2.56e+01
c	97	17	9.9	216916	17	AC004970	*** SEQUENCING IN PROG	2.56e+01
	98	17	9.9	313573	18	CEF57G11C	Caenorhabditis elegans	2.56e+01
	99	16	9.4	129	21	AO8897	H.sapiens (haplotype 2	1.14e+02
	100	16	9.4	264	19	GOTHBBE1	Goat germline-like bet	1.14e+02
	101	16	9.4	265	30	G37829	mefl Plasmodium falcip	1.14e+02
c	102	16	9.4	278	30	G02557	human STS WI-3495.	1.14e+02
	103	16	9.4	315	30	HS179XH10	H. sapiens (D3S1282) D	1.14e+02
c	104	16	9.4	332	18	PHU77867	Parantonae hispida 12S	1.14e+02
	105	16	9.4	350	30	DM63D12T	D. melanogaster STS de	1.14e+02
	106	16	9.4	400	30	G17872	human STS SHGC-14929 c	1.14e+02
c	107	16	9.4	400	30	G13700	human STS SHGC-12107 c	1.14e+02
c	108	16	9.4	481	18	APU36740	Acyrthosiphon pisum 16	1.14e+02
	109	16	9.4	566	27	HUMZC30H06	Homo sapiens full leng	1.14e+02
	110	16	9.4	642	23	ARYITS	Argyroxiphium caligini	1.14e+02
	111	16	9.4	642	24	AF061886	Argyroxiphium grayanum	1.14e+02
	112	16	9.4	642	24	AF061885	Argyroxiphium grayanum	1.14e+02
	113	16	9.4	642	24	AF061884	Argyroxiphium sandwice	1.14e+02
	114	16	9.4	642	24	AF061883	Argyroxiphium sandwice	1.14e+02
	115	16	9.4	642	24	AF061887	Argyroxiphium kauense	1.14e+02
	116	16	9.4	704	23	IHU34212	Iris hartwegii H-1 ATP	1.14e+02
	117	16	9.4	823	23	CAWODUT1	C.albicans (WO-1) DUT1	1.14e+02
c	118	16	9.4	835	28	RNEDN2805	Rattus norvegicus endo	1.14e+02
c	119	16	9.4	996	18	AF057985	Dendroctonus simplex c	1.14e+02
c	120	16	9.4	1022	25	HSPRKCSH07	Human protein kinase C	1.14e+02
	121	16	9.4	1106	15	MCAAK	M.capricolum DNA for C	1.14e+02
c	122	16	9.4	1156	18	AF044862	Phoebis agarithe large	1.14e+02
c	123	16	9.4	1158	18	AF044857	Colias scudderi large	1.14e+02
c	124	16	9.4	1159	18	AF044854	Colias eurytheme large	1.14e+02
c	125	16	9.4	1159	18	AF044850	Colias behri large sub	1.14e+02
c	126	16	9.4	1159	18	AF044853	Colias philodice eriph	1.14e+02
c	127	16	9.4	1160	18	AF044849	Colias alexandra alexa	1.14e+02
c	128	16	9.4	1161	18	AF044852	Colias alexandra edwar	1.14e+02
c	129	16	9.4	1162	18	AF044855	Colias meadii large su	1.14e+02
c	130	16	9.4	1162	18	AF044851	Colias croceus large s	1.14e+02
c	131	16	9.4	1165	18	AF044856	Colias erate poliograp	1.14e+02
c	132	16	9.4	1234	28	MMVICR	Mouse mRNA for vasoact	1.14e+02
	133	16	9.4	1356	20	CHKGALP	Gallus gallus G protei	1.14e+02
	134	16	9.4	1379	18	AF033934	Bactrocera xanthodes s	1.14e+02
	135	16	9.4	1508	18	S55234	CAR2=cAMP receptor sub	1.14e+02
	136	16	9.4	1545	18	PFACSPG	Plasmodium malariae (c	1.14e+02
	137	16	9.4	1546	18	DDU82512	Dictyostelium discoide	1.14e+02
	138	16	9.4	1585	26	AF038193	Homo sapiens clone 236	1.14e+02
	139	16	9.4	1612	23	SCYDL173W	S.cerevisiae chromosom	1.14e+02
	140	16	9.4	1712	25	HUMBCKDHAX	Human branched chain k	1.14e+02
c	141	16	9.4	1748	23	YSCMTCG17	Saccharomyces cerevisi	1.14e+02
	142	16	9.4	1754	18	PFEF1	P.falciparum MEF-1 gen	1.14e+02
c	143	16	9.4	1836	28	MMU52197	Mus musculus poly(A) p	1.14e+02
	144	16	9.4	1881	19	GOTHBBEVP	Goat embryonic beta-gl	1.14e+02

	145	16	9.4	2000	23	CPMTATP6	C.parapsilosos mitocho 1.14e+02
c	146	16	9.4	2011	18	BTMARLEA	B.terrestris for marin 1.14e+02
	147	16	9.4	2031	23	OSU89895	Oryza sativa pathogene 1.14e+02
c	148	16	9.4	2210	20	XELBRAT	Xenopus laevis brachyu 1.14e+02
	149	16	9.4	2221	19	GOTHBBEI	Goat embryonic beta-gl 1.14e+02
	150	16	9.4	2221	19	CHEBGLI	Goat epsilon I beta-gl 1.14e+02
c	151	16	9.4	2236	23	DPMTTRND	C.parapsilosis mitocho 1.14e+02
	152	16	9.4	2253	18	PFGR2	Plasmodium falciparum 1.14e+02
	153	16	9.4	2344	23	RCRICIN	Ricinus communis gene 1.14e+02
c	154	16	9.4	2561	20	AF077225	Danio rerio spadetail 1.14e+02
	155	16	9.4	2562	26	HSAMPD3S04	Human AMP deaminase (A 1.14e+02
	156	16	9.4	2579	18	TPPUBQ	T.pyriformis polyubiqu 1.14e+02
	157	16	9.4	2585	15	ECPEC	E.chrysanthemi DNA seq 1.14e+02
	158	16	9.4	2587	26	AF019114	Papio hamadryas X-ray 1.14e+02
	159	16	9.4	2638	23	SC130LP	S.commersonii (pOSML13 1.14e+02
	160	16	9.4	2707	23	AF030887	Ustilago maydis clone 1.14e+02
	161	16	9.4	2715	18	PFSC03116	Plasmodium falciparum 1.14e+02
	162	16	9.4	2773	16	S67974	RepS=replication prote 1.14e+02
	163	16	9.4	2797	25	HUMXRCC1	Human DNA-repair prote 1.14e+02
	164	16	9.4	2798	16	AF019139	Streptococcus thermoph 1.14e+02
c	165	16	9.4	3079	24	AF080600	Emericella nidulans ho 1.14e+02
	166	16	9.4	3116	26	AC002216	Homo sapiens (subclone 1.14e+02
	167	16	9.4	3168	18	CTBR6G	C.tentans BR6 gene. 1.14e+02
	168	16	9.4	3237	20	AF002983	Xenopus laevis cadheri 1.14e+02
c	169	16	9.4	3450	28	MUSIL8RB	Mouse interleukin-8 re 1.14e+02
c	170	16	9.4	3597	25	HSIL4R	Human IL-4-R mRNA for 1.14e+02
	171	16	9.4	3868	28	RNACPP1	R.norvegicus gene for 1.14e+02
	172	16	9.4	4122	18	DDU50042	Dictyostelium discoide 1.14e+02
	173	16	9.4	4432	18	TPPOLUBQ	T.pyriformis (CGL) pol 1.14e+02
	174	16	9.4	4544	18	PFAARPS2	P.falciparum mRNA for 1.14e+02
	175	16	9.4	4699	21	I59777	Sequence 6 from patent 1.14e+02
	176	16	9.4	4699	21	I07376	Sequence 28 from Paten 1.14e+02
	177	16	9.4	4699	23	TOBGLB13B	Nicotiana tabacum gluc 1.14e+02
	178	16	9.4	4699	21	I56911	Sequence 6 from patent 1.14e+02
	179	16	9.4	4699	21	I38436	Sequence 6 from patent 1.14e+02
	180	16	9.4	4699	21	I75104	Sequence 6 from patent 1.14e+02
	181	16	9.4	4736	20	CHKFASA	Chicken fatty acid syn 1.14e+02
c	182	16	9.4	4946	15	WSPCRABC	W.succinogenes genes p 1.14e+02
c	183	16	9.4	5585	28	MMGPXG	M.musculus GPX3 gene. 1.14e+02
	184	16	9.4	5762	18	PFAHSP86A	Plasmodium falciparum 1.14e+02
	185	16	9.4	5850	18	OGFTBB	O.gibsoni Tbb gene. 1.14e+02
	186	16	9.4	6119	23	PSU30841	Pisum sativum manganes 1.14e+02
	187	16	9.4	6499	15	AF027167	Streptococcus thermoph 1.14e+02
	188	16	9.4	6715	15	BSGALE	B.subtilis orfs 1,2,3, 1.14e+02
c	189	16	9.4	7002	25	HSTRE5REP	H.sapiens mRNA for TRE 1.14e+02
c	190	16	9.4	7056	23	SCGLUTSYN	S.cerevisiae DNA for g 1.14e+02
	191	16	9.4	7231	23	SCYDL171C	S.cerevisiae chromosom 1.14e+02
	192	16	9.4	7625	20	CHKFASB	Gallus gallus White Le 1.14e+02
c	193	16	9.4	7753	21	E03790	DNA sequence of exon X 1.14e+02
	194	16	9.4	8748	26	D67013	Homo sapiens DNA for a 1.14e+02
c	195	16	9.4	9313	23	CHY15839	Cochliobolus heterostr 1.14e+02

	196	16	9.4	9345	20	CHKFAS	Chicken fatty acid syn 1.14e+02
c	197	16	9.4	9496	18	CELF56D3	Caenorhabditis elegans 1.14e+02
	198	16	9.4	9985	18	PCU43145	Plasmodium chabaudi re 1.14e+02
c	199	16	9.4	9987	18	PFSC03090	Plasmodium falciparum 1.14e+02
	200	16	9.4	10412	15	U67541	Methanococcus jannasch 1.14e+02
	201	16	9.4	10888	15	U32720	Haemophilus influenzae 1.14e+02
	202	16	9.4	12950	15	AE000066	Rhizobium sp. NGR234 p 1.14e+02
	203	16	9.4	13232	18	DDID16466	Cellular slime mold Mi 1.14e+02
	204	16	9.4	13471	18	PFSC03023	Plasmodium falciparum 1.14e+02
	205	16	9.4	14000	23	SPAC23E2	S.pombe chromosome I c 1.14e+02
	206	16	9.4	14653	18	CEF28C12	Caenorhabditis elegans 1.14e+02
	207	16	9.4	16500	24	SPAC13C5	S.pombe chromosome I c 1.14e+02
	208	16	9.4	17286	26	HSRINGG	H.sapiens gene encodin 1.14e+02
c	209	16	9.4	18499	18	CEY7A9A	Caenorhabditis elegans 1.14e+02
c	210	16	9.4	18980	26	AC002049	Homo sapiens Chromosom 1.14e+02
c	211	16	9.4	20518	18	PFSC03030	Plasmodium falciparum 1.14e+02
c	212	16	9.4	21173	18	CEF21D5	Caenorhabditis elegans 1.14e+02
	213	16	9.4	22950	18	CEF55H2	Caenorhabditis elegans 1.14e+02
	214	16	9.4	23360	18	DHMMHC	D.hydei mMHC gene. 1.14e+02
	215	16	9.4	24494	26	AC004632	Homo sapiens chromosom 1.14e+02
c	216	16	9.4	24652	18	CEW04G5	Caenorhabditis elegans 1.14e+02
c	217	16	9.4	25329	23	SPAC3H8	S.pombe chromosome I c 1.14e+02
c	218	16	9.4	26527	18	CELT23F2	Caenorhabditis elegans 1.14e+02
c	219	16	9.4	27764	18	CET21E8	Caenorhabditis elegans 1.14e+02
	220	16	9.4	30088	18	PFSC03056	Plasmodium falciparum 1.14e+02
c	221	16	9.4	30335	17	AC005463	*** SEQUENCING IN PROG 1.14e+02
c	222	16	9.4	30782	18	CEF33A8	Caenorhabditis elegans 1.14e+02
	223	16	9.4	31520	26	HSXLSP1	Human Xq13 5' end of P 1.14e+02
c	224	16	9.4	32559	18	CELM02A10	Caenorhabditis elegans 1.14e+02
c	225	16	9.4	33539	18	CELF54E7	Caenorhabditis elegans 1.14e+02
c	226	16	9.4	35955	18	CEF53B7	Caenorhabditis elegans 1.14e+02
c	227	16	9.4	37328	26	HS67D6	Human DNA sequence fro 1.14e+02
	228	16	9.4	37785	25	HUMXRCC1G	Human XRCC1 DNA repair 1.14e+02
c	229	16	9.4	37801	18	CEAH6	Caenorhabditis elegans 1.14e+02
c	230	16	9.4	39569	26	AC000084	Homo sapiens Chromosom 1.14e+02
	231	16	9.4	40351	27	AC005620	Homo sapiens chromosom 1.14e+02
	232	16	9.4	40909	18	CEB0464	Caenorhabditis elegans 1.14e+02
	233	16	9.4	40998	27	AC005503	Homo sapiens clone UWG 1.14e+02
	234	16	9.4	41175	23	SC41KCIV	S.cerevisiae DNA (cosm 1.14e+02
	235	16	9.4	41730	26	HSAC000372	Human cosmid g1980a186 1.14e+02
	236	16	9.4	41778	25	AC002491	Homo sapiens Chromosom 1.14e+02
	237	16	9.4	42635	18	CELC29H12	Caenorhabditis elegans 1.14e+02
	238	16	9.4	42801	18	CELC42D8	Caenorhabditis elegans 1.14e+02
	239	16	9.4	43644	26	AC004092	Human Cosmid g1248a217 1.14e+02
c	240	16	9.4	43843	26	AC002054	Homo sapiens Chromosom 1.14e+02
	241	16	9.4	44312	27	AC005689	Homo sapiens clone UWG 1.14e+02
c	242	16	9.4	46872	17	AB018116	*** SEQUENCING IN PROG 1.14e+02
	243	16	9.4	47558	18	AF030694	Plasmodium falciparum 1.14e+02
	244	16	9.4	47777	26	AC004189	Homo sapiens clone UWG 1.14e+02
c	245	16	9.4	50849	23	SCE9669	Saccharomyces cerevisi 1.14e+02
c	246	16	9.4	52284	26	CH19R25187	Homo sapiens DNA from 1.14e+02

	247	16	9.4	58645	27	AC002068	Homo sapiens BAC clone 1.14e+02
c	248	16	9.4	61852	18	AC004445	Drosophila melanogaste 1.14e+02
	249	16	9.4	64923	26	AC002545	Homo sapiens chromosom 1.14e+02
c	250	16	9.4	65143	15	D83026	Bacillus subtilis geno 1.14e+02
	251	16	9.4	66084	24	AB015468	Arabidopsis thaliana g 1.14e+02
c	252	16	9.4	68415	24	ATAC005310	Arabidopsis thaliana c 1.14e+02
c	253	16	9.4	68889	23	AB010694	Arabidopsis thaliana g 1.14e+02
	254	16	9.4	69013	25	HS138A5	Human DNA sequence fro 1.14e+02
c	255	16	9.4	69034	18	AF007261	Reclinomonas americana 1.14e+02
c	256	16	9.4	72941	26	AC004227	Homo sapiens chromosom 1.14e+02
	257	16	9.4	73350	18	AC002493	Drosophila melanogaste 1.14e+02
c	258	16	9.4	73840	23	AB007644	Arabidopsis thaliana g 1.14e+02
	259	16	9.4	74613	23	AB005230	Arabidopsis thaliana g 1.14e+02
c	260	16	9.4	76819	26	AC004043	Homo sapiens chromosom 1.14e+02
	261	16	9.4	77183	26	HS299C21	Human DNA sequence fro 1.14e+02
	262	16	9.4	77713	26	HUAC004131	Homo sapiens Chromosom 1.14e+02
	263	16	9.4	80288	27	AC004901	Homo sapiens PAC clone 1.14e+02
c	264	16	9.4	81231	26	AC003044	Human PAC clone DJ1055 1.14e+02
c	265	16	9.4	83698	23	AB010072	Arabidopsis thaliana g 1.14e+02
c	266	16	9.4	87077	26	HUAC004097	Human Chromosome 16 BA 1.14e+02
	267	16	9.4	87585	18	AC004299	Drosophila melanogaste 1.14e+02
c	268	16	9.4	92636	26	AC005251	Homo sapiens BAC clone 1.14e+02
c	269	16	9.4	92975	26	AC002369	Homo sapiens Xp22 PAC 1.14e+02
	270	16	9.4	93695	23	ATF21O9	Arabidopsis thaliana D 1.14e+02
c	271	16	9.4	95968	18	CEY47D3B	Caenorhabditis elegans 1.14e+02
c	272	16	9.4	96217	26	AC002410	Human BAC clone RG264L 1.14e+02
	273	16	9.4	96589	23	ATF10N7	Arabidopsis thaliana D 1.14e+02
c	274	16	9.4	97006	24	ATF16G20	Arabidopsis thaliana D 1.14e+02
c	275	16	9.4	98343	26	AC005352	Homo sapiens chromosom 1.14e+02
	276	16	9.4	99781	26	AF031078	Homo sapiens chromosom 1.14e+02
c	277	16	9.4	101733	17	AC005518	*** SEQUENCING IN PROG 1.14e+02
c	278	16	9.4	102703	17	HS968D22	Human DNA sequence *** 1.14e+02
c	279	16	9.4	103268	26	HS67M12	Human DNA sequence fro 1.14e+02
	280	16	9.4	104710	26	AC002488	Homo sapiens chromosom 1.14e+02
	281	16	9.4	105984	17	AC005087	*** SEQUENCING IN PROG 1.14e+02
	282	16	9.4	106107	17	AC005663	*** SEQUENCING IN PROG 1.14e+02
	283	16	9.4	106921	26	AC004263	Homo sapiens PAC clone 1.14e+02
	284	16	9.4	110000	17	HSY738P9	Human DNA sequence *** 1.14e+02
c	285	16	9.4	110000	17	HSY313F4	Human DNA sequence *** 1.14e+02
c	286	16	9.4	110440	25	HS293E14	Human DNA sequence fro 1.14e+02
c	287	16	9.4	110441	17	HS23	*** SEQUENCING IN PROG 1.14e+02
	288	16	9.4	111398	26	HS381E2	Human DNA sequence fro 1.14e+02
c	289	16	9.4	111398	26	HS381E2	Human DNA sequence fro 1.14e+02
c	290	16	9.4	111507	26	HS236J17	Human DNA sequence fro 1.14e+02
c	291	16	9.4	112661	26	AC004668	Homo sapiens BAC clone 1.14e+02
	292	16	9.4	112756	26	AF030876	Homo sapiens chromosom 1.14e+02
c	293	16	9.4	114231	27	HS390C10	Homo sapiens DNA seque 1.14e+02
	294	16	9.4	116884	17	PFMALIP1	Plasmodium falciparum 1.14e+02
c	295	16	9.4	117939	27	HS211A9	Human DNA sequence fro 1.14e+02
	296	16	9.4	117968	27	HS272E8	Human DNA sequence *** 1.14e+02
c	297	16	9.4	119903	27	AC004662	Homo sapiens chromosom 1.14e+02

	298	16	9.4	120387	26	AC003051	Homo sapiens chromosom 1.14e+02
c	299	16	9.4	122302	26	AC003982	Human PAC clone 166H1 1.14e+02
	300	16	9.4	122325	27	HS1044O17	Human DNA sequence *** 1.14e+02
c	301	16	9.4	123234	26	AC002556	Human Chromosome 11 pa 1.14e+02
	302	16	9.4	123416	17	AC004847	*** SEQUENCING IN PROG 1.14e+02
c	303	16	9.4	124519	28	AC003019	Mus musculus Bac B6, c 1.14e+02
c	304	16	9.4	128117	26	AC004000	Human PAC clone DJ404F 1.14e+02
c	305	16	9.4	129098	26	HS46H23	Human DNA sequence fro 1.14e+02
	306	16	9.4	129969	26	HS204E5	Human DNA sequence fro 1.14e+02
	307	16	9.4	130020	26	HUAC004525	Homo sapiens Chromosom 1.14e+02
c	308	16	9.4	131673	17	HS336H9	Human DNA sequence *** 1.14e+02
c	309	16	9.4	133925	26	AC004562	Homo sapiens chromosom 1.14e+02
c	310	16	9.4	133965	26	AF064861	Homo sapiens PAC 128M1 1.14e+02
c	311	16	9.4	134929	27	AC004655	Homo sapiens Xp22-140- 1.14e+02
	312	16	9.4	135007	26	HS423N22	Homo sapiens DNA seque 1.14e+02
	313	16	9.4	135038	25	HUMYWXD703	Homo sapiens ADP/ATP c 1.14e+02
	314	16	9.4	135684	26	AC002463	Human BAC clone RG302F 1.14e+02
	315	16	9.4	135891	17	HS221G9	Human DNA sequence *** 1.14e+02
c	316	16	9.4	135891	17	HS221G9	Human DNA sequence *** 1.14e+02
	317	16	9.4	136142	17	HS37J18	Human DNA sequence *** 1.14e+02
	318	16	9.4	136580	17	AC005664	*** SEQUENCING IN PROG 1.14e+02
c	319	16	9.4	139259	18	CEY51A2D	Caenorhabditis elegans 1.14e+02
	320	16	9.4	142705	17	AC004972	*** SEQUENCING IN PROG 1.14e+02
c	321	16	9.4	142949	17	AC004960	*** SEQUENCING IN PROG 1.14e+02
c	322	16	9.4	147104	27	AC002472	Homo sapiens Chromosom 1.14e+02
	323	16	9.4	147831	17	AC005399	*** SEQUENCING IN PROG 1.14e+02
	324	16	9.4	149322	17	HS1119D9	Human DNA sequence *** 1.14e+02
c	325	16	9.4	150158	17	HS87B08	Human DNA sequence *** 1.14e+02
c	326	16	9.4	150894	15	D90904	Synechocystis sp. PCC6 1.14e+02
	327	16	9.4	153356	17	AC004995	*** SEQUENCING IN PROG 1.14e+02
c	328	16	9.4	153568	26	AC002981	Homo sapiens Xp22 BAC 1.14e+02
	329	16	9.4	155022	27	AC002395	Homo sapiens 12q24 PAC 1.14e+02
	330	16	9.4	155529	17	HS738P11	Human DNA sequence *** 1.14e+02
c	331	16	9.4	158172	17	AC004884	*** SEQUENCING IN PROG 1.14e+02
	332	16	9.4	161997	17	AC005026	*** SEQUENCING IN PROG 1.14e+02
	333	16	9.4	162327	17	AC004896	*** SEQUENCING IN PROG 1.14e+02
	334	16	9.4	162575	17	AC004086	*** SEQUENCING IN PROG 1.14e+02
c	335	16	9.4	163196	26	AC005220	Homo sapiens chromosom 1.14e+02
c	336	16	9.4	166832	26	AF070718	Homo sapiens BAC clone 1.14e+02
	337	16	9.4	170000	17	AC005599	*** SEQUENCING IN PROG 1.14e+02
	338	16	9.4	171905	26	HSAC002080	Human BAC clone RG253B 1.14e+02
c	339	16	9.4	172995	17	AC004926	*** SEQUENCING IN PROG 1.14e+02
	340	16	9.4	175825	26	HS106H8	Human DNA sequence fro 1.14e+02
	341	16	9.4	177014	26	AC004063	Homo sapiens chromosom 1.14e+02
c	342	16	9.4	177137	17	AC005089	*** SEQUENCING IN PROG 1.14e+02
c	343	16	9.4	177328	17	HS462D8	Human DNA sequence *** 1.14e+02
	344	16	9.4	183180	17	AC004929	*** SEQUENCING IN PROG 1.14e+02
c	345	16	9.4	185371	17	HS625H18	Human DNA sequence *** 1.14e+02
c	346	16	9.4	186609	23	MPOMTCG	Marchantia polymorpha 1.14e+02
c	347	16	9.4	187253	17	AC005486	*** SEQUENCING IN PROG 1.14e+02
	348	16	9.4	190482	26	AC004106	Homo sapiens Xp22 BAC 1.14e+02

c	349	16	9.4	191380	17	AC000052	*** SEQUENCING IN PROG 1.14e+02
c	350	16	9.4	191436	17	AC005722	*** SEQUENCING IN PROG 1.14e+02
c	351	16	9.4	193374	17	AC005139	*** SEQUENCING IN PROG 1.14e+02
c	352	16	9.4	193869	17	AC005043	*** SEQUENCING IN PROG 1.14e+02
	353	16	9.4	194985	28	AC002406	Mouse chromosome X BAC 1.14e+02
	354	16	9.4	196672	18	AC004767	Drosophila melanogaste 1.14e+02
	355	16	9.4	197391	17	AC005140	*** SEQUENCING IN PROG 1.14e+02
c	356	16	9.4	197391	17	AC005140	*** SEQUENCING IN PROG 1.14e+02
	357	16	9.4	200134	17	AC004806	*** SEQUENCING IN PROG 1.14e+02
	358	16	9.4	203583	17	AC004940	*** SEQUENCING IN PROG 1.14e+02
c	359	16	9.4	204596	17	AC005069	*** SEQUENCING IN PROG 1.14e+02
	360	16	9.4	204877	17	AC005375	*** SEQUENCING IN PROG 1.14e+02
	361	16	9.4	206547	17	AC005183	*** SEQUENCING IN PROG 1.14e+02
c	362	16	9.4	206784	26	HS93L7	Homo sapiens DNA seque 1.14e+02
	363	16	9.4	209499	17	AC005539	*** SEQUENCING IN PROG 1.14e+02
	364	16	9.4	212150	15	BSUB0020	Bacillus subtilis comp 1.14e+02
c	365	16	9.4	215609	26	AC002485	Human PAC clone RG212D 1.14e+02
c	366	16	9.4	220046	17	AC005343	*** SEQUENCING IN PROG 1.14e+02
	367	16	9.4	220633	26	HUU91321	Human Chromosome 16 BA 1.14e+02
	368	16	9.4	222543	17	HS477J10	Human DNA sequence *** 1.14e+02
c	369	16	9.4	224129	17	CEY7A9	Caenorhabditis elegans 1.14e+02
	370	16	9.4	224645	17	AC004153	*** SEQUENCING IN PROG 1.14e+02
c	371	16	9.4	243222	17	AC004019	*** SEQUENCING IN PROG 1.14e+02
c	372	16	9.4	244133	17	HS32B1	Human DNA sequence *** 1.14e+02
c	373	16	9.4	244649	17	AC005308	*** SEQUENCING IN PROG 1.14e+02
c	374	16	9.4	275180	17	CEY39G8	Caenorhabditis elegans 1.14e+02
c	375	16	9.4	310126	17	CEY48B6	Caenorhabditis elegans 1.14e+02
c	376	16	9.4	317311	17	AC004085	*** SEQUENCING IN PROG 1.14e+02
	377	16	9.4	326663	26	AF011889	Human Xq28 cosmids U12 1.14e+02
c	378	16	9.4	337546	17	CEY47D3	Caenorhabditis elegans 1.14e+02
	379	15	8.8	147	20	MMMICRO05	M.merlangus DNA (CA)n 4.68e+02
	380	15	8.8	161	28	MUSCEAPSP	Mus musculus transcrip 4.68e+02
	381	15	8.8	193	20	NAJNEURA	Naja naja (clone NTX2) 4.68e+02
	382	15	8.8	193	20	NAJNEURB	Naja naja (clone NTX1) 4.68e+02
	383	15	8.8	198	28	RNLINM1	Rat DNA with long inte 4.68e+02
	384	15	8.8	219	20	ONHHON51	Kokanee salmon DNA for 4.68e+02
	385	15	8.8	264	20	AF023271	Naja sputatrix neuroto 4.68e+02
	386	15	8.8	264	20	AF023272	Naja sputatrix neuroto 4.68e+02
c	387	15	8.8	303	30	G02594	human STS WI-5156. 4.68e+02
	388	15	8.8	315	28	RATHOMBXR4	Rat homeobox protein R 4.68e+02
c	389	15	8.8	316	25	HS27C12R	H.sapiens CpG island D 4.68e+02
	390	15	8.8	318	30	HS144ZH4	H. sapiens (D5S470) DN 4.68e+02
	391	15	8.8	321	26	AF058060	Homo sapiens clone ASP 4.68e+02
c	392	15	8.8	324	30	G06390	human STS WI-7093. 4.68e+02
c	393	15	8.8	353	30	G25364	human STS EST239117. 4.68e+02
	394	15	8.8	364	30	G07514	human STS SHGC-2837 cl 4.68e+02
	395	15	8.8	366	23	PNPCSP	Poncirus trifoliata (c 4.68e+02
c	396	15	8.8	371	30	G13062	sWSS2988 Eric D. Green 4.68e+02
c	397	15	8.8	376	30	G12379	Nile tilapia STS UNH22 4.68e+02
c	398	15	8.8	382	30	G28181	human STS SHGC-34176. 4.68e+02
c	399	15	8.8	385	30	G30227	human STS SHGC-36543, 4.68e+02

	400	15	8.8	390	30	G33510	human STS SHGC-50389, 4.68e+02
	401	15	8.8	390	30	HS349XC1	H.sapiens DNA segment 4.68e+02
c	402	15	8.8	400	30	G19483	human STS SHGC-18782. 4.68e+02
	403	15	8.8	417	20	NAU58521	Naja atra cobrotoxin'' 4.68e+02
	404	15	8.8	417	20	NAU77491	Naja atra cobrotoxin'' 4.68e+02
	405	15	8.8	419	30	G27830	human STS SHGC-33390. 4.68e+02
	406	15	8.8	425	18	TRBVSG172	T.brucei variant surfa 4.68e+02
	407	15	8.8	427	30	G26839	human STS STS_M33875. 4.68e+02
	408	15	8.8	429	18	TRBVSG18	T.brucei variant surfa 4.68e+02
	409	15	8.8	430	30	G36644	SHGC-53703 Human Homo 4.68e+02
	410	15	8.8	430	25	HSRPTAN	H.sapiens microsatelli 4.68e+02
	411	15	8.8	432	30	G06998	human STS WI-8795. 4.68e+02
c	412	15	8.8	444	30	G41517	Z4425 Zebrafish AB Dan 4.68e+02
	413	15	8.8	475	20	NAU77490	Naja atra cobrotoxin' 4.68e+02
c	414	15	8.8	501	19	BTMSAT60	B.taurus microsatellit 4.68e+02
	415	15	8.8	503	20	NAU77492	Naja atra cobrotoxin'' 4.68e+02
	416	15	8.8	503	20	NAU58520	Naja atra cobrotoxin'' 4.68e+02
	417	15	8.8	537	26	HSCARNPSE	Homo sapiens carnitine 4.68e+02
	418	15	8.8	539	30	G08600	human STS CHLC.GATA21C 4.68e+02
	419	15	8.8	539	30	G20197	sWSS3287 Eric D. Green 4.68e+02
	420	15	8.8	540	30	G31972	STS MCW177, sequence 4.68e+02
c	421	15	8.8	581	30	G41258	B316G4/SP6 Human Chrom 4.68e+02
	422	15	8.8	582	29	CEITRNAA	Ceratitis capitata mat 4.68e+02
	423	15	8.8	617	23	CSSOD	Citrus sinensis mRNA f 4.68e+02
c	424	15	8.8	624	18	MTLSCOIR2	Lasaea.sp (ir2) mitoch 4.68e+02
c	425	15	8.8	624	18	MTLSCOFR	Lasaea.sp (frmed) mito 4.68e+02
c	426	15	8.8	624	18	MTLSCOIRB	Lasaea.sp (ir1) mitoch 4.68e+02
c	427	15	8.8	624	18	MTLSCOIR3	Lasaea.sp (ir3) mitoch 4.68e+02
	428	15	8.8	647	26	AG000932	Homo sapiens genomic D 4.68e+02
c	429	15	8.8	648	25	HSU04688	Human olfactory recept 4.68e+02
c	430	15	8.8	648	27	U86272	Homo sapiens olfactory 4.68e+02
c	431	15	8.8	648	25	HSU04687	Human olfactory recept 4.68e+02
c	432	15	8.8	648	25	HSU04713	Human olfactory recept 4.68e+02
c	433	15	8.8	649	27	U86243	Homo sapiens olfactory 4.68e+02
c	434	15	8.8	649	27	U86244	Homo sapiens olfactory 4.68e+02
c	435	15	8.8	649	27	U86245	Homo sapiens olfactory 4.68e+02
	436	15	8.8	664	26	AG000289	Homo sapiens genomic D 4.68e+02
	437	15	8.8	665	26	AG000290	Homo sapiens genomic D 4.68e+02
	438	15	8.8	686	18	MIAPORI	Starfish mitochondrial 4.68e+02
	439	15	8.8	704	18	DDP24	D.discoideum mRNA for 4.68e+02
	440	15	8.8	707	26	AG000933	Homo sapiens genomic D 4.68e+02
c	441	15	8.8	739	26	HSUROKR7	Human urokinase-type p 4.68e+02
c	442	15	8.8	755	28	RATIGFII2	Rat insulin-like growt 4.68e+02
	443	15	8.8	792	25	HUM1RBPG4	Human (clone HGL-3) in 4.68e+02
	444	15	8.8	792	25	HSPST1CA	H.sapiens Pst 1 fragme 4.68e+02
c	445	15	8.8	848	18	ECU39390	Euphranta canadensis 1 4.68e+02
c	446	15	8.8	873	18	BMU39367	Blepharoneura manchest 4.68e+02
	447	15	8.8	911	30	G41986	SHGC-69031 Human Homo 4.68e+02
c	448	15	8.8	921	23	KLCEN1	K.lactis centromere 1 4.68e+02
	449	15	8.8	961	18	AF052048	Ostertagia ostertagi G 4.68e+02
	450	15	8.8	1020	26	HUMPROD04	Human prolyl 4-hydroxy 4.68e+02

	451	15	8.8	1022	28	MUSLRP01	Mouse leukocyte common 4.68e+02
c	452	15	8.8	1057	25	HSU09346	Human urokinase plasmi 4.68e+02
	453	15	8.8	1068	33	PUUNCPS	Puumala virus (strain 4.68e+02
c	454	15	8.8	1114	19	BTU95811	Bos taurus GRO-gamma m 4.68e+02
	455	15	8.8	1138	24	ATU16765	Arabidopsis thaliana T 4.68e+02
	456	15	8.8	1155	23	PUECHSY	Pueraria lobata gene f 4.68e+02
c	457	15	8.8	1177	18	AF067992	Dendroctonus adjunctus 4.68e+02
	458	15	8.8	1205	23	MIKTRRNA	Kluyveromyces thermoto 4.68e+02
	459	15	8.8	1231	28	MUSEN1S4	Mouse Edn1 gene for pr 4.68e+02
	460	15	8.8	1231	28	RNCACCP	R.norvegicus mRNA for 4.68e+02
	461	15	8.8	1304	28	MMHOX31	Mouse mRNA for Hox 3.1 4.68e+02
c	462	15	8.8	1314	15	CLO7AHD	Clostridium sordellii 4.68e+02
c	463	15	8.8	1319	23	LJRAB11D	L.japonicus mRNA for s 4.68e+02
c	464	15	8.8	1332	28	MUSTNF	Mouse tumor necrosis f 4.68e+02
c	465	15	8.8	1344	20	CHKCL3	Chicken glutathione tr 4.68e+02
c	466	15	8.8	1349	25	HUMMO3A	Human monocyte activat 4.68e+02
c	467	15	8.8	1363	25	HSUPARAA	H.sapiens urokinase pl 4.68e+02
	468	15	8.8	1377	18	PFAPHOSKIN	P.falciparum 3-phospho 4.68e+02
c	469	15	8.8	1400	21	A18757	u-PA receptor. 4.68e+02
c	470	15	8.8	1400	21	I21305	Sequence 12 from paten 4.68e+02
c	471	15	8.8	1407	21	I08456	Sequence 8 from Patent 4.68e+02
	472	15	8.8	1450	28	MMHMOXY1	M.musculus gene for ha 4.68e+02
	473	15	8.8	1457	23	SCCLB3	S.cerevisiae CLB3 gene 4.68e+02
	474	15	8.8	1459	15	BACFLIM	Bacillus subtilis flag 4.68e+02
	475	15	8.8	1470	28	MMHOX31R	Mouse mRNA for Hox-3.1 4.68e+02
	476	15	8.8	1481	25	HSG17R	Human mRNA for glutath 4.68e+02
c	477	15	8.8	1503	28	MUSGNA14A	Mouse G protein alpha 4.68e+02
c	478	15	8.8	1536	18	SLMMTATPAA	P.polycephalum mitocho 4.68e+02
	479	15	8.8	1586	28	D43775	Mouse mRNA for endothe 4.68e+02
c	480	15	8.8	1590	31	SYNMTATPAA	Synthetic P.polycephal 4.68e+02
c	481	15	8.8	1595	23	ZMU60507	Zea mays actin (Mat95) 4.68e+02
c	482	15	8.8	1608	26	HSU08839	Human urokinase-type p 4.68e+02
c	483	15	8.8	1612	18	CPRRNAA2	Cryptosporidium parvum 4.68e+02
	484	15	8.8	1632	20	GGVI01	Gallus gallus major vi 4.68e+02
c	485	15	8.8	1644	21	I07534	Sequence 20 from Paten 4.68e+02
c	486	15	8.8	1644	28	MMTNFR	Mouse mRNA for tumour 4.68e+02
c	487	15	8.8	1665	23	ATHAAT	Arabidopsis thaliana a 4.68e+02
c	488	15	8.8	1685	21	I88787	Sequence 1 from patent 4.68e+02
c	489	15	8.8	1685	21	A37879	Sequence 1 from Patent 4.68e+02
c	490	15	8.8	1685	23	ATPPPMR	A.thaliana AAP1 mRNA f 4.68e+02
	491	15	8.8	1704	18	TBGP04	Trypanosoma brucei mRN 4.68e+02
	492	15	8.8	1734	23	YSCMCYBA	Saccharomyces cerevisi 4.68e+02
c	493	15	8.8	1735	27	HSU48361	Human NGFI-A binding p 4.68e+02
c	494	15	8.8	1738	18	AF025793	Drosophila melanogaste 4.68e+02
c	495	15	8.8	1755	20	U01349	Torpedo californica pr 4.68e+02
c	496	15	8.8	1756	26	AF072321	Papio sp. clone 3 DAZL 4.68e+02
c	497	15	8.8	1764	28	HAMP450	Hamster major 3-methyl 4.68e+02
	498	15	8.8	1769	23	CDXFRA	Curdiea flabellata DNA 4.68e+02
	499	15	8.8	1770	28	MMU35233	Mus musculus preproend 4.68e+02
	500	15	8.8	1770	23	GRCFRE	Gracilaria chilensis D 4.68e+02
c	501	15	8.8	1771	28	HAMCYP41	Hamster cytochrome P-4 4.68e+02

c	502	15	8.8	1794	18	CTBRC54S	Chironomus thummi BRc 4.68e+02
	503	15	8.8	1836	23	MIAN09	Aspergillus nidlans mi 4.68e+02
	504	15	8.8	1847	23	SCYDL155W	S.cerevisiae chromosom 4.68e+02
c	505	15	8.8	1849	21	A45716	Sequence 1 from Patent 4.68e+02
	506	15	8.8	1900	16	CHTHSPR	Chlamydia trachomatis 4.68e+02
	507	15	8.8	1980	15	AGU03517	Arthrobacter globiform 4.68e+02
	508	15	8.8	1982	18	KSC228783	Kuhnia scombri 18S rRN 4.68e+02
c	509	15	8.8	2014	23	APU12757	Acer pseudoplatanus la 4.68e+02
	510	15	8.8	2030	20	GGVITUP	Chicken vitellogenin g 4.68e+02
c	511	15	8.8	2098	25	HSFASAPOB	H.sapiens of FAS/Apo1 4.68e+02
	512	15	8.8	2179	23	PEACL22	Pisum sativum ribosoma 4.68e+02
	513	15	8.8	2233	18	TRBVSGI	Trypanosma brucei bruc 4.68e+02
c	514	15	8.8	2242	18	EHU48386	Entamoeba histolytica 4.68e+02
	515	15	8.8	2282	15	CHTOMP3D	Chlamydia trachomatis 4.68e+02
	516	15	8.8	2362	23	RCRICIN2	Ricinus communis ricin 4.68e+02
c	517	15	8.8	2375	25	HSMDIPEPIV	H.sapiens dipeptidyl p 4.68e+02
	518	15	8.8	2402	20	NACOBROTX	N.atra gene encoding c 4.68e+02
	519	15	8.8	2444	21	E02164	Human protein disulfid 4.68e+02
	520	15	8.8	2444	21	E03087	cDNA encoding human pr 4.68e+02
	521	15	8.8	2454	21	I29876	Sequence 1 from patent 4.68e+02
	522	15	8.8	2454	21	E06719	cDNA encoding human pr 4.68e+02
	523	15	8.8	2514	25	HUMTHBP	Human thyroid hormone 4.68e+02
c	524	15	8.8	2546	15	PSU40485	Pseudomonas sp. alkali 4.68e+02
	525	15	8.8	2613	20	MIRPRNAR	R.pipiens mitochondria 4.68e+02
c	526	15	8.8	2617	28	RATIGFIIA1	Rat insulin-like growt 4.68e+02
	527	15	8.8	2621	28	RNU03416	Rattus norvegicus neur 4.68e+02
c	528	15	8.8	2645	24	YSCCTK1A	S.cerevisiae CTD kinas 4.68e+02
c	529	15	8.8	2673	23	SCYOR243C	S.cerevisiae chromosom 4.68e+02
c	530	15	8.8	2709	28	MUSTNFB	Mouse tumor necrosis f 4.68e+02
c	531	15	8.8	2748	23	TOBPOXAN	Nicotiana tabacum anio 4.68e+02
	532	15	8.8	2759	28	RNU03417	Rattus norvegicus neur 4.68e+02
	533	15	8.8	2786	22	AF090369	Lactococcus phage P008 4.68e+02
c	534	15	8.8	2821	23	GMRPRP3	Glycine max RPRP3 gene 4.68e+02
c	535	15	8.8	2828	15	BACORI43	B.thuringiensis replic 4.68e+02
c	536	15	8.8	2843	23	PCCUBI42	P.crispum gene Pcubi4- 4.68e+02
c	537	15	8.8	2850	23	SCYKL139W	S.cerevisiae chromosom 4.68e+02
c	538	15	8.8	2869	25	GCRDELGLB	G.crassicaudatus beta 4.68e+02
c	539	15	8.8	2892	28	MUSIGFII2	Mouse insulin-like gro 4.68e+02
	540	15	8.8	3025	20	CHKFRA2A1	Chicken fra-2 oncogene 4.68e+02
	541	15	8.8	3070	28	RNNFMMR	R.norvegicus mRNA for 4.68e+02
	542	15	8.8	3112	18	DROGATA	Drosophila melanogaste 4.68e+02
c	543	15	8.8	3115	20	D87871	Gallus gallus mRNA for 4.68e+02
	544	15	8.8	3119	21	I09339	Sequence 1 from Patent 4.68e+02
c	545	15	8.8	3195	23	SCYJL206C	S.cerevisiae chromosom 4.68e+02
	546	15	8.8	3216	25	HUMCBF	Human CCAAT-box-bindin 4.68e+02
c	547	15	8.8	3217	28	AF011336	Mus musculus putative 4.68e+02
	548	15	8.8	3289	15	RHMLCRATOE	Rhizobium sp. lcrABCDE 4.68e+02
c	549	15	8.8	3329	23	LSTISF1	Lactuca sativa T-DNA i 4.68e+02
	550	15	8.8	3400	16	BSU17292	Bacillus subtilis ATPa 4.68e+02
c	551	15	8.8	3407	25	HSPCHDP7	Human pcHDP7 mRNA for 4.68e+02
c	552	15	8.8	3500	28	MUSTNFC	Mouse tumor necrosis f 4.68e+02

c	553	15	8.8	3501	21	E01987	DNA encoding genomic m 4.68e+02
c	554	15	8.8	3543	23	GMGLY4MU	Soybean mutant Gy(4) g 4.68e+02
c	555	15	8.8	3592	23	SCYBL053W	S.cerevisiae chromosom 4.68e+02
c	556	15	8.8	3638	23	RCHCHS1	Rhizopus oligosporus c 4.68e+02
c	557	15	8.8	3815	28	MMU68414	Mus musculus strain A/ 4.68e+02
c	558	15	8.8	3821	28	MMU68415	Mus musculus strain BA 4.68e+02
c	559	15	8.8	3832	28	MMU68416	Mus musculus strain C5 4.68e+02
	560	15	8.8	3960	25	HUMP107B	Human retinoblastoma r 4.68e+02
c	561	15	8.8	3967	18	DME011114	Drosophila melanogaste 4.68e+02
c	562	15	8.8	4026	22	AF007792	Bacteriophage Mu late 4.68e+02
	563	15	8.8	4057	28	RATBAND33E	Rat band 3 Cl-/HCO3- e 4.68e+02
c	564	15	8.8	4110	23	SCGACOP	S.cerevisiae alpha-COP 4.68e+02
	565	15	8.8	4275	25	HUMIRBP	Human interphotorecept 4.68e+02
c	566	15	8.8	4294	15	AF042099	Sulfolobus metallicus 4.68e+02
	567	15	8.8	4308	20	GGRNACMET	G.gallus mRNA for c-me 4.68e+02
	568	15	8.8	4436	24	SCYDL145C	S.cerevisiae chromosom 4.68e+02
	569	15	8.8	4477	15	PPDMPEPGH	P.putida dmpE,P,G,H & 4.68e+02
c	570	15	8.8	4599	26	AB002358	Human mRNA for KIAA036 4.68e+02
	571	15	8.8	4600	15	SA1234	S.aureus orfs 1,2,3 & 4.68e+02
c	572	15	8.8	4650	18	DMRNAHDCP	D.melanogaster mRNA fo 4.68e+02
c	573	15	8.8	4915	26	HSZPHSAL2	H.sapiens mRNA for zin 4.68e+02
	574	15	8.8	5043	24	SCYDL144C	S.cerevisiae chromosom 4.68e+02
	575	15	8.8	5245	28	RATRGPI	Rat gene for regenerat 4.68e+02
c	576	15	8.8	5383	26	AB011126	Homo sapiens mRNA for 4.68e+02
	577	15	8.8	5407	28	RATLITHOST	Rat lithostathine gene 4.68e+02
	578	15	8.8	5449	28	MMU44725	Mus musculus mast cell 4.68e+02
c	579	15	8.8	5495	18	PFSC03099	Plasmodium falciparum 4.68e+02
	580	15	8.8	5498	18	PFSC04006	Plasmodium falciparum 4.68e+02
c	581	15	8.8	5526	23	MDUB9156	Malus domestica ACC sy 4.68e+02
c	582	15	8.8	5612	16	AF043470	Escherichia coli plasm 4.68e+02
c	583	15	8.8	5636	18	AB002270	Bombyx mori L1Bm DNA, 4.68e+02
c	584	15	8.8	5771	18	PFSC04079	Plasmodium falciparum 4.68e+02
c	585	15	8.8	5788	15	TMNUSGGE	T.maritima nusG gene a 4.68e+02
c	586	15	8.8	5798	23	AF031649	Arabidopsis thaliana n 4.68e+02
	587	15	8.8	5830	18	PFSC04105	Plasmodium falciparum 4.68e+02
	588	15	8.8	6104	18	PFSC04016	Plasmodium falciparum 4.68e+02
c	589	15	8.8	6393	18	CEC49G9	Caenorhabditis elegans 4.68e+02
c	590	15	8.8	6475	23	SCYJL207C	S.cerevisiae chromosom 4.68e+02
c	591	15	8.8	6477	23	SCASCOPI	S.cerevisiae DNA for a 4.68e+02
	592	15	8.8	6760	15	YYYXYLRN	Thermophillic bacteria 4.68e+02
c	593	15	8.8	6788	20	XELNMUSMYO	Xenopus laevis nonmusc 4.68e+02
	594	15	8.8	6892	23	SPURA1	S.pombe mRNA for URA1. 4.68e+02
	595	15	8.8	7078	25	HUMTGKVA	Human Val-tRNA and Lys 4.68e+02
c	596	15	8.8	7378	15	U39724	Mycoplasma genitalium 4.68e+02
c	597	15	8.8	7515	27	HSU26044	Human fatty acid synth 4.68e+02
	598	15	8.8	7642	26	AF002246	Homo sapiens neural ce 4.68e+02
c	599	15	8.8	8285	16	MMU47023	Methanococcus maripalu 4.68e+02
	600	15	8.8	8302	15	BSPLAAO	B. subtilis FlaA locus 4.68e+02
	601	15	8.8	8302	15	BACFLAAOA	B.subtilis flaA locus 4.68e+02
	602	15	8.8	8726	18	PFSC03088	Plasmodium falciparum 4.68e+02
	603	15	8.8	9170	18	CELT14A8	Caenorhabditis elegans 4.68e+02

	604	15	8.8	9711	26	HUMIRBPG	Homo sapiens interstit 4.68e+02
	605	15	8.8	9735	15	U67535	Methanococcus jannasch 4.68e+02
	606	15	8.8	9885	18	PFSC03054	Plasmodium falciparum 4.68e+02
c	607	15	8.8	10120	15	TMRPO	T.maritima ribosomal p 4.68e+02
	608	15	8.8	10272	28	MUSHOXMAA	Mouse Hox-3.1 gene and 4.68e+02
	609	15	8.8	10523	15	U67498	Methanococcus jannasch 4.68e+02
	610	15	8.8	10666	15	U67518	Methanococcus jannasch 4.68e+02
	611	15	8.8	10901	18	PFSC03070	Plasmodium falciparum 4.68e+02
	612	15	8.8	11020	15	U67512	Methanococcus jannasch 4.68e+02
c	613	15	8.8	11107	15	U32783	Haemophilus influenzae 4.68e+02
	614	15	8.8	11198	18	PFSC04057	Plasmodium falciparum 4.68e+02
c	615	15	8.8	11288	23	SPCEN1A	S.pombe cenl central r 4.68e+02
	616	15	8.8	11519	15	AE001159	Borrelia burgdorferi (4.68e+02
	617	15	8.8	11684	18	CELH22D07	Caenorhabditis elegans 4.68e+02
	618	15	8.8	11772	18	PFSC03022	Plasmodium falciparum 4.68e+02
	619	15	8.8	11958	18	AC002445	Drosophila melanogaste 4.68e+02
c	620	15	8.8	12043	15	AE001120	Borrelia burgdorferi (4.68e+02
	621	15	8.8	12794	17	CE4R79	Caenorhabditis elegans 4.68e+02
	622	15	8.8	12806	15	U67469	Methanococcus jannasch 4.68e+02
c	623	15	8.8	12965	26	HSN104C4	Human DNA sequence fro 4.68e+02
	624	15	8.8	13407	16	AE001312	Chlamydia trachomatis 4.68e+02
c	625	15	8.8	13742	23	ENU75347	Emericella nidulans fa 4.68e+02
c	626	15	8.8	13990	15	AE001166	Borrelia burgdorferi (4.68e+02
c	627	15	8.8	14001	18	PFCOMPIRB	P.falciparum complete 4.68e+02
	628	15	8.8	14326	26	HS549K18	Human DNA sequence *** 4.68e+02
	629	15	8.8	14440	23	EMEMTDNA	Emericella nidulans mt 4.68e+02
	630	15	8.8	15213	28	MMU06950	Mus musculus C57BL/6 l 4.68e+02
c	631	15	8.8	15688	15	HPAE000579	Helicobacter pylori se 4.68e+02
c	632	15	8.8	15719	18	MIALDNASQ	A.lixula mitochondrial 4.68e+02
	633	15	8.8	15921	18	AF034856	Drosophila melanogaste 4.68e+02
c	634	15	8.8	16260	18	D16387	Asterina pectinifera, 4.68e+02
	635	15	8.8	17576	18	CEY48E1A	Caenorhabditis elegans 4.68e+02
c	636	15	8.8	17816	18	PFSC03051	Plasmodium falciparum 4.68e+02
c	637	15	8.8	20535	18	CET04B2	Caenorhabditis elegans 4.68e+02
	638	15	8.8	20759	18	PFSC03057	Plasmodium falciparum 4.68e+02
c	639	15	8.8	20823	18	CELT24H7	Caenorhabditis elegans 4.68e+02
	640	15	8.8	21153	23	YSCMTCG13	Saccharomyces cerevisi 4.68e+02
	641	15	8.8	21173	18	CEF21D5	Caenorhabditis elegans 4.68e+02
c	642	15	8.8	21262	18	PFSC03028	Plasmodium falciparum 4.68e+02
	643	15	8.8	21961	18	CEF14F4	Caenorhabditis elegans 4.68e+02
	644	15	8.8	22617	26	HSA005577	Homo sapiens pfkfb2 ge 4.68e+02
	645	15	8.8	23096	23	SCIV23	S.cerevisiae chrIV gen 4.68e+02
c	646	15	8.8	23140	18	CELF18E3	Caenorhabditis elegans 4.68e+02
c	647	15	8.8	24342	18	CEF49A5	Caenorhabditis elegans 4.68e+02
	648	15	8.8	26343	18	CEY45P10A	Caenorhabditis elegans 4.68e+02
c	649	15	8.8	27263	24	AB016881	Arabidopsis thaliana g 4.68e+02
	650	15	8.8	27394	18	CEF32B6	Caenorhabditis elegans 4.68e+02
c	651	15	8.8	27874	28	MMU71085	Mus musculus insulin-l 4.68e+02
	652	15	8.8	27981	18	CE2K945	Caenorhabditis elegans 4.68e+02
c	653	15	8.8	28611	18	CEC02D4	Caenorhabditis elegans 4.68e+02
c	654	15	8.8	28650	23	SPAC22G7	S.pombe chromosome I c 4.68e+02

	655	15	8.8	29867	18	CEF46A9	Caenorhabditis elegans 4.68e+02
c	656	15	8.8	30000	28	RNIGF2	Rat IGFII gene for ins 4.68e+02
	657	15	8.8	30500	18	CER10E12	Caenorhabditis elegans 4.68e+02
	658	15	8.8	31049	18	CELF16H11	Caenorhabditis elegans 4.68e+02
c	659	15	8.8	31531	18	CELC27A12	Caenorhabditis elegans 4.68e+02
c	660	15	8.8	32560	23	SCI112260	S.cerevisiae DNA for o 4.68e+02
	661	15	8.8	32566	18	CEY94A7B	Caenorhabditis elegans 4.68e+02
c	662	15	8.8	32589	18	CEF25H8	Caenorhabditis elegans 4.68e+02
	663	15	8.8	32613	26	AF045449	Homo sapiens chromosom 4.68e+02
c	664	15	8.8	32816	18	CELF56E3	Caenorhabditis elegans 4.68e+02
c	665	15	8.8	32911	18	CELT23C6	Caenorhabditis elegans 4.68e+02
	666	15	8.8	33330	18	CELF37C12	Caenorhabditis elegans 4.68e+02
	667	15	8.8	33593	22	AF063097	Bacteriophage P2, comp 4.68e+02
c	668	15	8.8	33658	18	CEC33G3	Caenorhabditis elegans 4.68e+02
	669	15	8.8	33760	25	HSNFG9	Human DNA sequence fro 4.68e+02
	670	15	8.8	34461	25	HSU82205	Human chromosome 10 cl 4.68e+02
	671	15	8.8	34831	18	CEC47G2	Caenorhabditis elegans 4.68e+02
c	672	15	8.8	34836	17	AC002469	*** SEQUENCING IN PROG 4.68e+02
c	673	15	8.8	35221	18	CET04C12	Caenorhabditis elegans 4.68e+02
c	674	15	8.8	35379	18	CEF35G2	Caenorhabditis elegans 4.68e+02
c	675	15	8.8	35540	24	SPBC36	S.pombe chromosome II 4.68e+02
c	676	15	8.8	35673	18	CELK06A9	Caenorhabditis elegans 4.68e+02
	677	15	8.8	35743	18	CELF31E8	Caenorhabditis elegans 4.68e+02
c	678	15	8.8	36017	23	SCPRP21	S.cerevisiae PRP21, SP 4.68e+02
	679	15	8.8	36080	18	CER10D12	Caenorhabditis elegans 4.68e+02
c	680	15	8.8	36087	18	CELZC21	C. elegans cosmid ZC21 4.68e+02
	681	15	8.8	36849	15	AE000788	Borrelia burgdorferi p 4.68e+02
c	682	15	8.8	36922	18	PFSC03026	Plasmodium falciparum 4.68e+02
c	683	15	8.8	37195	18	CELR07G3	Caenorhabditis elegans 4.68e+02
	684	15	8.8	37476	18	CELF53B1	Caenorhabditis elegans 4.68e+02
	685	15	8.8	37599	18	CELK05B2	Caenorhabditis elegans 4.68e+02
	686	15	8.8	37628	18	CEF31D4	Caenorhabditis elegans 4.68e+02
	687	15	8.8	37700	26	AC000390	Genomic sequence from 4.68e+02
	688	15	8.8	37794	18	CELF13H8	Caenorhabditis elegans 4.68e+02
c	689	15	8.8	37799	18	CELC05E11	Caenorhabditis elegans 4.68e+02
c	690	15	8.8	37854	26	AC004172	Homo sapiens clone UWG 4.68e+02
c	691	15	8.8	38060	18	CET13F2	Caenorhabditis elegans 4.68e+02
c	692	15	8.8	38190	17	AC004193	*** SEQUENCING IN PROG 4.68e+02
	693	15	8.8	38514	18	CEC27A7	Caenorhabditis elegans 4.68e+02
	694	15	8.8	38757	26	AC005380	Homo sapiens chromosom 4.68e+02
c	695	15	8.8	39482	18	CET04A8	Caenorhabditis elegans 4.68e+02
c	696	15	8.8	39599	17	HSE140F8	Human DNA sequence *** 4.68e+02
	697	15	8.8	40176	27	HSU58675	Human olfactory recept 4.68e+02
c	698	15	8.8	40222	18	CELF22B7	C.aenorhabditis elegan 4.68e+02
	699	15	8.8	40573	27	HSU15422	Human protamine 1 (PRM 4.68e+02
	700	15	8.8	40873	17	CEH04I09	Caenorhabditis elegans 4.68e+02
	701	15	8.8	40876	18	CELM02D8	Caenorhabditis elegans 4.68e+02
c	702	15	8.8	41958	26	AC002380	Human BAC clone GS330J 4.68e+02
c	703	15	8.8	42297	18	CELC03B1	Caenorhabditis elegans 4.68e+02
c	704	15	8.8	42810	18	CELB0403	Caenorhabditis elegans 4.68e+02
c	705	15	8.8	43802	18	CEK07F5	Caenorhabditis elegans 4.68e+02

c	706	15	8.8	43910	18	CELC06G3	Caenorhabditis elegans 4.68e+02
	707	15	8.8	43952	26	HSU105G4	Human DNA sequence fro 4.68e+02
	708	15	8.8	44237	18	CER05H10	Caenorhabditis elegans 4.68e+02
c	709	15	8.8	44352	18	DMC4F1	Drosophila melanogaste 4.68e+02
c	710	15	8.8	44399	27	AC005795	Homo sapiens chromosom 4.68e+02
c	711	15	8.8	45911	24	AB017070	Arabidopsis thaliana g 4.68e+02
	712	15	8.8	47045	18	CELC17H12	Caenorhabditis elegans 4.68e+02
c	713	15	8.8	48096	18	CET20G5	Caenorhabditis elegans 4.68e+02
	714	15	8.8	51019	17	HS349H11	Human DNA sequence *** 4.68e+02
c	715	15	8.8	51498	26	HSU82694	Human cosmid Xq28_IA64 4.68e+02
	716	15	8.8	54724	17	AC002555	*** SEQUENCING IN PROG 4.68e+02
	717	15	8.8	55593	15	BSAJ2571	Bacillus subtilis 168 4.68e+02
c	718	15	8.8	57113	27	OCU60902	Otolemur crassicaudatu 4.68e+02
	719	15	8.8	59648	24	AB016870	Arabidopsis thaliana g 4.68e+02
	720	15	8.8	63495	18	AC004244	Drosophila melanogaste 4.68e+02
c	721	15	8.8	64767	26	U69730	Homo sapiens Xp22 Cosm 4.68e+02
c	722	15	8.8	65114	18	AB011813	Drosophila melanogaste 4.68e+02
c	723	15	8.8	66237	24	AB016892	Arabidopsis thaliana g 4.68e+02
	724	15	8.8	67531	17	AC005290	*** SEQUENCING IN PROG 4.68e+02
c	725	15	8.8	68326	17	HS359F1	Human DNA sequence *** 4.68e+02
	726	15	8.8	69034	18	AF007261	Reclinomonas americana 4.68e+02
c	727	15	8.8	69850	23	AB010700	Arabidopsis thaliana g 4.68e+02
c	728	15	8.8	70504	16	AF053946	Yersinia pestis plasmi 4.68e+02
c	729	15	8.8	72119	23	SCD9717	Saccharomyces cerevisi 4.68e+02
c	730	15	8.8	72732	18	AC004715	Drosophila melanogaste 4.68e+02
c	731	15	8.8	74432	26	HS465G10B	Human DNA sequence fro 4.68e+02
	732	15	8.8	74717	18	AC004563	Drosophila melanogaste 4.68e+02
	733	15	8.8	75270	27	AF023268	Homo sapiens clk2 kina 4.68e+02
	734	15	8.8	76329	23	AB007648	Arabidopsis thaliana g 4.68e+02
	735	15	8.8	76854	18	AC003052	Drosophila melanogaste 4.68e+02
	736	15	8.8	77905	24	T2L5	Arabidopsis thaliana B 4.68e+02
c	737	15	8.8	78439	27	HSU85197	Homo sapiens BAC480, c 4.68e+02
	738	15	8.8	78748	28	AC000400	Genomic sequence from 4.68e+02
c	739	15	8.8	79574	23	ATAC002340	Arabidopsis thaliana c 4.68e+02
c	740	15	8.8	80315	23	AB005232	Arabidopsis thaliana g 4.68e+02
c	741	15	8.8	80559	3	AC004570	*** SEQUENCING IN PROG 4.68e+02
c	742	15	8.8	80803	3	AC004275	*** SEQUENCING IN PROG 4.68e+02
	743	15	8.8	80993	18	AC004313	Drosophila melanogaste 4.68e+02
c	744	15	8.8	81053	18	AC005762	Drosophila melanogaste 4.68e+02
	745	15	8.8	81405	3	AC004717	*** SEQUENCING IN PROG 4.68e+02
c	746	15	8.8	81984	25	HSJ30E17	Human DNA sequence fro 4.68e+02
	747	15	8.8	82188	23	AB010076	Arabidopsis thaliana g 4.68e+02
c	748	15	8.8	82291	27	AC005376	Homo sapiens PAC clone 4.68e+02
c	749	15	8.8	83024	26	AC000389	Genomic sequence from 4.68e+02
	750	15	8.8	83822	17	HSAJ9611	Homo sapiens *** SEQUE 4.68e+02
c	751	15	8.8	84056	28	AC002121	Genomic sequence from 4.68e+02
	752	15	8.8	84485	18	AC004351	Drosophila melanogaste 4.68e+02
	753	15	8.8	84537	17	HSU80B1	Human DNA sequence *** 4.68e+02
	754	15	8.8	85444	26	AP000023	Homo sapiens genomic D 4.68e+02
	755	15	8.8	86004	27	HS51J23	Human DNA sequence fro 4.68e+02
	756	15	8.8	86491	26	AC004810	Homo sapiens chromosom 4.68e+02

c	757	15	8.8	86894	26	AC003676	Homo sapiens chromosom 4.68e+02
c	758	15	8.8	86896	26	HS93N13	Human DNA sequence fro 4.68e+02
	759	15	8.8	87768	24	AC005388	Arabidopsis thaliana c 4.68e+02
c	760	15	8.8	88274	17	AC005693	*** SEQUENCING IN PROG 4.68e+02
	761	15	8.8	88987	26	AC004541	Homo sapiens PAC clone 4.68e+02
	762	15	8.8	89521	17	HS751L1	Human DNA sequence *** 4.68e+02
c	763	15	8.8	89743	28	AC003062	Mouse Chromosome 16 Re 4.68e+02
	764	15	8.8	89791	18	AC005268	Drosophila melanogaste 4.68e+02
	765	15	8.8	91406	17	HS301K23	Human DNA sequence *** 4.68e+02
	766	15	8.8	91638	17	AC002489	*** SEQUENCING IN PROG 4.68e+02
	767	15	8.8	92720	16	AB011549	Escherichia coli plasm 4.68e+02
c	768	15	8.8	92802	17	AC004710	*** SEQUENCING IN PROG 4.68e+02
c	769	15	8.8	92879	17	AC004450	*** SEQUENCING IN PROG 4.68e+02
	770	15	8.8	93578	25	HSAC000055	Human PAC clone DJ073F 4.68e+02
	771	15	8.8	93932	18	AC004766	Drosophila melanogaste 4.68e+02
c	772	15	8.8	94482	24	ATF26P21	Arabidopsis thaliana D 4.68e+02
	773	15	8.8	94647	28	MUSTCRA	Mouse T-cell receptor 4.68e+02
c	774	15	8.8	94882	17	AC002317	*** SEQUENCING IN PROG 4.68e+02
c	775	15	8.8	95855	25	HSAC000115	Human BAC clone GS188P 4.68e+02
c	776	15	8.8	95886	17	CEH27C11	Caenorhabditis elegans 4.68e+02
c	777	15	8.8	96608	26	HS308I13	Human DNA sequence fro 4.68e+02
	778	15	8.8	96792	26	HSU95741	Human chromosome 16p13 4.68e+02
	779	15	8.8	97684	27	HS309I22	Human DNA sequence fro 4.68e+02
c	780	15	8.8	98025	26	AF003530	Homo sapiens homeobox 4.68e+02
c	781	15	8.8	98822	23	ATT6K21	Arabidopsis thaliana D 4.68e+02
	782	15	8.8	99033	27	AC005704	Homo sapiens Xp22 bins 4.68e+02
c	783	15	8.8	99121	24	AF069716	Arabidopsis Thaliana B 4.68e+02
	784	15	8.8	99359	23	ATAC003673	Arabidopsis thaliana c 4.68e+02
c	785	15	8.8	99359	23	ATAC003673	Arabidopsis thaliana c 4.68e+02
	786	15	8.8	100000	26	AP000038	Homo sapiens genomic D 4.68e+02
	787	15	8.8	100000	26	AP000056	Homo sapiens genomic D 4.68e+02
c	788	15	8.8	100000	26	AP000019	Homo sapiens genomic D 4.68e+02
	789	15	8.8	100375	26	HS193B12	Human DNA sequence fro 4.68e+02
	790	15	8.8	100783	17	HS344I7	Human DNA sequence *** 4.68e+02
	791	15	8.8	100793	26	AF070717	Homo sapiens BAC clone 4.68e+02
	792	15	8.8	100808	26	HS94G16	Human DNA sequence fro 4.68e+02
c	793	15	8.8	100849	25	HUMNEUROP	Human oligodendrocyte 4.68e+02
	794	15	8.8	101486	17	AC005228	*** SEQUENCING IN PROG 4.68e+02
c	795	15	8.8	102230	26	HS234H5	Homo sapiens DNA seque 4.68e+02
c	796	15	8.8	102312	24	ATAC004218	Arabidopsis thaliana c 4.68e+02
c	797	15	8.8	102750	26	AC004465	Homo sapiens 12q24 PAC 4.68e+02
	798	15	8.8	103319	26	AC004217	Homo sapiens 12q24.1 P 4.68e+02
c	799	15	8.8	103909	26	AC000100	Homo sapiens Chromosom 4.68e+02
c	800	15	8.8	103911	26	HUAC002565	Human Chromosome 16 BA 4.68e+02
c	801	15	8.8	103916	26	AC004225	Homo sapiens chromosom 4.68e+02
	802	15	8.8	104278	18	AC005269	Drosophila melanogaste 4.68e+02
c	803	15	8.8	104508	17	AC005662	*** SEQUENCING IN PROG 4.68e+02
c	804	15	8.8	105174	17	HS171H20	Human DNA sequence *** 4.68e+02
	805	15	8.8	105809	17	HS364I1	Human DNA sequence *** 4.68e+02
c	806	15	8.8	106172	26	AC004111	Homo sapiens BAC clone 4.68e+02
	807	15	8.8	106374	17	AC005623	*** SEQUENCING IN PROG 4.68e+02

c	808	15	8.8	107352	26	HS474A14	Human DNA sequence fro 4.68e+02
	809	15	8.8	107677	26	HS440021	Human DNA sequence fro 4.68e+02
c	810	15	8.8	107784	25	HSU82207	Human chromosome 10 cl 4.68e+02
c	811	15	8.8	108754	26	AC004226	Homo sapiens chromosom 4.68e+02
	812	15	8.8	108778	26	AC005194	Homo sapiens chromosom 4.68e+02
c	813	15	8.8	109276	17	AC005281	*** SEQUENCING IN PROG 4.68e+02
	814	15	8.8	109741	24	ATAC004005	Arabidopsis thaliana c 4.68e+02
	815	15	8.8	110000	17	CEY45F10	Caenorhabditis elegans 4.68e+02
	816	15	8.8	110000	17	CEY73F8	Caenorhabditis elegans 4.68e+02
	817	15	8.8	110000	17	CEY53F4	Caenorhabditis elegans 4.68e+02
	818	15	8.8	110149	24	ATAC005397	Arabidopsis thaliana c 4.68e+02
	819	15	8.8	110227	17	HS298J15	Human DNA sequence *** 4.68e+02
	820	15	8.8	111767	23	AF058914	Arabidopsis thaliana B 4.68e+02
c	821	15	8.8	111994	26	HS296K21	Human DNA sequence fro 4.68e+02
	822	15	8.8	112585	26	HS1059E7	Human DNA sequence fro 4.68e+02
	823	15	8.8	112661	26	AC004668	Homo sapiens BAC clone 4.68e+02
	824	15	8.8	113033	26	AC005070	Homo sapiens BAC clone 4.68e+02
	825	15	8.8	113950	24	ATAC004697	Arabidopsis thaliana c 4.68e+02
c	826	15	8.8	114173	26	HS884M20	Homo sapiens DNA seque 4.68e+02
	827	15	8.8	114231	27	HS390C10	Homo sapiens DNA seque 4.68e+02
	828	15	8.8	114827	26	AC004413	Homo sapiens PAC clone 4.68e+02
c	829	15	8.8	115835	27	HS28C20	Human DNA sequence fro 4.68e+02
	830	15	8.8	116561	27	HS753P9	Human DNA sequence fro 4.68e+02
c	831	15	8.8	117132	17	AC004954	*** SEQUENCING IN PROG 4.68e+02
	832	15	8.8	117161	17	CEY43F11	Caenorhabditis elegans 4.68e+02
c	833	15	8.8	118001	17	HS566L20	Human DNA sequence *** 4.68e+02
c	834	15	8.8	118684	26	AC004259	Human Chromosome 15q11 4.68e+02
	835	15	8.8	118995	26	AC005368	Homo sapiens chromosom 4.68e+02
c	836	15	8.8	120067	24	T25C13	Arabidopsis thaliana B 4.68e+02
c	837	15	8.8	120423	26	HS326L12	Homo sapiens DNA seque 4.68e+02
c	838	15	8.8	120703	26	AC002364	Homo sapiens Xp22 Cosm 4.68e+02
c	839	15	8.8	120821	27	AC004702	Homo sapiens chromosom 4.68e+02
	840	15	8.8	121055	17	AC004823	*** SEQUENCING IN PROG 4.68e+02
c	841	15	8.8	121780	27	HS774I24	Human DNA sequence fro 4.68e+02
c	842	15	8.8	122322	23	ATF5I18	Arabidopsis thaliana D 4.68e+02
c	843	15	8.8	122322	26	AC002448	Human PAC clone DJ0068 4.68e+02
	844	15	8.8	122733	17	AC004865	*** SEQUENCING IN PROG 4.68e+02
c	845	15	8.8	122809	17	HS881E24	Human DNA sequence *** 4.68e+02
c	846	15	8.8	122940	17	AC004709	*** SEQUENCING IN PROG 4.68e+02
c	847	15	8.8	123468	17	HS232N11	Human DNA sequence *** 4.68e+02
c	848	15	8.8	123521	27	HS75G13	Homo sapiens chromosom 4.68e+02
c	849	15	8.8	124321	26	HS417G15	Human DNA sequence fro 4.68e+02
c	850	15	8.8	124330	17	AC005062	*** SEQUENCING IN PROG 4.68e+02
	851	15	8.8	124497	26	HS232G24	Homo sapiens DNA seque 4.68e+02
c	852	15	8.8	126474	26	HS181N1	Human DNA sequence fro 4.68e+02
c	853	15	8.8	126939	17	HS159A19	Human DNA sequence *** 4.68e+02
c	854	15	8.8	127270	26	AC004740	Homo sapiens PAC clone 4.68e+02
c	855	15	8.8	128501	27	AC004675	Homo sapiens chromosom 4.68e+02
	856	15	8.8	128915	26	AC002418	Human Chromosome X, co 4.68e+02
	857	15	8.8	129169	17	AC005506	*** SEQUENCING IN PROG 4.68e+02
c	858	15	8.8	129169	17	AC005506	*** SEQUENCING IN PROG 4.68e+02

	859	15	8.8	129271	24	AC005687	Arabidopsis thaliana B 4.68e+02
c	860	15	8.8	129969	26	HS204E5	Human DNA sequence fro 4.68e+02
c	861	15	8.8	130491	26	HUAC003977	Homo sapiens Chromosom 4.68e+02
	862	15	8.8	130508	17	AC004858	*** SEQUENCING IN PROG 4.68e+02
c	863	15	8.8	130705	25	HS232L22	Human DNA sequence fro 4.68e+02
	864	15	8.8	131458	26	AC002386	Human BAC clone RG343H 4.68e+02
	865	15	8.8	133237	27	AC005550	Homo sapiens PAC clone 4.68e+02
c	866	15	8.8	134289	17	AC004905	*** SEQUENCING IN PROG 4.68e+02
c	867	15	8.8	134474	26	HS33F8	Human DNA sequence fro 4.68e+02
c	868	15	8.8	134525	23	CHOSXX	Rice complete chloropl 4.68e+02
c	869	15	8.8	134539	26	HS24M15	Human DNA sequence fro 4.68e+02
c	870	15	8.8	135007	26	HS423N22	Homo sapiens DNA seque 4.68e+02
c	871	15	8.8	135090	26	HS164L12	Homo sapiens DNA seque 4.68e+02
	872	15	8.8	135855	27	HS135E14	Homo sapiens chromosom 4.68e+02
	873	15	8.8	135886	26	HSXIST2	Human Xq13 3' end of P 4.68e+02
	874	15	8.8	135959	17	AC005383	*** SEQUENCING IN PROG 4.68e+02
c	875	15	8.8	136222	27	AC004099	Homo sapiens chromosom 4.68e+02
c	876	15	8.8	137277	26	AC002431	Human BAC clone RG180F 4.68e+02
c	877	15	8.8	138181	17	AC005171	*** SEQUENCING IN PROG 4.68e+02
	878	15	8.8	138799	26	AC004056	Homo sapiens chromosom 4.68e+02
	879	15	8.8	138849	26	HS95C20	Homo sapiens DNA seque 4.68e+02
c	880	15	8.8	138965	26	AC003998	Human PAC clone DJ0161 4.68e+02
	881	15	8.8	139042	17	AC004600	*** SEQUENCING IN PROG 4.68e+02
c	882	15	8.8	139097	17	AC004591	*** SEQUENCING IN PROG 4.68e+02
	883	15	8.8	139389	26	HS680J	Homo sapiens DNA seque 4.68e+02
	884	15	8.8	139919	26	AC002433	Human BAC clone RG317M 4.68e+02
c	885	15	8.8	140230	17	AC004845	*** SEQUENCING IN PROG 4.68e+02
	886	15	8.8	140896	17	HS163G9	Human DNA sequence *** 4.68e+02
	887	15	8.8	142123	26	AC004690	Homo sapiens PAC clone 4.68e+02
	888	15	8.8	142278	26	AC005177	Homo sapiens chromosom 4.68e+02
c	889	15	8.8	142600	26	AC005371	Homo sapiens chromosom 4.68e+02
	890	15	8.8	142799	18	AC004759	Drosophila melanogaste 4.68e+02
	891	15	8.8	143092	18	CEY17G7B	Caenorhabditis elegans 4.68e+02
c	892	15	8.8	143701	27	HS29C18	Homo sapiens DNA seque 4.68e+02
c	893	15	8.8	143819	26	AC002994	Homo sapiens chromosom 4.68e+02
	894	15	8.8	144545	17	AC005287	*** SEQUENCING IN PROG 4.68e+02
	895	15	8.8	147135	26	HS368A4	Human DNA sequence fro 4.68e+02
c	896	15	8.8	147476	17	AC004157	*** SEQUENCING IN PROG 4.68e+02
c	897	15	8.8	147831	17	AC005399	*** SEQUENCING IN PROG 4.68e+02
c	898	15	8.8	149316	17	AC005008	*** SEQUENCING IN PROG 4.68e+02
	899	15	8.8	149687	27	AC004981	Homo sapiens PAC clone 4.68e+02
c	900	15	8.8	149698	26	HS76C18	Human DNA sequence fro 4.68e+02
	901	15	8.8	149764	26	HSAC002076	Human BAC clone GS345D 4.68e+02
c	902	15	8.8	150140	17	AC005478	*** SEQUENCING IN PROG 4.68e+02
c	903	15	8.8	150620	25	HS203C2	Human DNA sequence fro 4.68e+02
	904	15	8.8	151018	17	PFMAL1P6	Plasmodium falciparum 4.68e+02
	905	15	8.8	151152	26	HS454M7	Homo sapiens DNA seque 4.68e+02
	906	15	8.8	151609	17	HS410I8	Human DNA sequence *** 4.68e+02
c	907	15	8.8	151750	26	AC003666	Homo sapiens Xp22 BAC 4.68e+02
c	908	15	8.8	151770	26	HSAC002066	Human BAC clone RG030H 4.68e+02
	909	15	8.8	151855	17	HS191J18	Human DNA sequence *** 4.68e+02

	910	15	8.8	152077	26	HS130G2	Human DNA sequence fro 4.68e+02
c	911	15	8.8	152184	17	CEY70G10	Caenorhabditis elegans 4.68e+02
	912	15	8.8	152386	17	AC005711	*** SEQUENCING IN PROG 4.68e+02
	913	15	8.8	152392	17	HS1018D12	Human DNA sequence *** 4.68e+02
c	914	15	8.8	152550	17	HSJ181N11	*** SEQUENCING IN PROG 4.68e+02
	915	15	8.8	153019	17	AC004969	*** SEQUENCING IN PROG 4.68e+02
c	916	15	8.8	153773	26	HUAC004020	Homo sapiens Chromosom 4.68e+02
	917	15	8.8	153875	26	AC003682	Homo sapiens chromosom 4.68e+02
	918	15	8.8	155821	18	AC005733	Drosophila melanogaste 4.68e+02
	919	15	8.8	155952	26	AC004534	Homo sapiens PAC clone 4.68e+02
	920	15	8.8	156432	26	HS581F7	Human DNA sequence fro 4.68e+02
	921	15	8.8	156811	27	AC005548	Homo sapiens chromosom 4.68e+02
c	922	15	8.8	157099	26	AC002451	Human BAC clone GS306C 4.68e+02
	923	15	8.8	158898	26	U82696	Homo sapiens chromosom 4.68e+02
	924	15	8.8	160850	17	AC004957	*** SEQUENCING IN PROG 4.68e+02
	925	15	8.8	161242	17	HS366B10	Human DNA sequence *** 4.68e+02
c	926	15	8.8	161652	26	AC004774	Homo sapiens BAC clone 4.68e+02
c	927	15	8.8	162455	17	AC005536	*** SEQUENCING IN PROG 4.68e+02
c	928	15	8.8	166001	17	AC005225	*** SEQUENCING IN PROG 4.68e+02
	929	15	8.8	166900	27	AC004053	Homo sapiens chromosom 4.68e+02
	930	15	8.8	166983	26	AC002541	Human BAC clone RG043K 4.68e+02
c	931	15	8.8	167302	17	HS377H14	Human DNA sequence *** 4.68e+02
c	932	15	8.8	167343	26	HSU96629	Human chromosome 8 BAC 4.68e+02
	933	15	8.8	168644	26	AC004544	Homo sapiens BAC clone 4.68e+02
	934	15	8.8	169045	27	AC005544	Homo sapiens chromosom 4.68e+02
c	935	15	8.8	169998	26	HS279H11	Human DNA sequence fro 4.68e+02
c	936	15	8.8	170000	17	AC005509	*** SEQUENCING IN PROG 4.68e+02
c	937	15	8.8	171681	26	HUAC002299	Homo sapiens Chromosom 4.68e+02
c	938	15	8.8	172151	25	HSAC000114	Human PAC clone DJ527C 4.68e+02
c	939	15	8.8	172657	26	AC004386	Homo Sapiens Chromosom 4.68e+02
c	940	15	8.8	172733	17	AC005082	*** SEQUENCING IN PROG 4.68e+02
	941	15	8.8	173321	17	HS242N11	Human DNA sequence *** 4.68e+02
	942	15	8.8	174045	25	HS107N3	Human DNA sequence fro 4.68e+02
c	943	15	8.8	175080	17	HS28O10	Human DNA sequence *** 4.68e+02
c	944	15	8.8	175488	17	AC004887	*** SEQUENCING IN PROG 4.68e+02
	945	15	8.8	177426	17	HS436K10	Human DNA sequence *** 4.68e+02
	946	15	8.8	177568	27	HS388M5	Human DNA sequence fro 4.68e+02
	947	15	8.8	177901	27	AC005084	Homo sapiens BAC clone 4.68e+02
	948	15	8.8	178615	17	HS742C19	Human DNA sequence *** 4.68e+02
c	949	15	8.8	179074	17	HS191D12	Human DNA sequence *** 4.68e+02
	950	15	8.8	180919	27	AF088219	Homo sapiens CC chemok 4.68e+02
	951	15	8.8	181588	26	AC002487	Human BAC clone RG369K 4.68e+02
	952	15	8.8	187051	17	AC004688	*** SEQUENCING IN PROG 4.68e+02
c	953	15	8.8	188362	27	AC005019	Homo sapiens BAC clone 4.68e+02
c	954	15	8.8	189357	17	HS596H12	Human DNA sequence *** 4.68e+02
	955	15	8.8	189370	15	AF010496	Rhodobacter capsulatus 4.68e+02
	956	15	8.8	190665	17	HS495010	Human DNA sequence *** 4.68e+02
	957	15	8.8	193374	17	AC005139	*** SEQUENCING IN PROG 4.68e+02
	958	15	8.8	193707	17	AC005035	*** SEQUENCING IN PROG 4.68e+02
c	959	15	8.8	197374	17	AC005701	*** SEQUENCING IN PROG 4.68e+02
	960	15	8.8	198285	27	HSEVMHC	H.sapiens DMA, DMB, HL 4.68e+02

	961	15	8.8	198552	23	ATAP22	Arabidopsis thaliana D 4.68e+02
	962	15	8.8	198605	17	AC005003	*** SEQUENCING IN PROG 4.68e+02
	963	15	8.8	199500	17	AC005384	*** SEQUENCING IN PROG 4.68e+02
	964	15	8.8	199665	26	AF064857	Homo sapiens BAC deriv 4.68e+02
c	965	15	8.8	199895	17	AC005748	*** SEQUENCING IN PROG 4.68e+02
c	966	15	8.8	200000	17	AC004661	*** SEQUENCING IN PROG 4.68e+02
c	967	15	8.8	200000	17	AC004664	*** SEQUENCING IN PROG 4.68e+02
	968	15	8.8	200349	27	HSU85195	Homo sapiens BAC129, c 4.68e+02
c	969	15	8.8	203583	17	AC004940	*** SEQUENCING IN PROG 4.68e+02
	970	15	8.8	207139	17	CEY17G7	Caenorhabditis elegans 4.68e+02
	971	15	8.8	208230	15	BSUB0008	Bacillus subtilis comp 4.68e+02
	972	15	8.8	208780	15	BSUB0009	Bacillus subtilis comp 4.68e+02
	973	15	8.8	210000	17	AC004599	*** SEQUENCING IN PROG 4.68e+02
	974	15	8.8	210000	17	AC005699	*** SEQUENCING IN PROG 4.68e+02
	975	15	8.8	211792	17	HS216E10	Human DNA sequence *** 4.68e+02
	976	15	8.8	216750	15	BSUB0007	Bacillus subtilis comp 4.68e+02
	977	15	8.8	217538	17	HS109G6	Human DNA sequence *** 4.68e+02
	978	15	8.8	220000	17	AC004480	*** SEQUENCING IN PROG 4.68e+02
	979	15	8.8	221160	15	BSUB0002	Bacillus subtilis comp 4.68e+02
c	980	15	8.8	224645	17	AC004153	*** SEQUENCING IN PROG 4.68e+02
	981	15	8.8	227730	17	CEY56A3	Caenorhabditis elegans 4.68e+02
	982	15	8.8	227892	17	AC005587	*** SEQUENCING IN PROG 4.68e+02
	983	15	8.8	234542	27	HUAC002041	Human Chromosome 16 BA 4.68e+02
	984	15	8.8	244582	17	CEY40H7	Caenorhabditis elegans 4.68e+02
	985	15	8.8	244649	17	AC005308	*** SEQUENCING IN PROG 4.68e+02
	986	15	8.8	247274	17	HS347H13	Human DNA sequence *** 4.68e+02
	987	15	8.8	250529	26	HUAE000658	Homo sapiens T-cell re 4.68e+02
	988	15	8.8	250611	28	MMAE000663	Mus musculus TCR beta 4.68e+02
	989	15	8.8	250685	17	CEY62E10	Caenorhabditis elegans 4.68e+02
c	990	15	8.8	251124	26	HUAE000660	Homo sapiens T-cell re 4.68e+02
	991	15	8.8	257867	18	AC005557	Drosophila melanogaste 4.68e+02
c	992	15	8.8	281376	17	CEY39C12	Caenorhabditis elegans 4.68e+02
c	993	15	8.8	291821	17	CEY48E1	Caenorhabditis elegans 4.68e+02
	994	15	8.8	300197	17	CEY54G9	Caenorhabditis elegans 4.68e+02
c	995	15	8.8	300197	17	CEY54G9	Caenorhabditis elegans 4.68e+02
	996	15	8.8	305444	17	CEY67H2	Caenorhabditis elegans 4.68e+02
c	997	15	8.8	310213	17	AC004526	*** SEQUENCING IN PROG 4.68e+02
	998	15	8.8	316170	17	CEY37A1	Caenorhabditis elegans 4.68e+02
c	999	15	8.8	319486	26	AF055066	Homo sapiens MHC class 4.68e+02
	1000	15	8.8	328808	17	CEY44A6	Caenorhabditis elegans 4.68e+02

ALIGNMENTS

```
RESULT    1
LOCUS       AF012048      1316 bp    DNA             INV       02-JUL-1998
DEFINITION  Drosophila yakuba retrotransposon Helena reverse transcriptase
            pseudogene, partial sequence.
ACCESSION   AF012048
NID         g3282364
KEYWORDS    .
```



```
SOURCE      Drosophila yakuba.
  ORGANISM  Drosophila yakuba
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Arthropoda;
            Tracheata; Hexapoda; Insecta; Pterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1316)
  AUTHORS   Petrov,D.A., Lozovskaya,E.R. and Hartl,D.L.
  TITLE     High intrinsic rate of DNA loss in Drosophila
  JOURNAL   Nature 384 (6607), 346-349 (1996)
  MEDLINE   97088609
REFERENCE   2  (bases 1 to 1316)
  AUTHORS   Petrov,D.A. and Hartl,D.L.
  TITLE     High rate of DNA loss in the D. melanogaster and D. virilis species
            groups
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1316)
  AUTHORS   Petrov,D.A. and Hartl,D.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUL-1997) OEB, Harvard University, 16 Divinity Ave,
            Cambridge, MA 02138, USA
FEATURES             Location/Qualifiers
     source          1. .1316
                     /organism="Drosophila yakuba"
                     /strain="Tai26"
                     /db_xref="taxon:7245"
                     /clone="H-yakuba349"
     repeat_region   <1. .>1316
                     /rpt_family="non-LTR retrotransposon Helena"
     CDS             <1. .>1316
                     /note="reverse transcriptase"
                     /codon_start=1
                     /pseudo
BASE COUNT      422 a    315 c    274 g    305 t
ORIGIN

  Query Match           11.1%;  Score 19;  DB 18;  Length 1316;
  Best Local Similarity  95.0%;  Pred. No. 1.09e+00;
  Matches    19;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db   1267 AGAGTCATAGCAGCCTTCTT 1286
           ||||| ||||||||||||||||
Cp    117 AGAGTNATAGCAGCCTTCTT 98


RESULT    2
LOCUS       AF012049      1317 bp    DNA             INV       02-JUL-1998
DEFINITION  Drosophila yakuba retrotransposon Helena reverse transcriptase,
            partial cds.
ACCESSION   AF012049
NID         g3282365
KEYWORDS    .
```



```
SOURCE      Drosophila yakuba.
  ORGANISM  Drosophila yakuba
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Arthropoda;
            Tracheata; Hexapoda; Insecta; Pterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1317)
  AUTHORS   Petrov,D.A., Lozovskaya,E.R. and Hartl,D.L.
  TITLE     High intrinsic rate of DNA loss in Drosophila
  JOURNAL   Nature 384 (6607), 346-349 (1996)
  MEDLINE   97088609
REFERENCE   2  (bases 1 to 1317)
  AUTHORS   Petrov,D.A. and Hartl,D.L.
  TITLE     High rate of DNA loss in the D. melanogaster and D. virilis species
            groups
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1317)
  AUTHORS   Petrov,D.A. and Hartl,D.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUL-1997) OEB, Harvard University, 16 Divinity Ave,
            Cambridge, MA 02138, USA
FEATURES             Location/Qualifiers
     source          1. .1317
                     /organism="Drosophila yakuba"
                     /strain="Tai26"
                     /db_xref="taxon:7245"
                     /clone="H-yakuba383"
     repeat_region   <1. .>1317
                     /rpt_family="non-LTR retrotransposon Helena"
     CDS             <1. .>1317
                     /codon_start=1
                     /product="reverse transcriptase"
                     /db_xref="PID:g3282366"
                     /translation="KAPGYDRICHATLKVLPIKAIIYIALIFNAILRVQVPPRQWKMA
                     VIMMIHKPGKPEADPESYRPISLLPSLSKLWERLLVKRINDIVRQGNILPDHQFGFRK
                     GHGTVEQVHRLVKHILQALDDCEYSNAVFIDMQQAFDKVWHVGLLCKIKTMLPAPYFC
                     ILKSYLEGREFKITVRNNYSTVYPMRAGVPQGSVLGPLLYSLYTADIPCPNFDHMEAP
                     YKALMATYADDIAVYYSSGDSREAAKGLQEYINALAAWCKRWNLKINATKTTNLCFTL
                     KTLIKKTPPLQLEGVTLDQPLQATTLGITLDRKLTFGPHLKNAVKKCGRRSQQLRWLM
                     NRRNTLSLRCKRAVYAHCILPIWLYGIQIWGIAAKSNYKRIQVMQNRALRQITNCPWY
                     VRNSTLHKDLNLYTVEEQIGRHTSRYSDRLQSHSSLLARRLIPARPL"
BASE COUNT      422 a    319 c    274 g    302 t
ORIGIN

  Query Match           11.1%;  Score 19;  DB 18;  Length 1317;
  Best Local Similarity  95.0%;  Pred. No. 1.09e+00;
  Matches    19;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db   1268 AGAGTCATAGCAGCCTTCTT 1287
           ||||| ||||||||||||||||
Cp    117 AGAGTNATAGCAGCCTTCTT 98
```



```
RESULT    3
LOCUS       AF012050      1317 bp    DNA             INV       02-JUL-1998
DEFINITION  Drosophila yakuba retrotransposon Helena reverse transcriptase
            pseudogene, partial sequence.
ACCESSION   AF012050
NID         g3282367
KEYWORDS    .
SOURCE      Drosophila yakuba.
  ORGANISM  Drosophila yakuba
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Arthropoda;
            Tracheata; Hexapoda; Insecta; Pterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1317)
  AUTHORS   Petrov,D.A., Lozovskaya,E.R. and Hartl,D.L.
  TITLE     High intrinsic rate of DNA loss in Drosophila
  JOURNAL   Nature 384 (6607), 346-349 (1996)
  MEDLINE   97088609
REFERENCE   2  (bases 1 to 1317)
  AUTHORS   Petrov,D.A. and Hartl,D.L.
  TITLE     High rate of DNA loss in the D. melanogaster and D. virilis species
            groups
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1317)
  AUTHORS   Petrov,D.A. and Hartl,D.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUL-1997) OEB, Harvard University, 16 Divinity Ave,
            Cambridge, MA 02138, USA
FEATURES             Location/Qualifiers
     source          1. .1317
                     /organism="Drosophila yakuba"
                     /strain="Tai26"
                     /db_xref="taxon:7245"
                     /clone="H-yakuba380"
     repeat_region   <1. .>1317
                     /rpt_family="non-LTR retrotransposon Helena"
     CDS             <1. .>1317
                     /note="reverse transcriptase"
                     /codon_start=1
                     /pseudo
BASE COUNT      422 a    318 c    274 g    302 t      1 others
ORIGIN

  Query Match           11.1%;  Score 19;  DB 18;  Length 1317;
  Best Local Similarity  95.0%;  Pred. No. 1.09e+00;
  Matches    19;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db   1268 AGAGTCATAGCAGCCTTCTT 1287
           ||||| ||||||||||||||||
Cp    117 AGAGTNATAGCAGCCTTCTT 98
```

```
RESULT    4
LOCUS       HASMT       27694 bp    DNA   circular  PLN       26-DEC-1996
DEFINITION  Hansenula wingei mitochondrial DNA, complete sequence.
ACCESSION   D31785
NID         g1000971
KEYWORDS    ATPase subunit 6; ATPase subunit 8; ATPase subunit 9; NADH
            dehydrogenase subunit 1; NADH dehydrogenase subunit 2; NADH
            dehydrogenase subunit 3; NADH dehydrogenase subunit 4; NADH
            dehydrogenase subunit 4L; NADH dehydrogenase subunit 5; NADH
            dehydrogenase subunit 6; apocytochrome b; cytochrome c oxidase
            subunit I; cytochrome c oxidase subunit II; cytochrome c oxidase
            subunit III; endonuclease homologue protein; large subunit
            ribosomal RNA; small subunit ribosomal RNA; tRNA-Ala; tRNA-Asn;
            tRNA-Asp; tRNA-Cys; tRNA-Gln; tRNA-Glu; tRNA-Gly; tRNA-His;
            tRNA-Ile; tRNA-Leu2; tRNA-Lys; tRNA-Phe; tRNA-Pro; tRNA-Thr;
            tRNA-Trp; tRNA-Tyr; tRNA-Val; transfer RNA-Arg-1; transfer
            RNA-Arg-2; transfer RNA-Leu1; transfer RNA-Met3; transfer RNA-Ser1;
            trnasfer RNA-Ser2.
SOURCE      Hansenula wingei (strain:21) mitochondrion DNA.
  ORGANISM  Mitochondrion Pichia canadensis
            Eukaryotae; mitochondrial eukaryotes; Fungi; Ascomycota;
            Hemiascomycetes; Saccharomycetales; Saccharomycetaceae; Pichia.
REFERENCE   1  (sites)
  AUTHORS   Yoshida,K.
  TITLE     The structure and organization of mitochondrial genome in Hansenula
            wingei
  JOURNAL   International Botanical Congress 503, 503-504 (1993)
REFERENCE   2  (sites)
  AUTHORS   Sekito,T., Okamoto,K., Kitano,H. and Yoshida,K.
  TITLE     Yeast Hansenula wingei mitochondria genome's complete DNA sequence
            demonstrated unique characteristics
  JOURNAL   Nucleic Acids Symp. Ser. 31, 233-234 (1994)
REFERENCE   3  (bases 1 to 27694)
  AUTHORS   Sekito,T., Okamoto,K., Kitano,H. and Yoshida,K.
  TITLE     The complete mitochondrial DNA sequence of Hansenula wingei reveals
            new characteristics of yeast mitochondria
  JOURNAL   Curr. Genet. 28 (1), 39-53 (1995)
  MEDLINE   96022424
REFERENCE   4  (sites)
  AUTHORS   Sekito,T., Okamoto,K., Kitano,H. and Yoshida,K.
  TITLE     Clone bank of the mitochondrial genome of yeast Hansenula wingei
  JOURNAL   Yeast 11 (13), 1317-1321 (1995)
  MEDLINE   96132035
REFERENCE   5  (sites)
  AUTHORS   Okamoto,K., Sekito,T. and Yoshida,K.
  TITLE     The secondary structure and phylogenetic relationship deduced from
            complete nucleotide sequence of mitochondrial small subunit rRNA in
            yeast Hansenula wingei
  JOURNAL   Genes Genet. Syst. (1996) In press
REFERENCE   6  (bases 1 to 27694)
```



```
  AUTHORS   Yoshida,K.
  JOURNAL   Unpublished (1996)
REFERENCE   7  (bases 1 to 27694)
  AUTHORS   Yoshida,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-1994) to the DDBJ/EMBL/GenBank databases. Kazuo
            Yoshida, Hiroshima University, Department of Biological Science;
            1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739, Japan
            (E-mail:R5155828@JPNHIROA.BINET, Tel:0824-24-7455, Fax:0824-24-0734)
FEATURES             Location/Qualifiers
     source          1. .27694
                     /organism="Pichia canadensis"
                     /mitochondrion
                     /strain="21"
                     /db_xref="taxon:4907"
     exon            1. .502
                     /gene="LSU"
                     /number=2
     tRNA            547. .619
                     /gene="tRNA-T"
                     /note="anticodon=ugu"
                     /evidence=not_experimental
                     /product="transfer RNA-Thr"
     gene            547. .619
                     /gene="tRNA-T"
     tRNA            714. .785
                     /gene="tRNA-E"
                     /note="anticodon=uuc"
                     /evidence=not_experimental
                     /product="transfer RNA-Glu"
     gene            714. .785
                     /gene="tRNA-E"
     tRNA            832. .903
                     /gene="tRNA-C"
                     /note="anticodon=gca"
                     /evidence=not_experimental
                     /product="transfer RNA-Cys"
     gene            832. .903
                     /gene="tRNA-C"
     exon            1241. .1963
                     /gene="COX1"
                     /number=1
                     /evidence=not_experimental
     gene            join(1241. .1963,2963. .3847)
                     /gene="COX1"
     CDS             join(1241. .1963,2963. .3847)
                     /gene="COX1"
                     /codon_start=1
                     /transl_table=3
                     /evidence=not_experimental
                     /product="cytochrome c oxidase subunit 1"
```



```
                     /db_xref="PID:d1007134"
                     /db_xref="PID:g1215789"
                     /translation="MYIQRWLYSTNAKDMAMLYFMFAMFSGVMGSTMSLIIRLELAAP
                     GNQMLHGNHQLFNVLVVGHALLMIFFLVMPGLVGGFGNYMLPLLIGASDMSFARLNNM
                     SFWLLPPALVCLVASTLIESWAGTGWTIYPPLSGMQAHSSPSVDLGIFAIHLTSMSSL
                     LGAINFMATSTNMRTNGMSTSKMPLFVWAIMITAVMLLLSLPVLTAGVTMLLMDRNFN
                     TSFFEVAGGGDPVLYQHLFWFFGHPEVYILIVPGFGIISHIVSTYSKKPVFGEMSWVY
                     AMASIAFLGFLVWSHHMFIVGLDADTRAYFTSSTMVIAVPTGIKMFSWLATLYGGSIR
                     LAVPMTYAIAFLFLFTMGGLTGVALANASLDVAFHDTTYYVVGHFHYVLSMGAIFSLFA
                     GYYWSPQILGLIFNERLAQMQFWLIFVGANVIFMPMHFLGTQGMPRRMPDYPDAYAG
                     WNYVSSMGSVMAIMSLALFIYIMYDQLINGLTNKMDNKSVVYSKAPDFVESNTMFANN
                     SIKSASIEFLLNSPPAMHSFNTPAVQS"
     CDS             <1964. .2923
                     /note="no ATG start codon"
                     /codon_start=1
                     /transl_table=3
                     /evidence=not_experimental
                     /product="intronic ORF"
                     /db_xref="PID:d1007135"
                     /db_xref="PID:g1000973"
                     /translation="TYWWPIWYEIINLYFAMCWKLLIMIKYKFKLITMYMSTLYNNML
                     YKLAMLVKMLIYYNNNQQMTNNMIKMMKRYKSYLVGISETMRKNSDLDKKDIKFNQWL
                     AGLMDGDGYLGVSKAGYTSCEMTVALEDEKALRQMQNKFGGSMKMRSGIKAIRYRLHN
                     KEGMIKLINAINGNIRNTKRLPQLHKVCSMLNLPIIEPINLTINNSWFMGFFDADGTM
                     NYYFKNNYPQLTISVSNKKYMDLKPFMDMFGGNIYYDKSNNGCYKWSMQSKTDMLFFI
                     NSYVKFNSSRTTKFNKLMLCSKYYDLINIKAWLIFNKNFNSKE"
     intron          1964. .2962
                     /cons_splice=(5'site:no, 3'site:no)
                     /number=1
                     /evidence=not_experimental
     exon            2963. .3847
                     /gene="COX1"
                     /number=2
                     /evidence=not_experimental
     gene            4115. .4261
                     /gene="ATP8"
     CDS             4115. .4261
                     /gene="ATP8"
                     /codon_start=1
                     /transl_table=3
                     /evidence=not_experimental
                     /product="ATPase subunit 8"
                     /db_xref="PID:d1007136"
                     /db_xref="PID:g1215790"
                     /translation="MPQLVPFYFLNQLTYGFLLLMMLLVLFSQFLLPRMLRLYMSRLF
                     MSKL"
     gene            4441. .5211
                     /gene="ATP6"
     CDS             4441. .5211
                     /gene="ATP6"
                     /codon_start=1
```



```
                     /transl_table=3
                     /evidence=not_experimental
                     /product="ATPase subunit 6"
                     /db_xref="PID:d1007137"
                     /db_xref="PID:g1215791"
                     /translation="MTNNMVNSPLDQFDIKVFMGFVSPFMDLTNMNMTTFTVYSVFML
                     MVNLGLTLLTDNNGRMMGNRWYVSQEAMYDTINNMVSGQMGGKLGGYYFPLMYTFFMF
                     MFTANLMGMMPYSFAMTSHMVFIMSLSVVIWLGVTMMGLYTHGLTFFALFVPAGCPLA
                     LAPLLVLIELLSYSARAMSLGLRLSANTTSGHLLMVILGGLVFNLMSVSIVTFVLGFM
                     PLAGILAMVALEFAIAMIQSYVFAILASGYIKDGLYLH"
     tRNA            5299. .5369
                     /gene="tRNA-I"
                     /note="anticodon=gau"
                     /evidence=not_experimental
                     /product="transfer RNA-Ile"
     gene            5299. .5369
                     /gene="tRNA-I"
     tRNA            5460. .5546
                     /gene="tRNA-S1"
                     /note="anticodon=uga"
                     /evidence=not_experimental
                     /product="transfer RNA-Ser1"
     gene            5460. .5546
                     /gene="tRNA-S1"
     tRNA            5610. .5680
                     /gene="tRNA-A"
                     /note="anticodon=ugc"
                     /evidence=not_experimental
                     /product="transfer RNA-Ala"
     gene            5610. .5680
                     /gene="tRNA-A"
     tRNA            5803. .5889
                     /gene="tRNA-Y"
                     /note="anticodon=gua"
                     /evidence=not_experimental
                     /product="transfer RNA-Tyr"
     gene            5803. .5889
                     /gene="tRNA-Y"
     tRNA            6034. .6107
                     /gene="tRNA-N"
                     /note="anticodon=guu"
                     /evidence=not_experimental
                     /product="transfer RNA-Asn"
     gene            6034. .6107
                     /gene="tRNA-N"
     gene            6330. .6953
                     /gene="ND6"
     CDS             6330. .6953
                     /gene="ND6"
...
Note: remainder of annotations omitted.
```

```
  Query Match          11.1%;  Score 19;  DB 23;  Length 27694;
  Best Local Similarity 100.0%;  Pred. No. 1.09e+00;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  15401 AAGATAAAATAAATAATAA 15419
          |||||||||||||||||||
Qy    148 AAGATAAAATAAATAATAA 166


RESULT    5
LOCUS       AC003986   188205 bp    DNA             PRI       24-SEP-1998
DEFINITION  Homo sapiens BAC clone RG370M10 from 7p15, complete sequence.
ACCESSION   AC003986
NID         g3645945
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 188205)
  AUTHORS   Strong,C., Graves,T., Duckels,G. and Ozersky,P.
  TITLE     The sequence of Homo sapiens BAC clone RG370M10
  JOURNAL   Unpublished (1998)
REFERENCE   2  (bases 1 to 188205)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JAN-1998) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   3  (bases 1 to 188205)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-SEP-1998) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Sep 24, 1998 this sequence version replaced gi:2769691.
            SUBMITTED BY:  WUGSC
            Genome Sequencing Center
            Department of Genetics
            Washington University
            St. Louis MO 63108, USA
            http://genome.wustl.edu/gsc
            mailto:sapiens@watson.wustl.edu

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded or sequenced with an alternate
            chemistry; an attempt was made to resolve all sequencing problems,
```



```
            such as compressions and repeats; all regions were covered by
            sequence from more than one subclone; and the assembly was
            confirmed by restriction digest.

            MAPPING INFORMATION:
            The sequence of this clone was established as part of a mapping and
            sequencing collaboration between the NHGRI Chromosome 7 Mapping
            Project (Eric D. Green, Director), John D. McPherson in the
            Department of Genetics (Washington University), and the Washington
            University Genome Sequencing Center.  For additional information
            about the map position of this sequence, see
            http://www.nhgri.nih.gov/DIR/GTB/CHR7 , send
            mailto:egreen@nhgri.nih.gov , or see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            Clone RG370M10 is from a release of the human BAC library
            CITB-HS-A.  The library contains cloned DNA from human sperm.  See:
             Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J.
            Kim et al., Genomics 34:213-8 (1996).  The clone is available from
            Research Genetics, Inc. (http://www.resgen.com).
            VECTOR:  pBeloBAC11
            Selection:  chloramphenicol

            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the right is DJ0531J23.  Actual start of
            this clone is at base position 1 of RG370M10; actual end is at
            188205 of RG370M10.

            This clone contains a polymorphic bases with DJ0531J23.
FEATURES            Location/Qualifiers
     source         1. .188205
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="7"
                    /map="7p15"
                    /clone_lib="CITB-HS-A"
                    /clone="RG370M10"
     repeat_region  1. .60
                    /rpt_family="MER2_type"
     repeat_region  156. .190
                    /rpt_family="AT_rich"
     repeat_region  822. .844
                    /rpt_family="AT_rich"
     repeat_region  1847. .2110
                    /rpt_family="Alu"
     repeat_region  2875. .2959
                    /rpt_family="MIR"
     repeat_region  2960. .2984
                    /rpt_family="(TA)n"
     repeat_region  2985. .3078
                    /rpt_family="MIR"
```



```
     repeat_region  3578. .3771
                    /rpt_family="MIR"
     repeat_region  3933. .4359
                    /rpt_family="L1"
     repeat_region  4824. .5299
                    /rpt_family="L1"
     repeat_region  5318. .5780
                    /rpt_family="L1"
     repeat_region  5810. .6097
                    /rpt_family="L1"
     repeat_region  6101. .6761
                    /rpt_family="L1"
     repeat_region  6786. .7639
                    /rpt_family="L1"
     repeat_region  7685. .7861
                    /rpt_family="L1"
     repeat_region  7949. .8534
                    /rpt_family="L1"
     repeat_region  8534. .13327
                    /rpt_family="L1"
     repeat_region  13439. .13949
                    /rpt_family="L1"
     repeat_region  13988. .14084
                    /rpt_family="MIR"
     repeat_region  14549. .15158
                    /rpt_family="MER4-group"
     repeat_region  17267. .17297
                    /rpt_family="(TAAA)n"
     repeat_region  17931. .18044
                    /rpt_family="L1"
     repeat_region  18067. .18335
                    /rpt_family="Alu"
     repeat_region  18366. .18522
                    /rpt_family="L1"
     repeat_region  18726. .20448
                    /rpt_family="L1"
     repeat_region  20449. .20736
                    /rpt_family="Alu"
     repeat_region  20737. .21567
                    /rpt_family="L1"
     repeat_region  21568. .21866
                    /rpt_family="Alu"
     repeat_region  21867. .23834
                    /rpt_family="L1"
     repeat_region  25247. .25290
                    /rpt_family="(CA)n"
     repeat_region  25586. .25764
                    /rpt_family="MIR"
     repeat_region  26838. .27027
                    /rpt_family="MIR"
     repeat_region  27850. .28230
```



```
                    /rpt_family="MaLR"
     repeat_region  31354. .31399
                    /rpt_family="AT_rich"
     repeat_region  31449. .31475
                    /rpt_family="AT_rich"
     repeat_region  32764. .32810
                    /rpt_family="(CAAAAA)n"
     repeat_region  32968. .33093
                    /rpt_family="T2_type"
     repeat_region  33208. .33272
                    /rpt_family="MIR"
     repeat_region  34667. .34738
                    /rpt_family="(CATA)n"
     repeat_region  35116. .35225
                    /rpt_family="MIR"
     repeat_region  35267. .35572
                    /rpt_family="Alu"
     repeat_region  36644. .36751
                    /rpt_family="MER1_type?"
     repeat_region  38067. .38377
                    /rpt_family="Alu"
     repeat_region  43162. .43203
                    /rpt_family="(CA)n"
     repeat_region  49905. .49927
                    /rpt_family="AT_rich"
     misc_feature   50860. .52851
                    /note="CpG_island (%GC=69.7, o/e=0.80, #CpGs=201)"
     gene           complement(51145. .51750)
                    /gene="SCS"
     CDS            complement(51145. .51750)
                    /gene="SCS"
                    /note="match to Q15672 (PID:g2498009); H_RG370M10.1"
                    /codon_start=1
                    /product="TWIST RELATED PROTEIN (H-TWIST)"
                    /db_xref="PID:g3645946"
                    /translation="MQDVSSSPVSPADDSLSNSEEEPDRQQPPSGKRGGRKRRSSRRS
                    AGGGAGPGGAAGGGVGGGDEPGSPAQGKRGKKSAGCGGGGGAGGGGGSSSGGGSPQSY
                    EELQTQRVMANVRERQRTQSLNEAFAALRKIIPTLPSDKLSKIQTLKLAARYIDFLYQ
                    VLQSDELDSKMASCSYVAHERLSYAFSVWRMEGAWSMSASH"
     repeat_region  51461. .51510
                    /rpt_family="(CGG)n"
     repeat_region  51576. .51620
                    /rpt_family="GC_rich"
     repeat_region  51828. .51854
                    /rpt_family="GC_rich"
     misc_feature   53540. .53877
                    /note="match to EST AA255760 (NID:g1891318) zr82e08.r1"
     misc_feature   53561. .53756
                    /note="match to EST AA463599 (NID:g2188483) zw54h07.r1"
     misc_feature   complement(53900. .54234)
                    /note="match to EST AA824330 (NID:g2896400) aj30d10.s1"
```

```
     repeat_region   59233. .59542
                     /rpt_family="Alu"
     repeat_region   60192. .60409
                     /rpt_family="MIR"
     repeat_region   60697. .60774
                     /rpt_family="L2"
     repeat_region   60845. .60913
...
Note: remainder of annotations omitted.

  Query Match           11.1%;  Score 19;  DB 27;  Length 188205;
  Best Local Similarity 100.0%;  Pred. No. 1.09e+00;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db 168808 GTAGAAGGTGGAAGAAGGC 168826
          ||||||||||||||||||| |
Qy     87 GTAGAAGGTGGAAGAAGGC 105


RESULT    6
LOCUS       AF012051     1196 bp    DNA              INV       02-JUL-1998
DEFINITION  Drosophila teissieri retrotransposon Helena reverse transcriptase
            pseudogene, partial sequence.
ACCESSION   AF012051
NID         g3282368
KEYWORDS    .
SOURCE      Drosophila teissieri.
  ORGANISM  Drosophila teissieri
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Arthropoda;
            Tracheata; Hexapoda; Insecta; Pterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1196)
  AUTHORS   Petrov,D.A., Lozovskaya,E.R. and Hartl,D.L.
  TITLE     High intrinsic rate of DNA loss in Drosophila
  JOURNAL   Nature 384 (6607), 346-349 (1996)
  MEDLINE   97088609
REFERENCE   2  (bases 1 to 1196)
  AUTHORS   Petrov,D.A. and Hartl,D.L.
  TITLE     High rate of DNA loss in the D. melanogaster and D. virilis species
            groups
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1196)
  AUTHORS   Petrov,D.A. and Hartl,D.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUL-1997) OEB, Harvard University, 16 Divinity Ave,
            Cambridge, MA 02138, USA
FEATURES             Location/Qualifiers
     source          1. .1196
                     /organism="Drosophila teissieri"
                     /strain="Brazzaville8"
                     /db_xref="taxon:7243"
```



```
                     /clone="H-teissieri44"
     repeat_region   <1. .>1196
                     /rpt_family="non-LTR retrotransposon Helena"
     CDS             <1. .>1196
                     /note="reverse transcriptase"
                     /codon_start=1
                     /pseudo
BASE COUNT      387 a    286 c    242 g    281 t
ORIGIN

  Query Match           10.5%;  Score 18;  DB 18;  Length 1196;
  Best Local Similarity  94.7%;  Pred. No. 5.43e+00;
  Matches   18;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db   1148 GAGTCATAGCAGCCTTCTT 1166
           |||| ||||||||||||||
Cp    116 GAGTNATAGCAGCCTTCTT 98


RESULT    7
LOCUS       AF012047     1316 bp    DNA              INV       02-JUL-1998
DEFINITION  Drosophila teissieri retrotransposon Helena reverse transcriptase
            pseudogene, partial sequence.
ACCESSION   AF012047
NID         g3282363
KEYWORDS    .
SOURCE      Drosophila teissieri.
  ORGANISM  Drosophila teissieri
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Arthropoda;
            Tracheata; Hexapoda; Insecta; Pterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1316)
  AUTHORS   Petrov,D.A., Lozovskaya,E.R. and Hartl,D.L.
  TITLE     High intrinsic rate of DNA loss in Drosophila
  JOURNAL   Nature 384 (6607), 346-349 (1996)
  MEDLINE   97088609
REFERENCE   2  (bases 1 to 1316)
  AUTHORS   Petrov,D.A. and Hartl,D.L.
  TITLE     High rate of DNA loss in the D. melanogaster and D. virilis species
            groups
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1316)
  AUTHORS   Petrov,D.A. and Hartl,D.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUL-1997) OEB, Harvard University, 16 Divinity Ave,
            Cambridge, MA 02138, USA
FEATURES             Location/Qualifiers
     source          1. .1316
                     /organism="Drosophila teissieri"
                     /strain="Brazzaville8"
                     /db_xref="taxon:7243"
```



```
                     /clone="H-teissieri423"
     repeat_region   <1. .>1316
                     /rpt_family="non-LTR retrotransposon Helena"
     CDS             <1. .>1316
                     /note="reverse transcriptase"
                     /codon_start=1
                     /pseudo
BASE COUNT      423 a    314 c    269 g    310 t
ORIGIN

  Query Match           10.5%;  Score 18;  DB 18;  Length 1316;
  Best Local Similarity  94.7%;  Pred. No. 5.43e+00;
  Matches   18;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db   1268 GAGTCATAGCAGCCTTCTT 1286
           |||| ||||||||||||||
Cp    116 GAGTNATAGCAGCCTTCTT 98


RESULT    8
LOCUS       VIBTCPJA     1396 bp    DNA              BCT       11-JUL-1995
DEFINITION  Vibrio cholerae toxin co-regulated pilus (tcpJ) gene, complete cds.
ACCESSION   L25661
NID         g431740
KEYWORDS    tcpJ gene; toxin co-regulated pilus.
SOURCE      Vibrio cholerae (strain Ogawa 395) DNA.
  ORGANISM  Vibrio cholerae
            Eubacteria; Proteobacteria; gamma subdivision; Vibrionaceae;
            Vibrio.
REFERENCE   1  (bases 1 to 1396)
  AUTHORS   Everiss,K.D., Hughes,K.J. and Peterson,K.M.
  TITLE     The accessory colonization factor and toxin-coregulated pilus gene
            clusters are physically linked on the Vibrio cholerae 0395
            chromosome
  JOURNAL   DNA Seq. 5 (1), 51-55 (1994)
  MEDLINE   95201292
FEATURES             Location/Qualifiers
     source          1. .1396
                     /organism="Vibrio cholerae"
                     /strain="Ogawa 395"
                     /db_xref="taxon:666"
     gene            135. .896
                     /gene="tcpJ"
     CDS             135. .896
                     /gene="tcpJ"
                     /codon_start=1
                     /transl_table=11
                     /product="toxin co-regulated pilus"
                     /db_xref="PID:g431741"
                     /translation="MEYVYLILFSIVSLILGSFSNVVIYRLPRKILLKNHFFYDIDSN
                     RSMCPKCGNKISWYDNVPLLSYLLLHGKCRHCDEKISLSYPIVELSFFIIAFPIYWLS
```



```
                     TDWVDSFVLLGLYFILFNLFVIDFKSMLLPNLLTYPIFMLAFIIVVQQNPALTVESSII
                     GGFAAFIISYVSNFIVRLFKRIDVMGGGDIKLYTAIGTLIGVEFVPYLFLLSSIIAFI
                     HWFFARVSCRYCLYIPLGPSIIISFVIVFFSIRLM"
BASE COUNT      400 a    213 c    243 g    540 t
ORIGIN

  Query Match           10.5%;  Score 18;  DB 15;  Length 1396;
  Best Local Similarity 100.0%;  Pred. No. 5.43e+00;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1047 TTGGAAGATAAAATAAAT 1064
           ||||||||||||||||||
Qy    144 TTGGAAGATAAAATAAAT 161


RESULT    9
LOCUS       VIBACFB      2041 bp    DNA              BCT       12-SEP-1994
DEFINITION  Vibrio cholerae accessory colonization factor (acfB) gene, complete
            cds.
ACCESSION   L25660
NID         g431738
KEYWORDS    accessory colonization factor; acfB gene.
SOURCE      Vibrio cholerae (strain Ogawa 395) DNA.
  ORGANISM  Vibrio cholerae
            Eubacteria; Proteobacteria; gamma subdivision; Vibrionaceae;
            Vibrio.
REFERENCE   1  (bases 1 to 2041)
  AUTHORS   Everiss,K.D., Hughes,K.J., Kovach,M.E. and Peterson,K.M.
  TITLE     The Vibrio cholerae acfB colonization determinant encodes an inner
            membrane protein that is related to a family of signal-transducing
            proteins
  JOURNAL   Infect. Immun. 62 (8), 3289-3298 (1994)
  MEDLINE   94314447
REFERENCE   2  (bases 1 to 2041)
  AUTHORS   Kovach,M.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-DEC-1993) Michael E. Kovach, Department of
            Microbiology, Louisiana State University Medical Center, 1501 Kings
            Highway, Shreveport, LA 71130, USA
FEATURES             Location/Qualifiers
     source          1. .2041
                     /organism="Vibrio cholerae"
                     /strain="Ogawa 395"
                     /db_xref="taxon:666"
     gene            64. .1944
                     /gene="acfB"
     CDS             64. .1944
                     /gene="acfB"
                     /codon_start=1
                     /transl_table=11
                     /product="accessory colonization factor"
```

```
                     /db_xref="PID:g431739"
                     /translation="MKFSIKLLMIFTSIIITSSILTYFQMHGTNEYLDRTIKDTISE
                     TLDSLEDKINLSIESKVDVSKSVIDIYENLLEGESPLSKFKDVDVNHLSNVFQLFGYA
                     DERTGEIITNDPNFKVPTGFDPRTRSWYLNAKKLNTFSLSEPYVDLITEKLMVTTSAP
                     IYNKNDLTGVIFFDIPLDDVQELIKSYNPFDAGTIFIVDNSGKIIFGNKNDISGKNLF
                     GDFDSFPLSVSESKTKDKNGVNYDVFIKMSDFGDWNLVSIIDHDKARSDIITLRWNSI
                     FTAVILASVFFAILLFTMRLMLKPLHQLTDAMVNISSGSADLTVRIPNSTDQEFSKII
                     NSFNIFVGNLQSIVSEVKNMSEKINCITTEYQELVEVCNNSVMDQYRELDMLASSMNE
                     MVATSNQIAQITSEASEITSKINGQVNEGVGAVSSVTESVGNLVEKLDKTKSVIQDLN
                     RQTQNIDVILKAINDIADQTNLLALNAAIEAARAGENGRGFAVVADEVRSLAIKTQES
                     TKNIGSIIHILQENSLLSVHVMDESFNIASETMTISADSKQCLDNISQSVIQIVDITN
                     QVATAAYEQSHVSEEINSNSISIKDKADTLSSLGNKISQQAYSQKALIGHQDDLISKF
                     II"
BASE COUNT      684 a    317 c    371 g    669 t
ORIGIN

  Query Match           10.5%;  Score 18;  DB 15;  Length 2041;
  Best Local Similarity 100.0%;  Pred. No. 5.43e+00;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    208 TTGGAAGATAAAATAAAT 225
          ||||||||||||||||||
Qy    144 TTGGAAGATAAAATAAAT 161


RESULT   10
LOCUS       YSCMTCG19    4785 bp    DNA             PLN       26-MAY-1995
DEFINITION  Saccharomyces cerevisiae mitochondrion 21S ribosomal RNA (21S rRNA)
            gene, 21S R2 ribosomal RNA (21S R2 rRNA) gene, and ORF RI.
ACCESSION   L36903 M62622
NID         g559279
KEYWORDS    21S ribosomal RNA.
SEGMENT     19 of 19
SOURCE      Mitochondrion Saccharomyces cerevisiae (organelle Mitochondrion
            Saccharomyces cerevisiae) DNA.
  ORGANISM  Mitochondrion Saccharomyces cerevisiae
            Eukaryotae; mitochondrial eukaryotes; Fungi; Ascomycota;
            Hemiascomycetes; Saccharomycetales; Saccharomycetaceae;
            Saccharomyces.
REFERENCE   1  (bases 1 to 4785)
  AUTHORS   de Zamaroczy,M. and Bernardi,G.
  TITLE     The primary structure of the mitochondrial genome of Saccharomyces
            cerevisiae--a review
  JOURNAL   Gene 47 (2-3), 155-177 (1986)
  MEDLINE   87163488
FEATURES             Location/Qualifiers
     source          1. .4785
                     /organism="Saccharomyces cerevisiae"
                     /mitochondrion
                     /db_xref="taxon:4932"
     rRNA            901. .2687
```



```
                     /gene="21S rRNA"
                     /product="21S ribosomal RNA"
     gene            901. .2687
                     /gene="21S rRNA"
     CDS             3886. .4593
                     /note="ORF RI"
                     /pseudo
                     /codon_start=1
                     /transl_table=3
                     /translation="MKNIKKNQVMNLGPNSKLLKEYKSXLIELNIEQFEAGIGLILGD
                     AYIRSRDEGKTYCMQFEWKNKAYMDHVCLLYDQWVLSPPHKKERVNHLGNLVITWGAQ
                     TFKHQAFNKLANLFIVNNKKTIPNNLVENYLTPMSLAYWFMDDGGKWDYNKNSTNKSI
                     VLNTQSFTPEEVEYLVKGLRNKFQLNCYVKINKNKPIIYIDSMSYLIFYNLIKPYLIP
                     QMMYKLPNTISSETFLK"
BASE COUNT     2066 a    399 c    568 g   1752 t
ORIGIN

  Query Match           10.5%;  Score 18;  DB 23;  Length 4785;
  Best Local Similarity 100.0%;  Pred. No. 5.43e+00;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1836 AATAAATAATAAGTGAAA 1853
          ||||||||||||||||||
Qy    155 AATAAATAATAAGTGAAA 172


RESULT   11
LOCUS       MISC19       5692 bp    DNA             PLN       01-OCT-1996
DEFINITION  Yeast mitochondrial gene of the large ribosomal RNA.
ACCESSION   V00699 J01526 J01527 J01528 V00684 V00688 V00689
NID         g13539
KEYWORDS    ribosomal RNA.
SOURCE      baker's yeast.
  ORGANISM  Mitochondrion Saccharomyces cerevisiae
            Eukaryotae; mitochondrial eukaryotes; Fungi; Ascomycota;
            Hemiascomycetes; Saccharomycetales; Saccharomycetaceae;
            Saccharomyces.
REFERENCE   1  (bases 1 to 5692)
  AUTHORS   Sor,F. and Fukuhara,H.
  TITLE     Complete DNA sequence coding for the large ribosomal RNA of yeast
            mitochondria
  JOURNAL   Nucleic Acids Res. 11 (2), 339-348 (1983)
  MEDLINE   83143306
FEATURES             Location/Qualifiers
     source          1. .5692
                     /organism="Saccharomyces cerevisiae"
                     /mitochondrion
                     /db_xref="taxon:4932"
     precursor_RNA   901. .>5692
                     /note="primary transcript"
     rRNA            901. .3588
```



```
     intron          3589. .4731
                     /note="intervening sequence"
     rRNA            4732. .5316
BASE COUNT     2444 a    473 c    646 g   2129 t
ORIGIN

  Query Match           10.5%;  Score 18;  DB 23;  Length 5692;
  Best Local Similarity 100.0%;  Pred. No. 5.43e+00;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1836 AATAAATAATAAGTGAAA 1853
          ||||||||||||||||||
Qy    155 AATAAATAATAAGTGAAA 172


RESULT   12
LOCUS       YSCMTRG2P    5692 bp    DNA             PLN       30-SEP-1988
DEFINITION  yeast (s.cerevisiae) mitochondrial 21s rrna omega(+) gene & flanks.
ACCESSION   J01527 J01526 J01528 V00684 V00688 V00689 V00699
NID         g343910
KEYWORDS    21S ribosomal RNA; chloramphenicol resistance; drug resistance
            protein; erythromycin resistance protein; ribosomal RNA.
SOURCE      yeast (saccharomyces cerevisiae; omega+ strains i18-8c/r53 (rho-),
            js1-3d, k114-4a, i18-8c (rho- mutant f11)) mitochondrial dna (see
            comment).
  ORGANISM  Mitochondrion Saccharomyces cerevisiae
            Eukaryotae; mitochondrial eukaryotes; Fungi; Ascomycota;
            Hemiascomycetes; Saccharomycetales; Saccharomycetaceae;
            Saccharomyces.
REFERENCE   1  (bases 3241 to 4790)
  AUTHORS   Dujon,B.
  TITLE     sequence of the intron and flanking exons of the mitochondrial 21s
            rrna gene of yeast strains having different alleles at the omega
            and rib-1 loci
  JOURNAL   Cell 20, 185-197 (1980)
  MEDLINE   80222854
REFERENCE   2  (bases 3429 to 3715; 4653 to 4791)
  AUTHORS   Bos,J.L., Osinga,K.A., Van der Horst,G., Hecht,N.B., Tabak,H.F.,
            van Ommen,G.-J.B. and Borst,P.A.
  TITLE     splice point sequence and transcripts of the intervening sequence
            in the mitochondrial 21s ribosomal rna gene of yeast
  JOURNAL   Cell 20, 207-214 (1980)
  MEDLINE   80222857
REFERENCE   3  (bases 689 to 957)
  AUTHORS   Osinga,K.A. and Tabak,H.F.
  TITLE     initiation of transcription of genes for mitochondrial ribosomal
            rna in yeast: comparison of the nucleotide sequence around the
            5'-ends of both genes reveals a homologous stretch of 17
            nucleotides
  JOURNAL   Nucleic Acids Res. 10, 3617-3626 (1982)
  MEDLINE   82274204
```



```
REFERENCE   4  (bases 1 to 5692)
  AUTHORS   Sor,F. and Fukuhara,H.
  TITLE     complete dna sequence coding for the large ribosomal rna of yeast
            mitochondria
  JOURNAL   Nucleic Acids Res. 11, 339-348 (1983)
  MEDLINE   83143306
COMMENT     [1] also sequenced the 21S rRNA gene sequences of omega alleles
            omega(-) and omega(n). [1] describes a long uninterrupted reading
            frame (3886-4593) able to specify a 235 amino acid polypeptide
            within the intron, and discusses the possibly recent insertion of
            introns and a mechanism for splicing after the assembly of the
            large ribosomal subunit. [2] determined the gene splice points by
            comparison to the s.carlsbergensis 21s rrna gene sequence. [4]
            found many stretches of given sequence homologous to E.coli 23S
            rRNA.
            The short (66 bp) exonic insert (mini-insert) found in omega(+)
            strains occurs at 3366-3432; this insert is transcribed into the
            mature rRNA without excision [4].
            r-c-321 refers to a chloramphenicol-resistant mutation. The r-e-514
            (erythromycin resistant) mutant characteristic of the i18-8c strain
            is at 2851 (located by [4], but wild-type allele not given).
               EMBL features not translated to GenBank features:
               key       from    to      description
               TRANSCR   901  >5692      primary transcript Complete source
            information:
            yeast (saccharomyces cerevisiae; omega+ strains i18-8c/r53 (rho-)
            [1], js1-3d [2], k114-4a [3], i18-8c (rho- mutant f11) [4])
            mitochondrial dna.
FEATURES             Location/Qualifiers
     source          1. .5692
                     /organism="Saccharomyces cerevisiae"
                     /mitochondrion
                     /db_xref="taxon:4932"
     rRNA            901. .3588
                     /note="21S rRNA"
     misc_RNA        901. .>5692
                     /note="21S rRNA precursor"
     rRNA            join(901. .3588,4732. .5316)
                     /note="21S rRNA"
     mutation        2851
                     /note="unknown in wild-type; g in r-e-514"
     intron          3589. .4731
                     /note="21S rRNA intron omega (no splice consensus);
                     putative"
     rRNA            4732. .5316
                     /note="21S rRNA"
     mutation        4784
                     /note="a in wild-type; c in r-c-321"
BASE COUNT     2445 a    472 c    645 g   2129 t      1 others
ORIGIN      139 bp upstream from avaii site.
```

```
  Query Match          10.5%;  Score 18;  DB 23;  Length 5692;
  Best Local Similarity 100.0%;  Pred. No. 5.43e+00;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1836 AATAAATAATAAGTGAAA 1853
          ||||||||||||||||||
Qy    155 AATAAATAATAAGTGAAA 172


RESULT  13
LOCUS       AE001074    13065 bp    DNA             BCT       15-DEC-1997
DEFINITION  Archaeoglobus fulgidus section 33 of 172 of the complete genome.
ACCESSION   AE001074 AE000782
NID         g2689397
KEYWORDS    .
SOURCE      Archaeoglobus fulgidus.
  ORGANISM  Archaeoglobus fulgidus
            Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
            Archaeoglobus.
REFERENCE   1  (bases 1 to 13065)
  AUTHORS   Klenk,H.P., Clayton,R.A., Tomb,J., White,O., Nelson,K.E.,
            Ketchum,K.A., Dodson,R.J., Gwinn,M., Hickey,E.K., Peterson,J.D.,
            Richardson,D.L., Kerlavage,A.R., Graham,D.E., Kyrpides,N.C.,
            Fleischmann,R.D., Quackenbush,J., Lee,N.H., Sutton,G.G., Gill,S.,
            Kirkness,E.F., Dougherty,B.A., McKenney,K., Adams,M.D., Loftus,B.,
            Peterson,S., Reich,C.I., McNeil,L.K., Badger,J.H., Glodek,A.,
            Zhou,L., Overbeek,R., Gocayne,J.D., Weidman,J.F., McDonald,L.,
            Utterback,T., Cotton,M.D., Spriggs,T., Artiach,P., Kaine,B.P.,
            Sykes,S.M., Sadow,P.W., D'Andrea,K.P., Bowman,C., Fujii,C.,
            Garland,S.A., Mason,T.M., Olsen,G.J., Fraser,C.M., Smith,H.O.,
            Woese,C.R. and Venter,J.C.
  TITLE     The complete genome sequence of the hyperthermophilic,
            sulphate-reducing archaeon Archaeoglobus fulgidus
  JOURNAL   Nature 390 (6658), 364-370 (1997)
  MEDLINE   98049343
REFERENCE   2  (bases 1 to 13065)
  AUTHORS   Klenk,H.P., Clayton,R.A., Tomb,J., White,O., Nelson,K.E.,
            Ketchum,K.A., Dodson,R.J., Gwinn,M., Hickey,E.K., Peterson,J.D.,
            Richardson,D.L., Kerlavage,A.R., Graham,D.E., Kyrpides,N.C.,
            Fleischmann,R.D., Quackenbush,J., Lee,N.H., Sutton,G.G., Gill,S.,
            Kirkness,E.F., Dougherty,B.A., McKenney,K., Adams,M.D., Loftus,B.,
            Peterson,S., Reich,C.I., McNeil,L.K., Badger,J.H., Glodek,A.,
            Zhou,L., Overbeek,R., Gocayne,J.D., Weidman,J.F., McDonald,L.,
            Utterback,T., Cotton,M.D., Spriggs,T., Artiach,P., Kaine,B.P.,
            Sykes,S.M., Sadow,P.W., D'Andrea,K.P., Bowman,C., Fujii,C.,
            Garland,S.A., Mason,T.M., Olsen,G.J., Fraser,C.M., Smith,H.O.,
            Woese,C.R. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-DEC-1997) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
  REMARK    In order to show the genes in ascending order on the genome, the
```



```
            origin of this version has been moved by TIGR to position 2093570
            of the original version and the opposite strand is shown from the
            original version.
  COMMENT   On Dec 16, 1997 this sequence version replaced gi:2650182.
  FEATURES  Location/Qualifiers
     source 1. .13065
            /organism="Archaeoglobus fulgidus"
            /db_xref="taxon:2234"
     gene   complement(90. .1004)
            /gene="AF0434"
     CDS    complement(90. .1004)
            /gene="AF0434"
            /note="similar to GB:M31799 PID:144733 percent identity:
            40.73; identified by sequence similarity; putative"
            /codon_start=1
            /transl_table=11
            /product="3-hydroxyacyl-CoA dehydrogenase (hbd-3)"
            /db_xref="PID:g2650191"
            /translation="MFEKIGVVGFGLMGTQITQFFAQQGLEVVAIDVSEERLRKGMEA
            IKAGRFGLQRLVEKGKITEEEMNAVLSRISTSTSHSALKDCDLVIEAVFEDVNLKLKV
            LREIDAVTDAVIGSNTSSISITKLSSAVSNPERFLGIHFFNPAQIQKLVELVKGLLSD
            EKLVNGIRDWFLKLGKVPIVVNDSPGFATSRLGLVLGKEALLMVQEGVATPQDIDIGM
            MLGYGYSMGPIETGDLVGLDTRLRIYEAMFEATKDPKWAPPKLLVQLVDAGYLGDPSL
            KGGSFCNYEYPQQERATEVLRRLGVRK"
     gene   complement(1009. .1779)
            /gene="AF0435"
     CDS    complement(1009. .1779)
            /gene="AF0435"
            /note="similar to PID:755067 percent identity: 47.64;
            identified by sequence similarity; putative"
            /codon_start=1
            /transl_table=11
            /product="enoyl-CoA hydratase (fad-1)"
            /db_xref="PID:g2650190"
            /translation="MGEEVKLLDGEIAVATLNRPEKLNALDTKTRMELAEVIEGIEE
            VARVLIITGSGKAFAAGADINELLQRDAIKAFEATKLGTDLPSRIEELEIPVIAAVNG
            YTLGGGCELAMACDIRIASEKAKFGQPEINLAIIPGAGGTQRLPRLVGLGMAKKLVLT
            GEIIDAQTALRIGLVEEVVEHERLMERAKEVAAKIIEKSPLAVKVAKKALNASINMPL
            KEGLRYEASLFALLFSSEDAKEGMRAFLEKRKPEFRGR"
     gene   complement(1789. .2937)
            /gene="AF0436"
     CDS    complement(1789. .2937)
            /gene="AF0436"
            /note="similar to SP:P45867 PID:853760 GB:AL009126 percent
            identity: 44.06; identified by sequence similarity;
            putative"
            /codon_start=1
            /transl_table=11
            /product="acyl-CoA dehydrogenase (acd-2)"
            /db_xref="PID:g2650189"
            /translation="MDFKLTPEQEDIRRAAREFAEKEFTPEVIEECDREEKYPMEIVR
```



```
            KARDLGFSTVKIPEETGGMGLSLFEEALVTEEFYRVSPGIGNACLSSTPGTELLILFG
            SEEQKEKYLRWVTERHGISAMATTEPEAGSDVANVKTKIVKEGDEWVINGTKMFITNG
            TVCDFVLVLGRTYEGEKRHHGLTFAIVESKSKGFKATKIKNKLGLRASDTAEIVLKDV
            KVPEENILGEPGKGFYYLMEFFNHTRPRVAAQAVGIAQGAFELALNYVKQRKQFGQPI
            AAFQHTQFKIAEMATRIQAARLLTWQAAWQCSIGQSRPELSSMAKWFAGETAVTVADW
            CVQLHGGYGYIGEYPAERYYRDAKLMEIYEGTKEVQKQIIARRLIGKL"
     gene   complement(2938. .3333)
            /gene="AF0437"
     CDS    complement(2938. .3333)
            /gene="AF0437"
            /note="similar to GB:L77117 PID:1592182 percent identity:
            32.98; identified by sequence similarity; putative"
            /codon_start=1
            /transl_table=11
            /product="conserved hypothetical protein"
            /db_xref="PID:g2650188"
            /translation="MFIESRELVMKHRIPVAKTERTWKGLEEGKVFKTVCKSCGKEHH
            PPRAECLCGGEVEWVEITGEGVVEAPTTVHTIPAGPEWHREPYTIAIARFGEVRVMGW
            AKEVRVGDVVRLKPGKDETGVWEVYPEVV"
     gene   complement(3338. .4501)
            /gene="AF0438"
     CDS    complement(3338. .4501)
            /gene="AF0438"
            /note="similar to GB:L77117 PID:1592180 percent identity:
            42.41; identified by sequence similarity; putative"
            /codon_start=1
            /transl_table=11
            /product="3-ketoacyl-CoA thiolase (acaB-8)"
            /db_xref="PID:g2650194"
            /translation="MRVGVIGVGCSRPGRRDDAVIQELAYEAVKEALDDAGITQDEVE
            LSVVGSVNTRGYELMPAVPVNEYCGFAGKGPLRVEAACATGAAAVYTAYTSIASGMAD
            VAIAIGVEKMTEVDTATSLAIGGRAGNYLWEFHQYGTTFPAYTAMHATAHMAKYGTTE
            EQMAKVAVKAHRNAARNPKAHFQKEIAVEDVLNSRYIITPLKLYDCSPISDGSAAAIL
            ASEEKIRELGIDDVVWIDSVGYASDTSNMTNRESFVGLKATVLAREMAYRKAGIEEPV
            KQLDFATVHDCFTIAEIMAYEDLGFCRKGEGGKLVESGETELGGKLPVNTFGGLKAKG
            HPLAATGVAMVYEIVKQLRGEAGELQVELRNNRALLHNVGGTGHFAPVMIFGR"
     gene   complement(4581. .5078)
            /gene="AF0439"
     CDS    complement(4581. .5078)
            /gene="AF0439"
            /note="similar to GB:L77117 SP:Q57615 PID:1498922 percent
            identity: 29.80; identified by sequence similarity;
            putative"
            /codon_start=1
            /transl_table=11
            /product="transcriptional regulatory protein, AsnC family"
            /db_xref="PID:g2650187"
            /translation="MRKGQLDDIDMAIIENILRGKTQSEIAEMLGITLRTVQNRVRAL
            EDNGYLIKIKEGYWKANYQKLGLSMLTVTFVDLDIESKTRIDEIIEHPNKLDYVENVL
            ELVGSAYDLCFVVRYRDIEEFNEERRKFMEWFVKKGIKINHLATYISSKTHKDHRRTT
            IRTKT"
```



```
     gene   complement(5161. .5508)
            /gene="AF0440"
     CDS    complement(5161. .5508)
            /gene="AF0440"
            /note="similar to PID:1001734 PID:1001713 percent
            identity: 42.99; identified by sequence similarity;
            putative"
            /codon_start=1
            /transl_table=11
            /product="conserved hypothetical protein"
            /db_xref="PID:g2650193"
            /translation="MEEEMIIGISTSFSAAHSIPGHKKCGKVHGHNFKVVEISGKVK
            ENGMVMDFFDLKRIVNEVVAKFDHTLLNEQIEIPTSENICLRIFSELAEKGLNVRRVR
            VAENEDKWAEIRR"
     gene   complement(5508. .6185)
            /gene="AF0441"
     CDS    complement(5508. .6185)
            /gene="AF0441"
            /note="similar to GB:L77117 PID:1592229 percent identity:
            33.33; identified by sequence similarity; putative"
            /codon_start=1
            /transl_table=11
            /product="conserved hypothetical protein"
            /db_xref="PID:g2650195"
            /translation="MRANLIEIPESIQGEGFYIGVRQLFVRFAGCNLNCYYCDTPKTS
            ENCLDLTANRTLKNPVSAETVQGRIDSSKVHSVCFTGGEPMLQAEFIASLSKTHPFYL
            ESNMTLPEKAKKLKFCDYVAGDLKVREAGLKNYDEVFQKTVKCFKVLRNTRRRKTFCK
            IVLPDKFDADEVLNSAYEIKNYVFGFVLQPVFGSRVEKILKLQKRMIDFADTRVIPQV
            HKYLGVR"
     gene   complement(6166. .6885)
            /gene="AF0442"
     CDS    complement(6166. .6885)
            /gene="AF0442"
            /note="similar to GB:L77117 PID:1591989 percent identity:
            37.16; identified by sequence similarity; putative"
            /codon_start=1
            /transl_table=11
            /product="succinoglycan biosynthesis regulator (exsB)"
            /db_xref="PID:g2650185"
            /translation="MKAVMLLSGGIDSSTLLYYLLDGGYEVHALTFFYGQKHSKEIES
            AEKVAKAAKVRHLKVDISTIHDLISYGALTGEEEVPKAFYSEEVQRRTIVPNRNMILL
            SIAAGYAVKIGAKEVHYAAHLSDYSIYPDCRKEFVKALDTAVYLANIWIPVEVRAPFV
            DMTKADIVRLGLKLGVPYELTWSCYEGGDRPCLSCGTCLERTEAFLANGVKDPLLSDE
            EWKNAVKIYEEMARNEGKSD"
     tRNA   6971. .7053
            /gene="tRNA-Leu-2"
            /product="Leu GAG"
     gene   6971. .7053
            /gene="tRNA-Leu-2"
     gene   9515. .9859
            /gene="AF0443"
```

```
     CDS             9515. .9859
                     /gene="AF0443"
                     /note="hypothetical protein; identified by GeneMark;
                     putative"
                     /codon_start=1
                     /transl_table=11
...
Note: remainder of annotations omitted.

  Query Match           10.5%;  Score 18;  DB 15;  Length 13065;
  Best Local Similarity 100.0%;  Pred. No. 5.43e+00;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   6880 TTTCATGTAGAAGGTGGA 6897
          ||||||||||||||||||
Qy     81 TTTCATGTAGAAGGTGGA 98


RESULT   14
LOCUS       GORAFPA     26833 bp    DNA             PRI       31-JUL-1995
DEFINITION  Gorilla alpha-fetoprotein (AFP) gene, complete cds.
ACCESSION   M38272
NID         g817963
KEYWORDS    alpha-fetoprotein.
SOURCE      gorilla.
  ORGANISM  Gorilla gorilla
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Gorilla.
REFERENCE   1  (bases 1 to 26833)
  AUTHORS   Ryan,S.C., Zielinski,R. and Dugaiczyk,A.
  TITLE     Structure of the gorilla alpha-fetoprotein gene and the divergence
            of primates
  JOURNAL   Genomics 9 (1), 60-72 (1991)
  MEDLINE   91169517
REFERENCE   2  (bases 1 to 26833)
  AUTHORS   Gibbs,P.E., Zielinski,R., Boyd,C. and Dugaiczyk,A.
  TITLE     Structure, polymorphism, and novel repeated DNA elements revealed
            by a complete sequence of the human alpha-fetoprotein gene
  JOURNAL   Biochemistry 26 (5), 1332-1343 (1987)
  MEDLINE   87185438
REFERENCE   3  (bases 1 to 26833)
  AUTHORS   Dugaiczyk,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-SEP-1990) Dugaiczyk A., University of California,
            Biochemistry, Riverside, Caliifornia, USA, 92521
COMMENT     On May 22, 1995 this sequence version replaced gi:790519.
FEATURES             Location/Qualifiers
     source          1. .26833
                     /organism="Gorilla gorilla"
                     /db_xref="taxon:9593"
                     /clone="lambda gafp22; lambda gafp 9"
```



```
                     /clone_lib="lambda charon 40 from Jerry L.Slightom (Upjohn
                     Corp)"
                     /map="q11-13"
     repeat_region   2565. .2656
                     /standard_name="x-1"
                     /citation=[1]
                     /rpt_type=direct
                     /evidence=experimental
                     /rpt_family="x"
                     /rpt_unit=2565. .2656
     repeat_region   2791. .2885
                     /standard_name="x-2"
                     /citation=[1]
                     /rpt_type=direct
                     /evidence=experimental
                     /rpt_family="x"
                     /rpt_unit=2791. .2885
     CAAT_signal     3027. .3031
                     /note="putative"
     TATA_signal     3069. .3075
                     /note="putative"
     mRNA            join(3096. .3224,4034. .4085,5048. .5180,7471. .7682,
                     8853. .8985,9894. .9991,11541. .11670,13640. .13854,
                     15514. .15646,16215. .16312,16794. .16932,18583. .18806,
                     19927. .20059,21402. .21456,21797. .21941)
                     /gene="gafp"
                     /citation=[1]
                     /evidence=experimental
     gene            join(3096. .3224,4034. .4085,5048. .5180,7471. .7682,
                     8853. .8985,9894. .9991,11541. .11670,13640. .13854,
                     15514. .15646,16215. .16312,16794. .16932,18583. .18806,
                     19927. .20059,21402. .21456,21797. .21941)
                     /gene="gafp"
     prim_transcript 3096. .21941
     exon            3096. .3224
                     /gene="gafp"
     5'UTR           3096. .3139
                     /gene="gafp"
     CDS             join(3140. .3224,4034. .4085,5048. .5180,7471. .7682,
                     8853. .8985,9894. .9991,11541. .11670,13640. .13854,
                     15514. .15646,16215. .16312,16794. .16932,18583. .18806,
                     19927. .20059,21402. .21446)
                     /gene="gafp"
                     /citation=[1]
                     /codon_start=1
                     /evidence=experimental
                     /product="alpha-fetoprotein"
                     /db_xref="PID:g177041"
                     /translation="MKWVESIPLIFLLNFTESRTLHRNEYGIASILDSYQCTAEISLA
                     DLATIFFAQFVQEATYKEVSKMVKDALTAIEKPTGDEQSAGCLENQLPAFLEELCHEK
                     EILEKYGLSDCCSQSEEGRHNCFLAHKKPTPASIPLFQVPEPVTSCEAYEEDRETFMN
```



```
                     KPIYEIARRHPFLYAPTILLWAARYDKIIPSCCKAENAVECFQTKAATVTKELRESSL
                     LNQHACAVMKNFGTRTFQAITVTKLSQKFTKVNFTEIQKLVLDVAHVHEHCCRGDVLD
                     CLQDGEKIMSYICSQQDTLSNKITECCKLTTLERGQCIIHAENDEKPEGLSPNLNRFL
                     GDRDFNQFSSGEKNIFLASFVHEYSRRHPQLAVSVILRVAKGYQELLEKCFQTENPLE
                     CQDKGEEELQKYIQESQALAKRSCGLFQKLGEYYLQNAFLVATTKKAPQLTSSELMAI
                     TRKMAATAACCQLSEDKLLACGEGAADIIIGHLCIRHEMTPVNPGVGQCCTSSYANR
                     RPCFSSLVVDETYVPPAFSDDKFIFHKDLCQAQGVALQTMKQEFLINLVKQKPQITEE
                     QLETVIADFSGLLEKCCQGQEQEVCFAEEGQKLISKTRTALGV"
     sig_peptide     3140. .3196
                     /gene="gafp"
     mat_peptide     join(3197. .3224,4034. .4085,5048. .5180,7471. .7682,
                     8853. .8985,9894. .9991,11541. .11670,13640. .13854,
                     15514. .15646,16215. .16312,16794. .16932,18583. .18806,
                     19927. .20059,21402. .21443)
                     /gene="gafp"
                     /citation=[1]
                     /evidence=experimental
                     /product="alpha-fetoprotein"
     intron          3225. .4033
     exon            4034. .4085
                     /gene="gafp"
     intron          4086. .5047
     exon            5048. .5180
                     /gene="gafp"
     intron          5181. .7470
     repeat_region   complement(6912. .7136)
                     /citation=[1]
                     /rpt_type=dispersed
                     /evidence=experimental
                     /rpt_family="kpn"
                     /rpt_unit=6912. .7136
     exon            7471. .7682
                     /gene="gafp"
     intron          7683. .8852
     exon            8853. .8985
                     /gene="gafp"
     intron          8986. .9893
     exon            9894. .9991
                     /gene="gafp"
     intron          9992. .11540
     exon            11541. .11670
                     /gene="gafp"
     intron          11671. .13639
     exon            13640. .13854
                     /gene="gafp"
     intron          13855. .15513
     misc_feature    14418. .14719
                     /standard_name="xba repeat"
                     /note="in the human afp gene, this region is repeated"
                     /citation=[1]
                     /citation=[2]
```



```
                     /evidence=experimental
     exon            15514. .15646
                     /gene="gafp"
     intron          15647. .16214
     exon            16215. .16312
                     /gene="gafp"
     intron          16313. .16793
     exon            16794. .16932
                     /gene="gafp"
     intron          16933. .18582
     exon            18583. .18806
                     /gene="gafp"
     intron          18807. .19926
     exon            19927. .20059
                     /gene="gafp"
     intron          20060. .21401
     exon            21402. .21456
                     /gene="gafp"
     terminator      21444. .21446
                     /gene="gafp"
     3'UTR           join(21447. .21456,21797. .21941)
                     /gene="gafp"
     intron          21457. .21796
     exon            21797. .21941
                     /gene="gafp"
     polyA_site      21941
                     /gene="gafp"
     repeat_region   23075. .23086
                     /citation=[1]
                     /rpt_type=flanking
                     /evidence=experimental
                     /rpt_family="Alu"
                     /rpt_unit=23075. .23086
     repeat_region   complement(23087. .23216)
                     /citation=[1]
                     /rpt_type=dispersed
                     /evidence=experimental
                     /rpt_family="Alu"
                     /rpt_unit=23086. .23216
     repeat_region   23217. .23227
                     /citation=[1]
                     /rpt_type=flanking
                     /evidence=experimental
                     /rpt_family="Alu"
                     /rpt_unit=23217. .23227
     repeat_region   23515. .23533
                     /citation=[1]
                     /rpt_type=flanking
                     /evidence=experimental
                     /rpt_family="Alu"
                     /rpt_unit=23515. .23533
```

```
     repeat_region   complement(23534. .23841)
                     /citation=[1]
                     /rpt_type=dispersed
                     /evidence=experimental
                     /rpt_family="Alu"
                     /rpt_unit=23534. .23841
     repeat_region   23842. .23859
                     /citation=[1]
                     /rpt_type=flanking
                     /evidence=experimental
                     /rpt_family="Alu"
                     /rpt_unit=23842. .23859
     repeat_region   25681. .25693
                     /citation=[1]
                     /rpt_type=flanking
...
Note: remainder of annotations omitted.

  Query Match           10.5%;  Score 18;  DB 25;  Length 26833;
  Best Local Similarity 100.0%;  Pred. No. 5.43e+00;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   6056 TTATTATTTATTTTATCT 6073
            ||||||||||||||||||
Cp    166 TTATTATTTATTTTATCT 149


RESULT   15
LOCUS       PTU21916     26874 bp    DNA             PRI       08-MAR-1996
DEFINITION  Pan troglodytes alpha-fetoprotein precursor (AFP) gene, complete
            cds.
ACCESSION   U21916
NID         g841311
KEYWORDS    alpha-fetoprotein; hnf1; nf1; glucocorticoid; ap-1.
SOURCE      chimpanzee.
  ORGANISM  Pan troglodytes
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 26874)
  AUTHORS   Nishio,H., Gibbs,P.E., Minghetti,P.P., Zielinski,R. and
            Dugaiczyk,A.
  TITLE     The chimpanzee alpha-fetoprotein-encoding gene shows structural
            similarity to that of gorilla but distinct differences from that of
            human
  JOURNAL   Gene 162 (2), 213-220 (1995)
  MEDLINE   96032345
REFERENCE   2  (bases 1 to 26874)
  AUTHORS   Gibbs,P.E., Zielinski,R., Boyd,C. and Dugaiczyk,A.
  TITLE     Structure, polymorphism, and novel repeated DNA elements revealed
            by a complete sequence of the human alpha-fetoprotein gene
  JOURNAL   Biochemistry 26 (5), 1332-1343 (1987)
```



```
  MEDLINE   87185438
REFERENCE   3  (bases 1 to 26874)
  AUTHORS   Ryan,S.C., Zielinski,R. and Dugaiczyk,A.
  TITLE     Structure of the gorilla alpha-fetoprotein gene and the divergence
            of primates
  JOURNAL   Genomics 9 (1), 60-72 (1991)
  MEDLINE   91169517
REFERENCE   4  (bases 1 to 26874)
  AUTHORS   Dugaiczyk,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-FEB-1995) Achilles Dugaiczyk, Biochemistry,
            University of California, Riverside, CA 92521, USA
FEATURES             Location/Qualifiers
     source          1. .26874
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="cAFP lambda31; cAFP lambda30; cAFP lambda28"
                     /clone_lib="lambda charon 30 (J.L.Slightom, The Upjohn
                     Corporation, Kalamazoo, MI);lambda FixII (Stratagene
                     custom made library)"
     repeat_region   2591. .2683
                     /citation=[1]
                     /citation=[2]
                     /citation=[3]
                     /rpt_type=direct
                     /evidence=experimental
                     /rpt_family="Xfda"
     repeat_region   2817. .2912
                     /citation=[1]
                     /citation=[2]
                     /citation=[3]
                     /rpt_type=direct
                     /evidence=experimental
                     /rpt_family="X"
     misc_binding    2955. .2972
                     /note="putative"
                     /citation=[1]
                     /bound_moiety="glucocorticoid"
     protein_bind    2962. .2970
                     /note="putative"
                     /citation=[1]
                     /bound_moiety="ap-1"
     protein_bind    2992. .3004
                     /note="putative"
                     /citation=[1]
                     /bound_moiety="hnf1"
     protein_bind    3001. .3012
                     /note="putative"
                     /citation=[1]
                     /bound_moiety="nf1"
     CAAT_signal     3053. .3057
```



```
     protein_bind    3063. .3075
                     /note="putative"
                     /citation=[1]
                     /bound_moiety="hnf1"
     TATA_signal     3095. .3101
     exon            3122. .3250
                     /gene="AFP"
                     /number=1
     mRNA            join(3122. .3250,4068. .4119,5082. .5214,7507. .7718,
                     8889. .9021,9936. .10033,11582. .11711,13680. .13894,
                     15553. .15685,16262. .16359,16841. .16979,18630. .18853,
                     19974. .20106,21449. .21503,21844. .21988)
                     /gene="AFP"
                     /product="alpha-fetoprotein"
     5'UTR           3122. .3165
                     /gene="AFP"
     prim_transcript 3122. .21988
     gene            join(3122. .3250,4068. .4119,5082. .5214,7507. .7718,
                     8889. .9021,9936. .10033,11582. .11711,13680. .13894,
                     15553. .15685,16262. .16359,16841. .16979,18630. .18853,
                     19974. .20106,21449. .21503,21844. .21988)
                     /gene="AFP"
     sig_peptide     3166. .3223
                     /gene="AFP"
     CDS             join(3166. .3250,4068. .4119,5082. .5214,7507. .7718,
                     8889. .9021,9936. .10033,11582. .11711,13680. .13894,
                     15553. .15685,16262. .16359,16841. .16979,18630. .18853,
                     19974. .20106,21449. .21493)
                     /gene="AFP"
                     /citation=[1]
                     /codon_start=1
                     /evidence=experimental
                     /product="alpha-fetoprotein precursor"
                     /db_xref="PID:g841312"
                     /translation="MKWVESIFLIFLLNFTESRTLHRNEYGIASILDSYQCTAEINLT
                     DLATIFFAQFVQEATYKEVSKMVKDALTAIEKPTGDEQSAGCLENQLPAFLEELCREK
                     EILEKYGHSDCCSQSEEGRHNCFLAHKKPTPASIPFFQVPEPVTSCEAYEEDRETFMN
                     KFIYEIARRHPFLYAPTILLWAARYDKIIPSCCKAENAVECFQTKAATVTKELRESSL
                     LNQHACAVMKNFGTRTFQAITVTKLSQKFTKVNFTEIQKLVLDVAHVHEHCCRGDVLD
                     CLQDGEKIMSYICSQQDTLSNKITECCKLTTLERGQCIIHAENDEKPEGLSPNLNRFL
                     GDRDFNQFSSGEKNIFLASFVHEYSRRHPQLAVSVILRVAKGYQELLEKCFQTENPLE
                     CQDKGEEELQKYIQESQALAKRSCGLFQKLGEYYLQNAFLVATTKKAPQLTSSELMAI
                     TRKMAATAATCCQLSEDKLLACGEGAADIIIGHLCIRHETTPVNPGVGQCCTSSYANR
                     RPCFSSLVVDETYVPPAFSDDKFIFHKDLCQAQGVALQTMKQEFLINLVKQKPQITEE
                     QLEAVIADFSGLLEKCCQGQEQEVCFAEEGQKLISKTRAALGV"
     mat_peptide     join(3224. .3250,4068. .4119,5082. .5214,7507. .7718,
                     8889. .9021,9936. .10033,11582. .11711,13680. .13894,
                     15553. .15685,16262. .16359,16841. .16979,18630. .18853,
                     19974. .20106,21449. .21490)
                     /gene="AFP"
                     /citation=[1]
```



```
                     /evidence=experimental
                     /product="alpha-fetoprotein"
     intron          3251. .4067
                     /number=1
     exon            4068. .4119
                     /gene="AFP"
                     /number=2
     intron          4120. .5081
                     /number=2
     exon            5082. .5214
                     /gene="AFP"
                     /number=3
     intron          5215. .7506
                     /number=3
     repeat_region   6761. .6881
                     /note="degenerate repeat"
                     /citation=[1]
                     /rpt_type=other
                     /rpt_family="(TG)n"
                     /rpt_unit=6831. .6845
     repeat_region   6831. .6845
                     /citation=[1]
                     /rpt_type=direct
                     /rpt_family="(tg)n"
     repeat_region   complement(6947. .7172)
                     /citation=[1]
                     /citation=[2]
                     /citation=[3]
                     /rpt_type=inverted
                     /evidence=experimental
                     /rpt_family="Kpn"
     exon            7507. .7718
                     /gene="AFP"
                     /number=4
     intron          7719. .8888
                     /number=4
     misc_feature    8179. .8189
                     /note="Alu target site"
                     /citation=[1]
                     /citation=[2]
                     /citation=[3]
     exon            8889. .9021
                     /gene="AFP"
                     /number=5
     intron          9022. .9935
                     /number=5
     exon            9936. .10033
                     /gene="AFP"
                     /number=6
     intron          10034. .11581
                     /number=6
```

```
     exon            11582. .11711
                     /gene="AFP"
                     /number=7
     intron          11712. .13679
                     /number=7
     misc_feature    13485. .13493
                     /note="Xba target site"
                     /citation=[1]
                     /citation=[2]
                     /citation=[3]
     exon            13680. .13894
                     /gene="AFP"
                     /number=8
     intron          13895. .15552
                     /number=8
     repeat_region   14456. .14759
                     /citation=[1]
                     /citation=[2]
                     /citation=[3]
                     /rpt_type=direct
                     /evidence=experimental
                     /rpt_family="Xba"
     exon            15553. .15685
                     /gene="AFP"
...
Note: remainder of annotations omitted.

  Query Match           10.5%;  Score 18;  DB 25;  Length 26874;
  Best Local Similarity 100.0%;  Pred. No. 5.43e+00;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   6090 TTATTATTTATTTTATCT 6107
            ||||||||||||||||||
Cp    166 TTATTATTTATTTTATCT 149


RESULT   16
LOCUS       HUMAFP      27553 bp    DNA             PRI       03-JUN-1997
DEFINITION  Homo sapiens alpha-fetoprotein gene, complete cds.
ACCESSION   M16110
NID         g773678
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 27553)
  AUTHORS   Gibbs,P.E., Zielinski,R., Boyd,C. and Dugaiczyk,A.
  TITLE     Structure, polymorphism, and novel repeated DNA elements revealed
            by a complete sequence of the human alpha-fetoprotein gene
```



```
  JOURNAL   Biochemistry 26 (5), 1332-1343 (1987)
  MEDLINE   87185438
REFERENCE   2  (bases 1 to 27553)
  AUTHORS   Dugaiczyk,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-APR-1987) Dugaiczyk A., University of California at
            Riverside, Biochemistry, Riverside, CA 92521 USA
COMMENT     On Apr 18, 1995 this sequence version replaced gi:178235.
FEATURES             Location/Qualifiers
     source          1. .27553
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="charon 4a library of T. Maniatis"
                     /map="4q11-22"
                     /tissue_type="liver"
                     /dev_stage="fetus"
     repeat_unit     2609. .2701
                     /citation=[1]
                     /rpt_type=direct
                     /evidence=experimental
                     /rpt_family="x"
     repeat_unit     2835. .2930
                     /citation=[1]
                     /rpt_type=direct
                     /evidence=experimental
                     /rpt_family="x"
     CAAT_signal     3072. .3076
     TATA_signal     3114. .3119
                     /note="putative"
                     /citation=[1]
                     /evidence=experimental
     prim_transcript 3141. .22629
     mRNA            join(3141. .3269,4082. .4133,5096. .5228,7516. .7727,
                     9214. .9346,10265. .10362,11911. .12040,14316. .14530,
                     16188. .16320,16889. .16986,17469. .17607,19255. .19478,
                     20619. .20751,22090. .22144,22485. .22629)
     5'UTR           3141. .3184
     exon            3141. .3269
     CDS             join(3185. .3269,4082. .4133,5096. .5228,7516. .7727,
                     9214. .9346,10265. .10362,11911. .12040,14316. .14530,
                     16188. .16320,16889. .16986,17469. .17607,19255. .19478,
                     20619. .20751,22090. .22134)
                     /citation=[1]
                     /codon_start=1
                     /evidence=experimental
                     /product="alpha-fetoprotein"
                     /db_xref="PID:g178236"
                     /translation="MKWVESIFLIFLLNFTESRTLHRNEYGIASILDSYQCTAEISLA
                     DLATIFFAQFVQEATYKEVSKMVKDALTAIEKPTGDEQSSGCLENQLPAFLEELCHEK
                     EILEKYGHSDCCSQSEEGRHNCFLAHKKPTPASIPLFQVPEPVTSCEAYEEDRETFMN
                     KFIYEIARRHPFLYAPTILLWAARYDKIIPSCCKAENAVECFQTKAATVTKELRESSL
```



```
                     LNQHACAVMKNFGTRTFQAITVTKLSQKFTKVNFTEIQKLVLDVAHVHEHCCRGDVLD
                     CLQDGEKIMSYICSQQDTLSNKITECCKLTTLERGQCIIHAENDEKPEGLSPNLNRFL
                     GDRDFNQFSSGEKNIFLASFVHEYSRRHPQLAVSVILRVAKGYQELLEKCFQTENPLE
                     CQDKGEEELQKYIQESQALAKRSCGLFQKLGEYYLQNAFLVAYTKKAPQLTSSELMAI
                     TRKMAATAATCCQLSEDKLLACGEGAADIIIGHLCIRHEMTPVNPGVGQCCTSSYANR
                     RPCFSSLVVDETYVPPAFSDDKFIFHKDLCQAQGVALQTMKQEFLINLVKQKPQITEE
                     QLEAVIADFSGLLEKCCQGQEQEVCFAEEGQKLISKTRAALGV"
     sig_peptide     3185. .3241
     mat_peptide     join(3242. .3269,4082. .4133,5096. .5228,7516. .7727,
                     9214. .9346,10265. .10362,11911. .12040,14316. .14530,
                     16188. .16320,16889. .16986,17469. .17607,19255. .19478,
                     20619. .20751,22090. .22131)
                     /note="alpha-fetoprotein"
                     /evidence=experimental
     intron          3270. .4081
     exon            4082. .4133
     intron          4134. .5095
     exon            5096. .5228
     intron          5229. .7515
     repeat_unit     complement(6956. .7181)
                     /citation=[1]
                     /rpt_type=inverted
                     /evidence=experimental
                     /rpt_family="Kpn"
     exon            7516. .7727
     intron          7728. .9213
     repeat_region   8188. .8198
                     /citation=[1]
                     /rpt_type=dispersed
                     /evidence=experimental
                     /rpt_family="Alu"
     repeat_region   8199. .8502
                     /citation=[1]
                     /rpt_type=dispersed
                     /evidence=experimental
                     /rpt_family="Alu"
     allele          8402. .8403
                     /citation=[1]
                     /replace="aaa"
     repeat_region   8503. .8513
                     /citation=[1]
                     /rpt_family="Alu"
                     /rpt_type=dispersed
                     /evidence=experimental
     allele          8691. .8693
                     /citation=[1]
                     /replace="tc"
     allele          8722. .8724
                     /citation=[1]
                     /replace="at"
     allele          8922
```



```
                     /citation=[1]
                     /replace="g"
     exon            9214. .9346
     intron          9347. .10264
     allele          9691. .9695
                     /citation=[1]
                     /replace="ac"
     exon            10265. .10362
     intron          10363. .11910
     allele          10673
                     /citation=[1]
                     /replace="g"
     exon            11911. .12040
     intron          12041. .14315
     allele          12634
                     /citation=[1]
                     /replace="g"
     repeat_region   13818. .14120
                     /citation=[1]
                     /rpt_type=direct
                     /evidence=experimental
                     /rpt_family="xba"
     exon            14316. .14530
     intron          14531. .16187
     repeat_region   15092. .15394
                     /citation=[1]
                     /rpt_type=direct
                     /evidence=experimental
                     /rpt_family="xba"
     exon            16188. .16320
     intron          16321. .16888
     exon            16889. .16986
     intron          16987. .17468
     exon            17469. .17607
     intron          17608. .19254
     exon            19255. .19478
     intron          19479. .20618
     exon            20619. .20751
     intron          20752. .22089
     exon            22090. .22144
     terminator      22132
     3'UTR           join(22132. .22144,22485. .22629)
     intron          22145. .22484
     exon            22485. .22629
     polyA_signal    22629
     repeat_region   complement(23761. .23771)
                     /citation=[1]
                     /rpt_type=dispersed
                     /evidence=experimental
                     /rpt_family="Alu"
     repeat_region   complement(23772. .23901)
```

```
                     /citation=[1]
                     /rpt_type=dispersed
                     /evidence=experimental
                     /rpt_family="Alu"
     repeat_region   complement(23902. .23912)
                     /citation=[1]
                     /rpt_type=dispersed
                     /evidence=experimental
                     /rpt_family="Alu"
     repeat_region   complement(24201. .24219)
                     /citation=[1]
                     /rpt_type=dispersed
                     /evidence=experimental
                     /rpt_family="Alu"
     repeat_region   complement(24220. .24527)
                     /citation=[1]
                     /rpt_type=dispersed
                     /evidence=experimental
                     /rpt_family="Alu"
     repeat_region   complement(24528. .24545)
                     /citation=[1]
                     /rpt_type=dispersed
                     /evidence=experimental
                     /rpt_family="Alu"
     repeat_region   26397. .26409
                     /citation=[1]
                     /rpt_type=dispersed
                     /evidence=experimental
                     /rpt_family="Kpn"
     repeat_region   complement(26409. .26481)
                     /citation=[1]
                     /rpt_type=inverted
                     /evidence=experimental
...
Note: remainder of annotations omitted.

  Query Match           10.5%;  Score 18;  DB 26;  Length 27553;
  Best Local Similarity 100.0%;  Pred. No. 5.43e+00;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   6098 TTATTATTTATTTTATCT 6115
            ||||||||||||||||||
Cp    166 TTATTATTTATTTTATCT 149


RESULT   17
LOCUS       AC005504    71803 bp    DNA             HTG       21-AUG-1998
DEFINITION  *** SEQUENCING IN PROGRESS *** Plasmodium falciparum 3D7 chromosome
            12 PFTAC1383 genomic sequence; HTGS phase 1, 12 unordered pieces.
ACCESSION   AC005504
NID         g3449066
```



```
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      malaria parasite P. falciparum.
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 71803)
  AUTHORS   Hyman,R.W., Qin,F., Fung,E.L., Conway,A.B. and Davis,R.W.
  TITLE     Plasmodium falciparum 3D7 chromosome 12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 71803)
  AUTHORS   Hyman,R.W., Qin,F., Fung,E.L., Conway,A.B. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-AUG-1998) Stanford DNA Sequencing and Technology
            Center, Stanford University, 855 California Avenue, Palo Alto, CA
            94304, USA
COMMENT     ***                                                        ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                        ***
            * This sequence is unfinished. It consists of 12 contigs for
            * which the order is not known; their order in this record is
            * arbitrary. In some cases, the exact lengths of the gaps
            * between the contigs are also unknown; these gaps are  presented
            * as runs of N as a convenience only. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
            *         1    10151: contig of 10151 bp in length
            *                     gap of unknown length
            *     10152    14350: contig of 4199 bp in length
            *                     gap of unknown length
            *     14351    27200: contig of 12850 bp in length
            *                     gap of unknown length
            *     27201    35836: contig of 8636 bp in length
            *                     gap of unknown length
            *     35837    43587: contig of 7751 bp in length
            *                     gap of unknown length
            *     43588    49180: contig of 5593 bp in length
            *                     gap of unknown length
            *     49181    54500: contig of 5320 bp in length
            *                     gap of unknown length
            *     54501    59215: contig of 4715 bp in length
            *                     gap of unknown length
            *     59216    63569: contig of 4354 bp in length
            *                     gap of unknown length
            *     63570    66543: contig of 2974 bp in length
            *                     gap of unknown length
            *     66544    69377: contig of 2834 bp in length
            *                     gap of unknown length
            *     69378    71803: contig of 2426 bp in length.
FEATURES                Location/Qualifiers
     source             1. .71803
                        /organism="Plasmodium falciparum"
                        /db_xref="taxon:5833"
```



```
                     /chromosome="12"
BASE COUNT    29972 a    6817 c    7270 g   27728 t     16 others
ORIGIN

  Query Match           10.5%;  Score 18;  DB 17;  Length 71803;
  Best Local Similarity 100.0%;  Pred. No. 5.43e+00;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  17411 TATTATTTATTTTATCTT 17428
            ||||||||||||||||||
Cp    165 TATTATTTATTTTATCTT 148


RESULT   18
LOCUS       AC004413   114827 bp    DNA             PRI       20-MAR-1998
DEFINITION  Homo sapiens PAC clone DJ0650P09 from 7q21, complete sequence.
ACCESSION   AC004413
NID         g2979593
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 114827)
  AUTHORS   Edwards,J.
  TITLE     The sequence of Homo sapiens PAC clone DJ0650P09
  JOURNAL   Unpublished (1998)
REFERENCE   2  (bases 1 to 114827)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-MAR-1998) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     SUBMITTED BY:  WUGSC
            Genome Sequencing Center
            Department of Genetics
            Washington University
            St. Louis MO 63108, USA
            http://genome.wustl.edu/gsc
            mailto:sapiens@watson.wustl.edu

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded or sequenced with an alternate
            chemistry; an attempt was made to resolve all sequencing problems,
            such as compressions and repeats; all regions were covered by
            sequence from more than one subclone; and the assembly was
            confirmed by restriction digest.
```



```
            MAPPING INFORMATION:
            The sequence of this clone was established as part of a mapping and
            sequencing collaboration between the NHGRI Chromosome 7 Mapping
            Project (Eric D. Green, Director), John D. McPherson in the
            Department of Genetics (Washington University), and the Washington
            University Genome Sequencing Center.  For additional information
            about the map position of this sequence, see
            http://www.nhgri.nih.gov/DIR/GTB/CHR7 , send
            mailto:egreen@nhgri.nih.gov , or see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            This clone was derived from human PAC library RPCI-4, prepared by
            Pieter de Jong and coworkers at Roswell Park Cancer Institute,
            using the method described by Ioannou et al., Nature Genetics
            6:84-9 (1994).  The library is from one male donor.  For further
            details, see http://bacpac.med.buffalo.edu/
            The clone is available from Genome Systems, Inc.
            (http://www.genomesystems.com).
            VECTOR:  pCYPAC2
            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is H_DJ0650P09.  Actual start of
            this clone is at base position 1 of DJ0650P09.

            This clone contains STS sWSS2917 (NID:g1113644).
FEATURES                Location/Qualifiers
     source             1. .114827
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="7"
                        /map="7q21"
                        /clone_lib="RPCI-4"
                        /clone="DJ0650P09"
     repeat_region      1. .218
                        /rpt_family="Alu"
     repeat_region      775. .949
                        /rpt_family="MER1_type"
     repeat_region      5086. .5143
                        /rpt_family="L2"
     repeat_region      5988. .6556
                        /rpt_family="L1"
     repeat_region      6678. .6978
                        /rpt_family="Alu"
     repeat_region      7534. .7717
                        /rpt_family="MER1_type"
     repeat_region      9505. .9542
                        /rpt_family="L2"
     repeat_region      9633. .9840
                        /rpt_family="L2"
     repeat_region      10170. .10455
                        /rpt_family="L2"
```

```
     repeat_region   10495. .10557
                     /rpt_family="L1"
     repeat_region   10536. .10656
                     /rpt_family="L1"
     repeat_region   10691. .10843
                     /rpt_family="Alu"
     repeat_region   11319. .11716
                     /rpt_family="MaLR"
     repeat_region   12065. .12361
                     /rpt_family="L1"
     repeat_region   12377. .12784
                     /rpt_family="L1"
     repeat_region   12815. .13059
                     /rpt_family="L1"
     repeat_region   13060. .14525
                     /rpt_family="MER2_type"
     repeat_region   14528. .17145
                     /rpt_family="L1"
     repeat_region   17146. .17215
                     /rpt_family="MER2_type"
     repeat_region   17216. .17957
                     /rpt_family="L1"
     repeat_region   17960. .18258
                     /rpt_family="Alu"
     repeat_region   18259. .18404
                     /rpt_family="L1"
     repeat_region   18385. .18731
                     /rpt_family="L1"
     repeat_region   18988. .19118
                     /rpt_family="L1"
     repeat_region   19192. .22751
                     /rpt_family="L1"
     repeat_region   22752. .23051
                     /rpt_family="Alu"
     repeat_region   23052. .25707
                     /rpt_family="L1"
     repeat_region   25730. .27914
                     /rpt_family="L1"
     repeat_region   27913. .28906
                     /rpt_family="L1"
     repeat_region   28904. .29073
                     /rpt_family="L1"
     repeat_region   29593. .29764
                     /rpt_family="Alu"
     repeat_region   30338. .30527
                     /rpt_family="Alu"
     repeat_region   30857. .30894
                     /rpt_family="L2"
     repeat_region   31916. .32058
                     /rpt_family="L2"
     repeat_region   32194. .32363
```



```
                     /rpt_family="L2"
     repeat_region   36551. .36739
                     /rpt_family="MIR"
     repeat_region   38119. .38163
                     /rpt_family="(CA)n"
     repeat_region   38429. .38568
                     /rpt_family="MaLR"
     repeat_region   38566. .38909
                     /rpt_family="MaLR"
     repeat_region   39459. .39754
                     /rpt_family="Alu"
     repeat_region   41063. .41230
                     /rpt_family="MER4-group"
     repeat_region   41232. .41691
                     /rpt_family="MER4-group"
     repeat_region   41688. .41810
                     /rpt_family="MER4-group"
     repeat_region   42147. .42364
                     /rpt_family="MER51A"
     repeat_region   45566. .47189
                     /rpt_family="MER52"
     repeat_region   47507. .47789
                     /rpt_family="Alu"
     repeat_region   47895. .48123
                     /rpt_family="L1"
     repeat_region   48500. .48677
                     /rpt_family="L1"
     repeat_region   48809. .48900
                     /rpt_family="L1"
     repeat_region   49829. .50216
                     /rpt_family="L1"
     repeat_region   51403. .51994
                     /rpt_family="L1"
     repeat_region   54360. .54525
                     /rpt_family="MER86"
     repeat_region   56047. .56808
                     /rpt_family="L1"
     repeat_region   57902. .58358
                     /rpt_family="L1"
     repeat_region   58368. .58633
                     /rpt_family="Alu"
     repeat_region   58717. .59065
                     /rpt_family="L1"
     repeat_region   59091. .59171
                     /rpt_family="L1"
     repeat_region   59172. .59484
                     /rpt_family="Alu"
     repeat_region   59498. .59629
                     /rpt_family="L1"
     repeat_region   59714. .59885
                     /rpt_family="MaLR"
```



```
     repeat_region   60037. .60274
                     /rpt_family="MaLR"
     repeat_region   60305. .61206
                     /rpt_family="L1"
     repeat_region   61962. .62181
                     /rpt_family="MIR"
     repeat_region   67700. .67754
                     /rpt_family="Alu"
     repeat_region   67846. .68312
                     /rpt_family="L1"
...
Note: remainder of annotations omitted.

  Query Match           10.5%;  Score 18;  DB 26;  Length 114827;
  Best Local Similarity 100.0%;  Pred. No. 5.43e+00;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  49676 TATTATTTATTTTATCTT 49693
          ||||||||||||||||||
Cp    165 TATTATTTATTTTATCTT 148


RESULT   19
LOCUS       AC002093   133687 bp    DNA             HTG       05-FEB-1998
DEFINITION  *** SEQUENCING IN PROGRESS *** Homo sapiens chromosome 17, clone
            HCIT145P4; HTGS phase 1, 8 unordered pieces.
ACCESSION   AC002093
NID         g2674258
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 133687)
  AUTHORS   Birren,B., Fasman,K., McKernan,K., Munro,C., Nusbaum,C.,
            Richardson,P. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone HCIT145P4
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 133687)
  AUTHORS   Hawkins,T.L., Reeve,M.P., Christoffersen,A., Birren,B.W.,
            Fasman,K.H., Lander,E.S., McKernan,K., Munro,C., Richardson,P.,
            Barna,N., Chang,A., Cooke,P., Daly,M.J., Forrest,C., Fripp,W.J.,
            Gage,D., Geraigery,K., Hagos,B., Jacotot,L., Lane,M., MacKenzie,J.,
            Marquis,N., McDermott,J., Moloney,N., Morrow,J., Nachman,A.,
            Naylor,J., O'Connor,T., Peterson,K., Rollins,G., Spencer,J.,
            Stilwell,J., Stone,C., Strickland,C., Sydney,K., Wilmer,F. and
            Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-1997) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Dec 11, 1997 this sequence version replaced gi:2257734.
```



```
            The Staden databases, finishing information, and all
            chromatographic files used in the assembly of this clone are
            available from our anonymous ftp site.

            All repeats were identified using RepeatMasker: Smit, A.F.A. &
            Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html.
            ***                                                          ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                          ***
            * This sequence is unfinished. It consists of 8 contigs for
            * which the order is not known; their order in this record is
            * arbitrary. In some cases, the exact lengths of the gaps
            * between the contigs are also unknown; these gaps are  presented
            * as runs of N as a convenience only. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
            *        1    43373: contig of 43373 bp in length
            *                    gap of unknown length
            *    43374    53096: contig of 9723 bp in length
            *                    gap of unknown length
            *    53097    64602: contig of 11506 bp in length
            *                    gap of unknown length
            *    64603    82151: contig of 17549 bp in length
            *                    gap of unknown length
            *    82152    94667: contig of 12516 bp in length
            *                    gap of unknown length
            *    94668   103743: contig of 9076 bp in length
            *                    gap of unknown length
            *   103744   107388: contig of 3645 bp in length
            *                    gap of unknown length
            *   107389   133687: contig of 26299 bp in length.
FEATURES              Location/Qualifiers
     source           1. .133687
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="HCIT145P4"
                      /clone_lib="Research Genetics/Cal Tech CITB978SK-B (plates
                      1-194)"
                      /map="17"
                      /chromosome="17"
BASE COUNT    32553 a  36202 c  34525 g  29033 t   1374 others
ORIGIN

  Query Match           10.5%;  Score 18;  DB 17;  Length 133687;
  Best Local Similarity 100.0%;  Pred. No. 5.43e+00;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  60902 GCAGCCTTCTTCCACCTT 60919
          ||||||||||||||||||
Cp    108 GCAGCCTTCTTCCACCTT 91
```

```
RESULT   20
LOCUS       AC004394   170000 bp    DNA             HTG       09-APR-1998
DEFINITION  *** SEQUENCING IN PROGRESS *** Homo sapiens chromosome 4, BAC clone
            C0103F14; HTGS phase 1, 23 unordered pieces.
ACCESSION   AC004394
NID         g3041827
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 170000)
  AUTHORS   Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 170000)
  AUTHORS   Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-MAR-1998) Department of Genetics, Stanford Human
            Genome Center, 855 California Avenue, Palo Alto, CA 94304, USA
COMMENT     On Apr 9, 1998 this sequence version replaced gi:3004541.
            ***                                                        ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                        ***
            * This sequence is unfinished. It consists of 23 contigs for
            * which the order is not known; their order in this record is
            * arbitrary. In some cases, the exact lengths of the gaps
            * between the contigs are also unknown; these gaps are  presented
            * as runs of N as a convenience only. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
            *      1     1507: contig of 1507 bp in length
            *   1508     4460: gap of unknown length
            *   4461     5578: contig of 1118 bp in length
            *   5579     8531: gap of unknown length
            *   8532    10313: contig of 1782 bp in length
            *  10314    13266: gap of unknown length
            *  13267    14749: contig of 1483 bp in length
            *  14750    17702: gap of unknown length
            *  17703    19224: contig of 1522 bp in length
            *  19225    22176: gap of unknown length
            *  22177    23586: contig of 1410 bp in length
            *  23587    26538: gap of unknown length
            *  26539    28412: contig of 1874 bp in length
            *  28413    31364: gap of unknown length
            *  31365    33052: contig of 1688 bp in length
            *  33053    36004: gap of unknown length
            *  36005    37443: contig of 1439 bp in length
            *  37444    40395: gap of unknown length
```



```
            *  40396    42441: contig of 2046 bp in length
            *  42442    45393: gap of unknown length
            *  45394    46604: contig of 1211 bp in length
            *  46605    49556: gap of unknown length
            *  49557    51233: contig of 1677 bp in length
            *  51234    54185: gap of unknown length
            *  54186    57257: contig of 3072 bp in length
            *  57258    60209: gap of unknown length
            *  60210    63159: contig of 2950 bp in length
            *  63160    66111: gap of unknown length
            *  66112    69387: contig of 3276 bp in length
            *  69388    72339: gap of unknown length
            *  72340    76313: contig of 3974 bp in length
            *  76314    79265: gap of unknown length
            *  79266    85820: contig of 6555 bp in length
            *  85821    88772: gap of unknown length
            *  88773    96587: contig of 7815 bp in length
            *  96588    99539: gap of unknown length
            *  99540   110208: contig of 10669 bp in length
            * 110209   113160: gap of unknown length
            * 113161   121620: contig of 8460 bp in length
            * 121621   124572: gap of unknown length
            * 124573   136454: contig of 11882 bp in length
            * 136455   139406: gap of unknown length
            * 139407   149790: contig of 10384 bp in length
            * 149791   152742: gap of unknown length
            * 152743   170000: contig of 17258 bp in length.
FEATURES             Location/Qualifiers
     source          1. .170000
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
BASE COUNT    25409 a  27687 c  27044 g  24897 t  64963 others
ORIGIN

  Query Match           10.5%;  Score 18;  DB 17;  Length 170000;
  Best Local Similarity 100.0%;  Pred. No. 5.43e+00;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db 168543 TCCTATTCCCCTCCTCCT 168560
          ||||||||||||||||||
Qy     33 TCCTATTCCCCTCCTCCT 50


RESULT   21
LOCUS       AC004579   170000 bp    DNA             HTG       17-APR-1998
DEFINITION  *** SEQUENCING IN PROGRESS *** Homo sapiens chromosome 4, BAC clone
            C0212P22; HTGS phase 1, 27 unordered pieces.
ACCESSION   AC004579
NID         g3056997
KEYWORDS    HTG; HTGS_PHASE1.
```



```
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 170000)
  AUTHORS   Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 170000)
  AUTHORS   Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-APR-1998) Department of Genetics, Stanford Human
            Genome Center, 855 California Avenue, Palo Alto, CA 94304, USA
COMMENT     ***                                                        ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                        ***
            * This sequence is unfinished. It consists of 27 contigs for
            * which the order is not known; their order in this record is
            * arbitrary. In some cases, the exact lengths of the gaps
            * between the contigs are also unknown; these gaps are  presented
            * as runs of N as a convenience only. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
            *      1     1891: contig of 1891 bp in length
            *   1892     3562: gap of unknown length
            *   3563     5221: contig of 1659 bp in length
            *   5222     6892: gap of unknown length
            *   6893     8031: contig of 1139 bp in length
            *   8032     9702: gap of unknown length
            *   9703    11259: contig of 1557 bp in length
            *  11260    12930: gap of unknown length
            *  12931    14177: contig of 1247 bp in length
            *  14178    15848: gap of unknown length
            *  15849    17357: contig of 1509 bp in length
            *  17358    19028: gap of unknown length
            *  19029    20303: contig of 1275 bp in length
            *  20304    21974: gap of unknown length
            *  21975    23701: contig of 1727 bp in length
            *  23702    25372: gap of unknown length
            *  25373    27009: contig of 1637 bp in length
            *  27010    28680: gap of unknown length
            *  28681    30188: contig of 1508 bp in length
            *  30189    31858: gap of unknown length
            *  31859    33215: contig of 1357 bp in length
            *  33216    34885: gap of unknown length
            *  34886    36494: contig of 1609 bp in length
            *  36495    38164: gap of unknown length
            *  38165    40053: contig of 1889 bp in length
            *  40054    41723: gap of unknown length
            *  41724    43481: contig of 1758 bp in length
            *  43482    45151: gap of unknown length
```



```
            *  45152    47049: contig of 1898 bp in length
            *  47050    48719: gap of unknown length
            *  48720    52368: contig of 3649 bp in length
            *  52369    54038: gap of unknown length
            *  54039    56950: contig of 2912 bp in length
            *  56951    58620: gap of unknown length
            *  58621    62530: contig of 3910 bp in length
            *  62531    64200: gap of unknown length
            *  64201    68199: contig of 3999 bp in length
            *  68200    69869: gap of unknown length
            *  69870    75999: contig of 6130 bp in length
            *  76000    77669: gap of unknown length
            *  77670    83810: contig of 6141 bp in length
            *  83811    85480: gap of unknown length
            *  85481    92098: contig of 6618 bp in length
            *  92099    93768: gap of unknown length
            *  93769   100377: contig of 6609 bp in length
            * 100378   102047: gap of unknown length
            * 102048   111902: contig of 9855 bp in length
            * 111903   113572: gap of unknown length
            * 113573   128399: contig of 14827 bp in length
            * 128400   130069: gap of unknown length
            * 130070   140142: contig of 10073 bp in length
            * 140143   141812: gap of unknown length
            * 141813   170000: contig of 28188 bp in length.
FEATURES             Location/Qualifiers
     source          1. .170000
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
BASE COUNT    29659 a  34404 c  33001 g  29491 t  43445 others
ORIGIN

  Query Match           10.5%;  Score 18;  DB 17;  Length 170000;
  Best Local Similarity 100.0%;  Pred. No. 5.43e+00;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  78413 AGGAGGAGGGGAATAGGA 78430
          ||||||||||||||||||
Cp     50 AGGAGGAGGGGAATAGGA 33


RESULT   22
LOCUS       AC002324   171952 bp    DNA             ROD       30-MAY-1998
DEFINITION  Mus musculus chromosome 11, clone 475_H_14, complete sequence.
ACCESSION   AC002324
NID         g3165207
KEYWORDS    HTG.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
```

```
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  REFERENCE 1  (bases 1 to 171952)
    AUTHORS   Birren,B., Fasman,K., Linton,L., Nusbaum,C. and Lander,E.
    TITLE     Mus musculus chromosome 11, clone 475_H_14
    JOURNAL   Unpublished
  REFERENCE 2  (bases 1 to 171952)
    AUTHORS   Hawkins,T.L., Reeve,M.P., Christoffersen,A., Birren,B.W.,
              Fasman,K.H., Lander,E.S., McKernan,K., Munro,C., Richardson,P.,
              Barna,N., Chang,A., Cooke,P., Daly,M.J., Devon,K., Dewar,K.,
              Forrest,C., Gage,D., Geraigery,K., Guitau,G., Hagos,B., Huang,J.,
              Jacotot,L., Lane,M., Lee,K., MacKenzie,J., Marquis,N.,
              McDermott,J., Molla,M., Moloney,N., Morrow,J., Nachman,A.,
              Naylor,J., O'Connor,T., Olotu,A., Peterson,K., Rollins,G.,
              Spencer,J., Stilwell,J., Stone,C., Strickland,C., Sydney,K.,
              Traish,A., Wilmer,F., Zemtseva,I. and Zody,M.
    TITLE     Direct Submission
    JOURNAL   Submitted (15-JUL-1997) Whitehead Institute/MIT Center for Genome
              Research, 320 Charles Street, Cambridge, MA 02141, USA
  REFERENCE 3  (bases 1 to 171952)
    AUTHORS   Birren,B., Fasman,K., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
              Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boatin,C.,
              Boutwell,C., Brown,A., Castle,A., Cerny,J., Cooke,P., Depayre,E.,
              Devon,K., Dewar,K., Donelan,L., Etemadi,S., Ferreira,P.,
              FitzHugh,W., Forrest,C., Funke,R., Gage,D., Gardyna,S.,
              Gensheimer,S., Geraigery,K., Gilmartin,T., Grant,G., Hagos,B.,
              Harris,K., Horton,L., Howland,J.C., Hui,L., Jacotot,L., Kann,L.,
              Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
              Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
              Nachman,A., Nahf,R., Naylor,J., O'Connor,T., Pavlin,B.,
              Peterson,K., Riley,R., Roberts,D., Rossello,R., Roy,A., Shyam,R.,
              Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C.,
              Strickland,C., Subramanian,A., Torruella-Miller,I., Vassiliev,H.,
              Vo,A., Wagner,A., Wang,B., Wheeler,J., Wu,Y., Ye,W.J., Zhao,J. and
              Zody,M.
    TITLE     Direct Submission
    JOURNAL   Submitted (30-MAY-1998) Whitehead Institute/MIT Center for Genome
              Research, 320 Charles Street, Cambridge, MA 02141, USA
  COMMENT     On May 30, 1998 this sequence version replaced gi:3142319.
              All repeats were identified using RepeatMasker: Smit, A.F.A. &
              Green, P. (1996-1997)
              http://ftp.genome.washington.edu/RM/RepeatMasker.html.
  FEATURES             Location/Qualifiers
       source          1. .171952
                       /organism="Mus musculus"
                       /db_xref="taxon:10090"
                       /clone="475_H_14"
                       /clone_lib="Research Genetics/Cal Tech CITB-HSP-C (plates
                       195-384)"
                       /map="11"
                       /chromosome="11"
       repeat_region   371. .456
```



```
                       /rpt_family="PB1D10"
       repeat_region   complement(712. .827)
                       /rpt_family="PB1D10"
       repeat_region   complement(990. .1029)
                       /rpt_family="PB1D9"
       repeat_region   complement(1309. .1451)
                       /rpt_family="B1_MM"
       repeat_region   complement(1882. .2071)
                       /rpt_family="URR1B"
       repeat_region   2523. .2563
                       /rpt_family="B3A"
       repeat_region   2596. .2717
                       /rpt_family="B3"
       repeat_region   complement(4356. .4383)
                       /rpt_family="POLY_A"
       repeat_region   5275. .5494
                       /rpt_family="B3"
       repeat_region   5501. .5674
                       /rpt_family="B3"
       repeat_region   6287. .6402
                       /rpt_family="PB1D10"
       repeat_region   complement(7238. .7371)
                       /rpt_family="L2"
       repeat_region   7659. .7737
                       /rpt_family="B1-F"
       repeat_region   complement(8284. .8429)
                       /rpt_family="B1_MM"
       repeat_region   8576. .8652
                       /rpt_family="L2"
       repeat_region   complement(11209. .11411)
                       /rpt_family="B3"
       repeat_region   complement(11436. .11605)
                       /rpt_family="B3"
       repeat_region   complement(11634. .11836)
                       /rpt_family="B2"
       repeat_region   complement(11861. .12013)
                       /rpt_family="B2"
       repeat_region   complement(12262. .12303)
                       /rpt_family="PB1D9"
       repeat_region   12306. .12347
                       /rpt_family="L2"
       repeat_region   complement(12573. .12596)
                       /rpt_family="(CCAA)n"
       repeat_region   complement(12639. .12762)
                       /rpt_family="B4"
       repeat_region   complement(12807. .12872)
                       /rpt_family="PB1D7"
       repeat_region   12920. .13046
                       /rpt_family="B1-F"
       repeat_region   complement(13101. .13294)
                       /rpt_family="B2"
```



```
       repeat_region   13310. .13532
                       /rpt_family="MER2"
       repeat_region   complement(13534. .13650)
                       /rpt_family="B1_MM"
       repeat_region   complement(13597. .13670)
                       /rpt_family="B4A"
       repeat_region   complement(13684. .13829)
                       /rpt_family="B1-F"
       repeat_region   complement(13804. .13837)
                       /rpt_family="MT2B"
       repeat_region   13897. .13922
                       /rpt_family="MT2B"
       unsure          complement(13899. .14050)
                       /note="Single-stranded terminator coverage."
       repeat_region   13923. .14070
                       /rpt_family="B1_MM"
       repeat_region   complement(14207. .14351)
                       /rpt_family="B1_MM"
       repeat_region   complement(14958. .15399)
                       /rpt_family="MTC"
       repeat_region   16021. .16143
                       /rpt_family="PB1D10"
       repeat_region   16153. .16236
                       /rpt_family="PB1D7"
       repeat_region   complement(17105. .17224)
                       /rpt_family="(CA)n"
       repeat_region   complement(17227. .17290)
                       /rpt_family="(CA)n"
       repeat_region   complement(18858. .18900)
                       /rpt_family="AT_rich"
       repeat_region   complement(18901. .18907)
                       /rpt_family="B1-F"
       repeat_region   complement(18908. .18966)
                       /rpt_family="B4"
       repeat_region   complement(18967. .19010)
                       /rpt_family="B1-F"
       repeat_region   complement(19543. .19690)
                       /rpt_family="RSINE1"
       repeat_region   complement(19713. .19815)
                       /rpt_family="PB1D10"
       repeat_region   19950. .20032
                       /rpt_family="PB1D9"
       repeat_region   21057. .21132
                       /rpt_family="PB1D7"
       repeat_region   21151. .21201
                       /rpt_family="ID4"
       repeat_region   21438. .21541
                       /rpt_family="B1-F"
       repeat_region   complement(24457. .24557)
                       /rpt_family="PB1D9"
       repeat_region   25092. .25226
```



```
                       /rpt_family="B1_MM"
       repeat_region   complement(25295. .25403)
                       /rpt_family="PB1D7"
       repeat_region   complement(25668. .25852)
                       /rpt_family="RSINE1"
       repeat_region   complement(25985. .26117)
                       /rpt_family="B1_MM"
       repeat_region   26126. .26176
                       /rpt_family="B3A"
       repeat_region   complement(26292. .26359)
                       /rpt_family="ID4"
       repeat_region   complement(28326. .28816)
                       /rpt_family="L1_MM"
       repeat_region   29393. .29599
                       /rpt_family="B2"
       repeat_region   29603. .29823
                       /rpt_family="B3"
       repeat_region   29996. .30193
                       /rpt_family="B3"
       repeat_region   complement(30412. .30563)
                       /rpt_family="B1_MM"
       repeat_region   complement(31242. .31293)
                       /rpt_family="PB1D7"
       repeat_region   31515. .31647
                       /rpt_family="B1_MM"
       repeat_region   32490. .32579
                       /rpt_family="PB1D10"
       unsure          33376. .33409
                       /note="Single-stranded terminator coverage."
       repeat_region   33580. .33671
                       /rpt_family="GC_rich"
       repeat_region   complement(35248. .35318)
                       /rpt_family="(TA)n"
       repeat_region   35664. .35809
                       /rpt_family="RSINE1"
       repeat_region   36199. .36304
                       /rpt_family="PB1D9"
       repeat_region   complement(36426. .36535)
                       /rpt_family="B2"
...
Note: remainder of annotations omitted.

  Query Match           10.5%;  Score 18;  DB 28;  Length 171952;
  Best Local Similarity 100.0%;  Pred. No. 5.43e+00;
  Matches   18;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  85774 ATAAAATAAATAATAAGT 85791
          ||||||||||||||||||
Qy    151 ATAAAATAAATAATAAGT 168
```

```
RESULT   23
LOCUS       AC004580   190000 bp    DNA             HTG       31-JUL-1998
DEFINITION  *** SEQUENCING IN PROGRESS *** Homo sapiens chromosome 4, BAC clone
            C0289H06; HTGS phase 1, 10 unordered pieces.
ACCESSION   AC004580
NID         g3366572
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 190000)
  AUTHORS   Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 190000)
  AUTHORS   Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-APR-1998) Department of Genetics, Stanford Human
            Genome Center, 855 California Avenue, Palo Alto, CA 94304, USA
COMMENT     On Jul 31, 1998 this sequence version replaced gi:3294513.
            ***                                                          ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                          ***
            * This sequence is unfinished. It consists of 10 contigs for
            * which the order is not known; their order in this record is
            * arbitrary. In some cases, the exact lengths of the gaps
            * between the contigs are also unknown; these gaps are  presented
            * as runs of N as a convenience only. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
            *       1     1249: contig of 1249 bp in length
            *    1250     3162: gap of unknown length
            *    3163     6300: contig of 3138 bp in length
            *    6301     8213: gap of unknown length
            *    8214    11823: contig of 3610 bp in length
            *   11824    13736: gap of unknown length
            *   13737    23882: contig of 10146 bp in length
            *   23883    25795: gap of unknown length
            *   25796    40584: contig of 14789 bp in length
            *   40585    42497: gap of unknown length
            *   42498    55147: contig of 12650 bp in length
            *   55148    57060: gap of unknown length
            *   57061    74037: contig of 16977 bp in length
            *   74038    75950: gap of unknown length
            *   75951   102967: contig of 27017 bp in length
            *  102968   104880: gap of unknown length
            *  104881   143544: contig of 38664 bp in length
            *  143545   145457: gap of unknown length
            *  145458   190000: contig of 44543 bp in length.
FEATURES             Location/Qualifiers
```



```
     source          1. .190000
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
BASE COUNT    40179 a  44650 c  45231 g  42722 t  17218 others
ORIGIN

  Query Match           10.5%;  Score 18;  DB 17;  Length 190000;
  Best Local Similarity 100.0%;  Pred. No. 5.43e+00;
  Matches    18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db 101721 TCCTATTCCCCTCCTCCT 101738
          ||||||||||||||||||
Qy     33 TCCTATTCCCCTCCTCCT 50


RESULT   24
LOCUS       AC004470   207957 bp    DNA             PRI       08-JUL-1998
DEFINITION  Homo sapiens Xp22 BAC GSHB-433024 (Genome Systems Human BAC
            library) complete sequence.
ACCESSION   AC004470
NID         g3294511
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 207957)
  AUTHORS   Muzny,D., Arenson,A.D., Adams,C., Brundage,E., Bunac,C.,
            Carvelli,K., Chacko,J., Chen,J., Di,W., Ding,Y., Dugan,S.,
            Durbin,J., Forcum,J., Ganesh,R., Garcia,C., Goodman,M.,
            Gorrell,J.H., Haywood,M., Hernandez,J., Jackson,L., Jin,S.,
            Kampal,R., Karpathy,S., Kovar,C., Leal,B., Li,Y., Lichtarge,O.,
            Liu,W., Logan,O., Lu,J., Ly,T., Martinez,C., Oswal,G., Perez,L.,
            Rashid,N.D., Rowland,K., Savage,L., Scherer,S.E., Shen,H.,
            Simon,M., Stovall,K., Timms,K.M., Todd,J., Vo,Q., Williamson,A.,
            Worley,K.C., Yu,W., Chinault,C., Nelson,D. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 207957)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-1998) Molecular and Human Genetics, Baylor
            College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 207957)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JUL-1998) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jul 8, 1998 this sequence version replaced gi:3219317.
```



```
            Sequencing is completed to a minimum standard of double strand
            coverage with a minimum of 2 clones and 2 reads with no ambiguities
            or 2 chemistries with a minimum of 2 clones and 3 reads with no
            ambiguities. If the sequence quality does not meet this standard,
            it will be indicated in the annotation.

            The repeat regions shown were identified using RepeatMasker by
            Adrian Smit.

            Sequence similarities were identified using Powerblast by Jinghui
            Zhang.

            Exon/Intron boundaries of identified genes were chosen if there
            were canonical splice junctions that maintained sequence continuity
            across the splice junctions.
FEATURES             Location/Qualifiers
     source          1. .207957
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="BAC GSHB-433024"
                     /chromosome="X"
                     /clone_lib="Genome Systems Human BAC library"
                     /map="Xp22"
     repeat_region   complement(13. .637)
                     /rpt_family="MER7B"
     repeat_region   816. .905
                     /rpt_family="L1MA10"
     repeat_region   complement(1284. .1568)
                     /rpt_family="L1MC/D"
     repeat_region   complement(1875. .2094)
                     /rpt_family="AluJo"
     repeat_region   complement(2128. .2192)
                     /rpt_family="LTR16A"
     repeat_region   complement(2387. .2408)
                     /rpt_family="AT_rich"
     repeat_region   complement(3195. .3512)
                     /rpt_family="L1"
     repeat_region   complement(3517. .3660)
                     /rpt_family="L1"
     repeat_region   complement(3701. .5580)
                     /rpt_family="L1MA"
     repeat_region   complement(5568. .6616)
                     /rpt_family="L1MA"
     repeat_region   complement(8596. .8683)
                     /rpt_family="MER63B"
     repeat_region   complement(9112. .9218)
                     /rpt_family="MSTB"
     repeat_region   9593. .9637
                     /rpt_family="(CAAAA)n"
     repeat_region   10874. .11256
                     /rpt_family="L1ME3A"
```



```
     repeat_region   complement(14179. .14461)
                     /rpt_family="AluSx"
     repeat_region   15013. .15215
                     /rpt_family="L1MA4A"
     repeat_region   15643. .15944
                     /rpt_family="AluSx"
     repeat_region   15945. .15969
                     /rpt_family="(CAAAA)n"
     repeat_region   complement(16603. .16676)
                     /rpt_family="(TAAA)n"
     repeat_region   17273. .17299
                     /rpt_family="AT_rich"
     repeat_region   complement(18248. .18399)
                     /rpt_family="MIR"
     repeat_region   19474. .19532
                     /rpt_family="L1MA5"
     repeat_region   complement(19740. .19815)
                     /rpt_family="MER20"
     repeat_region   complement(19876. .20365)
                     /rpt_family="L2"
     repeat_region   21517. .21886
                     /rpt_family="LTR16C"
     repeat_region   22416. .22529
                     /rpt_family="L2"
     repeat_region   24462. .24762
                     /rpt_family="AluSg"
     repeat_region   25169. .25201
                     /rpt_family="AT_rich"
     repeat_region   complement(25230. .25495)
                     /rpt_family="LTR16C"
     repeat_region   26138. .26169
                     /rpt_family="POLY_A"
     repeat_region   complement(26308. .26716)
                     /rpt_family="MSTA"
     repeat_region   complement(27493. .27516)
                     /rpt_family="(TAAA)n"
     repeat_region   complement(27517. .27816)
                     /rpt_family="AluSx"
     repeat_region   complement(30094. .30125)
                     /rpt_family="(CA)n"
     repeat_region   30776. .30875
                     /rpt_family="MIR"
     repeat_region   complement(31703. .31768)
                     /rpt_family="(CATA)n"
     repeat_region   complement(31892. .32060)
                     /rpt_family="MER5B"
     repeat_region   complement(32351. .32597)
                     /rpt_family="MIR"
     repeat_region   32697. .32765
                     /rpt_family="L2"
     repeat_region   complement(33799. .33856)
```

```
                     /rpt_family="AT_rich"
     repeat_region   34422. .34460
                     /rpt_family="AT_rich"
     repeat_region   34912. .35021
                     /rpt_family="MIR"
     repeat_region   35194. .35256
                     /rpt_family="(CA)n"
     repeat_region   35293. .35749
                     /rpt_family="MLT1C"
     repeat_region   36048. .36592
                     /rpt_family="TIGGER2"
     repeat_region   36594. .36883
                     /rpt_family="AluY"
     repeat_region   36884. .37105
                     /rpt_family="TIGGER2"
     repeat_region   37160. .37408
                     /rpt_family="TIGGER2"
     repeat_region   complement(37657. .38294)
                     /rpt_family="L1M4"
     repeat_region   38710. .38755
                     /rpt_family="L1M4"
     repeat_region   38755. .39465
                     /rpt_family="L1ME"
     repeat_region   complement(39603. .40801)
                     /rpt_family="L1M4"
     repeat_region   40801. .40852
                     /rpt_family="L1"
     repeat_region   41310. .41664
                     /rpt_family="THE1B"
     repeat_region   42024. .42058
                     /rpt_family="AT_rich"
     repeat_region   43394. .43535
                     /rpt_family="MIR"
     repeat_region   complement(43827. .43964)
                     /rpt_family="MIR"
     repeat_region   44283. .44581
                     /rpt_family="L1PA7"
     repeat_region   complement(44665. .44997)
                     /rpt_family="L1ME1"
     misc_feature    complement(44901. .45198)
                     /note="Region: T02862 FB13A8 Fetal brain, Stratagene Homo
                     sapiens cDNA clone FB13A8 3'end"
     repeat_region   complement(45177. .45247)
                     /rpt_family="MER5A"
     repeat_region   45492. .45527
                     /rpt_family="AT_rich"
     repeat_region   45645. .46308
                     /rpt_family="L1PA2"
     repeat_region   46311. .46350
                     /rpt_family="AT_rich"
     repeat_region   46443. .46660
```



```
                     /rpt_family="L1"
     repeat_region   46658. .46731
                     /rpt_family="L1PA4"
     repeat_region   47822. .48274
                     /rpt_family="L1PA15"
     repeat_region   49737. .49813
                     /rpt_family="(TAA)n"
     repeat_region   49856. .49910
                     /rpt_family="(CA)n"
     repeat_region   complement(50429. .50548)
                     /rpt_family="L2"
     repeat_region   51514. .51543
                     /rpt_family="AT_rich"
...
Note: remainder of annotations omitted.

  Query Match           10.5%;  Score 18;  DB 26;  Length 207957;
  Best Local Similarity 100.0%;  Pred. No. 5.43e+00;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db 102052 AGATAAAATAAATAATAA 102069
          ||||||||||||||||||
Qy    149 AGATAAAATAAATAATAA 166


RESULT   25
LOCUS       HUMTCVD1ES     38 bp    mRNA            PRI       10-FEB-1995
DEFINITION  Human (clone: 4coj22) T-cell receptor delta-chain (V-delta-1) mRNA.
ACCESSION   L32494
NID         g497564
KEYWORDS    T-cell receptor; delta chain.
SOURCE      Homo sapiens intestine cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 38)
  AUTHORS   Chowers,Y., Holtmeier,W., Harwood,J., Morzycka-Wroblewska,E. and
            Kagnoff,M.F.
  TITLE     The V delta 1 T cell receptor repertoire in human small intestine
            and colon
  JOURNAL   J. Exp. Med. 180 (1), 183-190 (1994)
  MEDLINE   94275371
FEATURES             Location/Qualifiers
     source          1. .38
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="intestine"
BASE COUNT       11 a       6 c      10 g      11 t
ORIGIN

  Query Match           9.9%;  Score 17;  DB 25;  Length 38;
```



```
  Best Local Similarity 100.0%;  Pred. No. 2.56e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     19 GGAAATTGGAAGATAAA 35
          |||||||||||||||||
Qy    139 GGAAATTGGAAGATAAA 155


RESULT   26
LOCUS       MMCKITI9      569 bp    DNA             ROD       15-FEB-1996
DEFINITION  M.musculus DNA for c-kit gene intron 9 enhancer.
ACCESSION   X80388
NID         g1197339
KEYWORDS    c-kit gene; enhancer.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;
            Murinae; Mus.
REFERENCE   1  (bases 1 to 569)
  AUTHORS   Hayashi,S., Kunisada,T., Ogawa,M. and Nishikawa,S.
  TITLE     Identification of the control regions for mouse c-kit gene
            transcription induced by retinoic acid
  JOURNAL   DNA Res. 2 (5), 211-218 (1995)
  MEDLINE   96366385
REFERENCE   2  (bases 1 to 569)
  AUTHORS   Hayashi,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUL-1995) S. Hayashi, Kyoto University, Faculty of
            Medicine, Department of Molecular Genetics, 53
            Shogoin-Kawahara-Cho, Sakyo-Ku, Kyoto, Kyoto, 606, Japan
FEATURES             Location/Qualifiers
     source          1. .569
                     /organism="Mus musculus"
                     /strain="DBA/2"
                     /db_xref="taxon:10090"
                     /cell_type="F9 teratocarcinoma"
                     /chromosome="5"
                     /clone_lib="lambda EMBL3"
     gene            1. .569
                     /gene="c_kit"
     enhancer        1. .569
                     /gene="c_kit"
                     /note="9th intron SacI_ApaI"
BASE COUNT      123 a    150 c    138 g    158 t
ORIGIN

  Query Match           9.9%;  Score 17;  DB 28;  Length 569;
  Best Local Similarity 100.0%;  Pred. No. 2.56e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```



```
Db    303 TCTTCCAATTTCCTCTT 319
          |||||||||||||||||
Cp    151 TCTTCCAATTTCCTCTT 135


RESULT   27
LOCUS       HIVU43136    1549 bp    DNA             VRL       09-JAN-1996
DEFINITION  Human immunodeficiency virus type 1, CAR isolate 1286CL1, envelope
            glycoprotein gp120 (env) gene, complete truncated cds.
ACCESSION   U43136
NID         g1150893
KEYWORDS    .
SOURCE      Human immunodeficiency virus type 1.
  ORGANISM  Human immunodeficiency virus type 1
            Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate
            lentivirus group.
REFERENCE   1  (bases 1 to 1549)
  AUTHORS   Blouin,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-DEC-1995) John Blouin, HIV Database, National
            Laboratory, Eniwetok, Los Alamos, NM 87545, USA
FEATURES             Location/Qualifiers
     source          1. .1549
                     /organism="Human immunodeficiency virus type 1"
                     /isolate="CAR isolate 1286CL1"
                     /note="HIVCAR286A"
                     /db_xref="taxon:11676"
     gene            1. .732
                     /gene="env"
     CDS             1. .732
                     /gene="env"
                     /codon_start=1
                     /product="envelope glycoprotein gp120"
                     /db_xref="PID:g1150894"
                     /translation="MNVKGTQRNWQNLWRWGLMIFGMLMICNATNNLWVTVYYGVPVW
                     KDRETTLPCASDAKAYSAENIMSGATHACVPTDPNPQEIPLENVTENFNMWRNSMVEQ
                     IDEDIISLWDESLKPCIELTPLCVTLNCTDAKIDCTNGRNNSAGGQTCTGVQNSTAEV
                     GQVEMKNCSFNITTEIRNRKKKEYALFYRLDVVPTNNDDNSTDNYTHNYRLINCNAST
                     IKQACPKVTFEPIPIHYCAPAGLRF"
BASE COUNT      579 a    258 c    336 g    376 t
ORIGIN

  Query Match           9.9%;  Score 17;  DB 33;  Length 1549;
  Best Local Similarity  94.4%;  Pred. No. 2.56e+01;
  Matches   17;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db   1360 CCCATCCAAAGAGTAATA 1377
          ||||||||||||| |||
Cp    126 CCCATCCAAAGAGTNATA 109
```

```
RESULT   28
LOCUS       HSNCFII01     1739 bp     DNA              PRI       21-NOV-1993
DEFINITION  Human neutrophil oxidase factor (NCF2) gene, exons 1 and 2.
ACCESSION   U00776
NID         g392841
KEYWORDS    p67phox.
SEGMENT     1 of 13
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1739)
  AUTHORS   Kenney,R.T., Malech,H.L., Epstein,N.D., Roberts,R.L. and Leto,T.L.
  TITLE     Characterization of the p67phox gene: genomic organization and
            restriction fragment length polymorphism analysis for prenatal
            diagnosis in chronic granulomatous disease
  JOURNAL   Blood 82 (12), 3739-3744 (1993)
  MEDLINE   94083655
REFERENCE   2  (bases 1 to 1739)
  AUTHORS   Kenney,R.T. and Leto,T.L.
  TITLE     A HindIII Polymorphism in the Human NCF2 Gene
  JOURNAL   Nucleic Acids Res 18, 7193 (1990)
  MEDLINE   91088352
REFERENCE   3  (bases 1 to 1739)
  AUTHORS   Leto,T.L., Lomax,K.J., Volpp,B.D., Nunoi,H., Sechler,J.M.,
            Nauseef,W.M., Clark,R.A., Gallin,J.I. and Malech,H.L.
  TITLE     Cloning of a 67-kD Neutrophil Oxidase Factor with Similarity to a
            Noncatalytic Region of p60c-src
  JOURNAL   Science 248, 727-730 (1990)
  MEDLINE   90239568
REFERENCE   4  (bases 1 to 1739)
  AUTHORS   Francke,U., Hsieh,C.L., Foellmer,B.E., Lomax,K.J., Malech,H.L. and
            Leto,T.L.
  TITLE     Genes for Two Autosomal Recessive Forms of Chronic Granulomatous
            Disease Assigned to 1q25 (NCF2) and 7q11.23 (NCF1)
  JOURNAL   Am J Hum Genet 47, 483-492 (1990)
  MEDLINE   90365041
REFERENCE   5  (bases 1 to 1739)
  AUTHORS   Kenney,R.T.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-AUG-1993) Richard T Kenney, Laboratory of Host
            Defenses, NIH/NIAID/LHD, Bldg. 4, Rm 126, 9000 Rockville Pike,
            Bethesda, MD, 20892, USA
FEATURES             Location/Qualifiers
     source          1. .1739
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="placental phage library from Stratagene and
                     Cosmid library of N. Epstein"
                     /chromosome="1"
                     /map="1q25"
```



```
     exon            389. .426
                     /number=1
     exon            905. .1108
                     /number=2
     allele          1293
                     /standard_name="HindIII polymorphism"
                     /replace="T"
     misc_feature    1739
                     /note="This is the first of 13 sequence files for this
                     gene. Insert 5kb virtual sequence between exon 2 and 3."
BASE COUNT      404 a    427 c    447 g    461 t
ORIGIN

  Query Match           9.9%;  Score 17;  DB 26;  Length 1739;
  Best Local Similarity 100.0%;  Pred. No. 2.56e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1630 CTGGCAAGAGGAAATTG 1646
         |||||||||||||||||
Qy    130 CTGGCAAGAGGAAATTG 146


RESULT   29
LOCUS       AF002732     2467 bp     DNA              VRL       01-AUG-1997
DEFINITION  Cydia pomonella granulovirus late expression factor 2 homolog lef-2
            gene, complete cds.
ACCESSION   AF002732
NID         g2232185
KEYWORDS    .
SOURCE      Cydia pomonella granulovirus.
  ORGANISM  Cydia pomonella granulovirus
            Viruses; dsDNA viruses, no RNA stage; Baculoviridae; Granulovirus.
REFERENCE   1  (bases 1 to 2467)
  AUTHORS   Jehle,J.A., van der Linden,I.F.A. and Vlak,J.M.
  TITLE     Identification and sequence analysis of the integration site of
            transposon TCp3.2 in the genome of Cydia pomonella granulovirus
  JOURNAL   Virus Res. (1997) In press
REFERENCE   2  (bases 1 to 2467)
  AUTHORS   Jehle,J.A., van der Linden,I.F.A. and Vlak,J.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAY-1997) Department of Virology, Wageningen
            Agricultural University, P.O.B. 8045, Wageningen 6700 EM,
            Wageningen
FEATURES             Location/Qualifiers
     source          1. .2467
                     /organism="Cydia pomonella granulovirus"
                     /db_xref="taxon:28289"
     CDS             342. .857
                     /codon_start=1
                     /product="late expression factor 2 homolog lef-2"
                     /db_xref="PID:g2232186"
```



```
                     /translation="MSSHYETYHPRRAIDPPKLYKVDMFLRHWVNVVNGNTAFLPGGRY
                     FLVVGRCLLNLVEKSKSFEEHTKVEENAKKVKVCPEDCGGNREAIIAKFRKLFYSRPN
                     AKTEPSFNKLCVRPRRQRYACRLSFNFLVVKRLQCCRCDNKCVYDALKRFYKMDSKCV
                     SQVTDGLVARLT"
     misc_feature    894^895
                     /note="insertion site of the Tcl-like insect transposon
                     TCp3.2 in the genome of the virus mutant CpGV-MCp4"
                     /note="Region: transposon TCp3.2 insertion site"
     CDS             949. .1197
                     /function="unknown"
                     /note="ORF35Ra"
                     /codon_start=1
                     /db_xref="PID:g2232187"
                     /translation="MDQVFSHKDTFIKLAKQLDTKSRAELKPFVSKIYMLIEVYCVTR
                     REKDLLKLLYFLRSCNSEMTRCLVNKCRNTHKIYCITE"
     CDS             1232. .1612
                     /function="unknown"
                     /note="ORF35Rb"
                     /codon_start=1
                     /db_xref="PID:g2232188"
                     /translation="MCTNLFINKHVHVVICKINYPILLPRHQLLHQLFLVHPYFNHLY
                     LMFSRAPIIPNTVYIVDATHQLTRVLSTQTLHLLNVTAHYRIVRRLHFCRRVYSVLDG
                     WWYLRGYSRNGWCGSGSLKYNEIK"
     CDS             complement(1654. .2277)
                     /function="unknown"
                     /note="ORF36L."
                     /codon_start=1
                     /db_xref="PID:g2232189"
                     /translation="MLLIVHSITVSSNTRVAINNTMEHYLVHIVDNIFWTRHSSHKII
                     QDQAILCVEDHATPTLYQTSNFMLLHTRALNFLKDIYKTHKFLKKYLHTTATIVCGDI
                     HLTLPTLAHFLRTHHGMRIESAVAFLEERCSEVVVPLTEMGRLQEVHDFVYEAGAEIL
                     NLDDNDTLNEEDLQPSITEDMIDNFLRSNNDHNYFSISTKFINTYLC"
BASE COUNT      667 a    493 c    564 g    743 t
ORIGIN

  Query Match           9.9%;  Score 17;  DB 33;  Length 2467;
  Best Local Similarity 100.0%;  Pred. No. 2.56e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1602 AGATAAAATAAATAATA 1618
         |||||||||||||||||
Qy    149 AGATAAAATAAATAATA 165


RESULT   30
LOCUS       SCU17018     2574 bp     DNA              PLN       07-FEB-1996
DEFINITION  Saccharomyces cerevisiae Cdc40p (CDC40) gene, complete cds.
ACCESSION   U17018
NID         g575456
KEYWORDS    .
SOURCE      baker's yeast.
```



```
  ORGANISM  Saccharomyces cerevisiae
            Eukaryotae; mitochondrial eukaryotes; Fungi; Ascomycota;
            Hemiascomycetes; Saccharomycetales; Saccharomycetaceae;
            Saccharomyces.
REFERENCE   1  (bases 1 to 2574)
  AUTHORS   Vaisman,N., Tsouladze,A., Robzyk,K., Ben-Yehuda,S., Kupiec,M. and
            Kassir,Y.
  TITLE     The role of Saccharomyces cerevisiae Cdc40p in DNA replication and
            mitotic spindle formation and/or maintenance
  JOURNAL   Mol. Gen. Genet. 247 (2), 123-136 (1995)
  MEDLINE   95272521
REFERENCE   2  (bases 1 to 2574)
  AUTHORS   Kassir,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-NOV-1994) Yona Kassir, Biology, Technion - Israel
            Institute of Technology, Technion City, Haifa, 32000, Israel
FEATURES             Location/Qualifiers
     source          1. .2574
                     /organism="Saccharomyces cerevisiae"
                     /db_xref="taxon:4932"
                     /map="4R"
                     /chromosome="IV"
     gene            477. .1844
                     /gene="CDC40"
                     /db_xref="SGD:L0000275"
     CDS             477. .1844
                     /gene="CDC40"
                     /codon_start=1
                     /db_xref="SGD:L0000275"
                     /evidence=experimental
                     /product="Cdc40p"
                     /db_xref="PID:g575457"
                     /translation="MGLVDGYDTSSDSDLNFDEGKSVHEKKNGNLHEDTSYEPSSNNI
                     HKRKSHFTKSELKRRRKTRKGDGPWGSWSSSDDETSQASETQKEDQDIFVHALAEDNL
                     DSEQIEVEEVSHFYGKSEKDYQGRGYLYPPNDVDVDLREERISFRCYLPKKVIRNYPG
                     HPEGTTALKFLPKTGHLILSGGNDHTIKIWDFTHDYECLRDFQGHNKPIKALRFTEDC
                     QSFLSSSFDRSVKIWDTETGKVKTRLHLNSTPADVESRPTNPHEFIVGLSNSKILHTD
                     DRVSENQGLVQTYDHHLSSILALKYFPDGSKFISSSEDKTVRIWENQINVPIKQISDT
                     AQHSMPFLNVHPSQNYFCAQSMDNRIYSFSLKPKYKRRPKIFKGHSSAGYGISLTPS
                     GDGRYICSGDSKSRLFTWDWNTSRLLNNIKIPGNKPITQVDWHPQETSKVICSGAAGK
                     IYVCD"
BASE COUNT      854 a    430 c    484 g    806 t
ORIGIN

  Query Match           9.9%;  Score 17;  DB 23;  Length 2574;
  Best Local Similarity 100.0%;  Pred. No. 2.56e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2174 TTATTATTTATTTTATC 2190
         |||||||||||||||||
Cp    166 TTATTATTTATTTTATC 150
```

```
RESULT   31
LOCUS       ECOLEP       3131 bp    DNA              BCT       15-SEP-1989
DEFINITION  E.coli lep operon: promoter, lepA and lep genes, complete cds,
            coding for LepA protein and signal peptidase I.
ACCESSION   K00426 J03295
NID         g146598
KEYWORDS    leader peptidase; lep gene; lep operon; peptidase; signal peptidase
            I.
SOURCE      E.coli K12 (strain HJM114) DNA, clone pTD101 [1],[2].
  ORGANISM  Escherichia coli
            Eubacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1835 to 3113)
  AUTHORS   Wolfe,P.B., Wickner,W. and Goodman,J.M.
  TITLE     Sequence of the leader peptidase gene of Escherichia coli and the
            orientation of leader peptidase in the bacterial envelope
  JOURNAL   J. Biol. Chem. 258, 12073-12080 (1983)
  MEDLINE   84008229
REFERENCE   2  (bases 1 to 2089; 3064 to 3131)
  AUTHORS   March,P.E. and Inouye,M.
  TITLE     Characterization of the lep operon of Escherichia coli:
            Identification of the promoter and the gene upstream of the signal
            peptidase I gene
  JOURNAL   J. Biol. Chem. 260, 7206-7213 (1985)
  MEDLINE   85207751
COMMENT     Positions 1 to 2220 and 3101 to 3131 are from [2]; positions 2221
            to 3100 are from [1].
            The data in [2] were kindly submitted on diskette by P.E.March,
            (02-SEP-1985).
            A promoter region containing a -35 region (positions 185-190) and a
            -10 region (positions 206-212) was identified for the lepA gene.  A
            potential ribosome binding site for the lep gene was identified at
            2092-2096.
FEATURES             Location/Qualifiers
     source          1. .3131
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
     mRNA            217. .>2086
                     /note="lepA mRNA (5' end +/- 1 bp) [2]"
     gene            290. .2086
                     /gene="lepA"
     CDS             290. .2086
                     /gene="lepA"
                     /note="LepA protein (lepA)"
                     /codon_start=1
                     /transl_table=11
                     /db_xref="PID:g146599"
                     /translation="MKNIRNFSIIAHIDHGKSTLSDRIIQICGGLSDREMEAQVLDSM
                     DLERERGITIKAQSVTLDYKASDGETYQLNFIDTPGHVDFSYEVSRSLAACEGALLVV
```



```
                     DAGQGVEAQTLANCYTAMEMDLEVVPVLNKIDLPAADPERVAEEIEDIVGIDAHRRGA
                     LFSENRRWCAGRSRTSGARHSAAGSDPEGPLQALIIDSWFDNYLGVVSLIRIKNGTLR
                     KGDKVKVMSTGQTYNADRLGIFTPKQVDRTELKCGEVGWLVCAIKDIGAPVGDTLTLA
                     RNPAEKALPGFKKVKPQVYAGLFPVSSDDYEAFRDALGKLSLNDASLFYEPESSSALG
                     FGFRCGFLGLLHMEIIQERLEREYDLDLITTAPTVVYEVETTSREVIYVDSPSKLPAV
                     NNIYELREPIAECHMLLPQAYLGNVITLCVEKRGVQTNMVYHGNQVALTYEIPMAEVV
                     LDFFDLKSTSRGYASLDYNFKRFQASDMVRVDVLINGERVDALALITHRDNSQNRGR
                     ELVERMKDLIPRQQFDIAIQAAIGTHIIARSTVKQLRRNVLAKCYGGDISRKKKLLQR
                     QKEGKKRMKQIGNVELPQEAPLAILHVGKDNK"
     gene            2102. .3073
                     /gene="lep"
     CDS             2102. .3073
                     /gene="lep"
                     /note="signal peptidase I (lep)"
                     /codon_start=1
                     /transl_table=11
                     /db_xref="PID:g146600"
                     /translation="MANMFALILVIATLVTGILWCVDKFFPAPKRRERQAAAQAARDS
                     LDKATLKKVAPKPGWLETGASVFPVLAIVLIVRSFIYEPPQIPSGSMMPTLLIGDFIL
                     VEKFAYGIKDPIYQKTLIENGHPKRGDIVVFKYPEDPKLDYIKRAVGLPGDKVTTDPV
                     SKELTIQPGCSSGQACENALPATYSNVEPSDFVQTFSRRNGGEATSGFFEVPKNETKE
                     NGIRLSERKETLGDVTHRILTVPIAQDQVGMYYQQPGQQLATWIVPPGQYFMMGDNRD
                     NSADSRYWGFVPEANLVGRATAIWMSFDKQEGEWPTGLRLSRIGGIH"
BASE COUNT      775 a    778 c    857 g    721 t
ORIGIN      1 bp upstream of AvaI site.

  Query Match           9.9%;  Score 17;  DB 15;  Length 3131;
  Best Local Similarity 100.0%;  Pred. No. 2.56e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2836 TCAGGTGGGGATGTATT 2852
            |||||||||||||||||
Qy     66 TCAGGTGGGGATGTATT 82


RESULT   32
LOCUS       LABORGRPL    3331 bp    DNA              BCT       24-APR-1996
DEFINITION  Plasmid pLAB1000 (from Lactobacillus hilgardii) replication (rep)
            and mobilization (mob) genes, complete cds.
ACCESSION   M55222 M37906
NID         g149352
KEYWORDS    mob gene; mobilization protein; origin of replication; rep gene;
            replication protein.
SOURCE      Plasmid pLAB1000 DNA.
  ORGANISM  Plasmid pLAB1000
            plasmids.
REFERENCE   1  (bases 1 to 3331)
  AUTHORS   Josson,K., Soetaert,P., Michiels,F., Joos,H. and Mahillon,J.
  TITLE     Lactobacillus hilgardii plasmid pLAB1000 consists of two functional
            cassettes commonly found in other gram-positive organisms
  JOURNAL   J. Bacteriol. 172 (6), 3089-3099 (1990)
```



```
  MEDLINE   90264300
FEATURES             Location/Qualifiers
     source          1. .3331
                     /organism="Plasmid pLAB1000"
                     /plasmid="Plasmid pLAB1000"
                     /specific_host="Lactobacillus hilgardii"
                     /db_xref="taxon:2591"
     gene            77. .2151
                     /gene="mob"
     CDS             77. .1162
                     /gene="mob"
                     /codon_start=1
                     /transl_table=11
                     /product="mobilization protein"
                     /db_xref="PID:g149353"
                     /translation="MSFAVARMTKLKADNLVGIGNHDQRKTTNHSNEDIDVSHSHLNY
                     DLVAGRTDNFKTDIEAYINENKASKRAVRKDAVLVNEWIITSDKDFFEQLDEAETRKY
                     FETAKQYFADNYGDENIRYAVVHMDEKTPHMHMGIVPFDDDKKLSAKRIFNRREALQRI
                     QEELPQYLKENGFDVQRGNKNKERKNLSVPEYKAMREELKKIETEKQETQAKLADTKK
                     QLDEIKPRDTKKIASKPTLMNKNKVTVDKSDLADLEQRAVTSDAYNFEKIHLEVENHS
                     LRNDLSEAKGRNYELRKENERLQKLVGTLQGIIRNVDEFLHKKLGINLPEKWLERAGL
                     KEPSKKAPESSSQERSEGHHEELDGPSL"
     terminator      1700. .1728
                     /gene="mob"
                     /note="putative"
     terminator      1756. .1777
                     /gene="mob"
                     /note="putative"
     terminator      2075. .2151
                     /gene="mob"
                     /note="putative"
     gene            2216. .3200
                     /gene="rep"
     CDS             2216. .3160
                     /gene="rep"
                     /codon_start=1
                     /transl_table=11
                     /product="replication protein"
                     /db_xref="PID:g149354"
                     /translation="MSKKILKDVSRNRKERPWRERKLENLQYAEYLRILNFKKANRVK
                     ECGEVLRFVADDEGRLRLYQTWFCKSRLCPLCNWRRSMGQSNQLMQVLDEAHKQRKTG
                     RFLPLTLTAENASGENLKQEVRKMGRAISKLFQYKKPAKNLLGYVRSTEITINKNGTY
                     HQHMHVLLFVKPTYFKDSANYINQAEWSKLWQRAMKLDYQPIVNVEAVRSNKAKGKNS
                     LIASAQETAKYQVKSKDILTNDQERDLQVVEDLEQGLAGSRQISYGGLFKEIRKQLQL
                     EDVDAHLINVDDDKVKIDEVVREVVAKWDYNKQNYFIW"
     terminator      3160. .3200
                     /gene="rep"
                     /note="putative"
BASE COUNT     1109 a    562 c    739 g    921 t
ORIGIN
```



```
  Query Match           9.9%;  Score 17;  DB 15;  Length 3331;
  Best Local Similarity 100.0%;  Pred. No. 2.56e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1316 ATAAAATAAATAATAAG 1332
            |||||||||||||||||
Qy    151 ATAAAATAAATAATAAG 167


RESULT   33
LOCUS       A14660       3331 bp    DNA              PAT       18-AUG-1994
DEFINITION  plasmid pLAB1000 or pHL1 sequence.
ACCESSION   A14660
NID         g580663
KEYWORDS    .
SOURCE      Lactobacillus hilgardii.
  ORGANISM  Lactobacillus hilgardii
            Eubacteria; Firmicutes; Low G+C gram-positive bacteria;
            Lactobacillaceae; Lactobacillus.
REFERENCE   1  (bases 1 to 3331)
  AUTHORS   Michiels,F., Delcour,J., Mahillon,J., Joos,H., Platteeuw,C. and
            Josson,K.
  TITLE     Transformed lactic acid bacteria
  JOURNAL   Patent: EP 0311469-A 4 12-APR-1989;
            PLANT GENETIC SYSTEMS N.V.; UNIVERSITE CATHOLIQUE DE LOUVAIN
FEATURES             Location/Qualifiers
     source          1. .3331
                     /organism="Lactobacillus hilgardii"
                     /db_xref="taxon:1588"
     exon            1. .78
                     /number=2
     CDS             complement(887. .1831)
                     /note="ORF 2"
                     /codon_start=1
                     /transl_table=11
                     /db_xref="PID:g580664"
                     /db_xref="SWISS-PROT:P35857"
                     /translation="MSKKILKDVSRNRKERPWRERKLENLQYAEYLRILNFKKANRVK
                     ECGEVLRFVADDEGRLRLYQTWFCKSRLCPLCNWRRSMGQSNQLMQVLDEAHKQRKTG
                     RFLPLTLTAENASGENLKQEVRKMGRAISKLFQYKKPAKNLLGYVRSTEITINKNGTY
                     HQHMHVLLFVKPTYFKDSANYINQAEWSKLWQRAMKLDYQPIVNVEAVRSNKAKGKNS
                     LIASAQETAKYQVKSKDILTNDQERDLQVVEDLEQGLAGSRQISYGGLFKEIRKQLQL
                     EDVDAHLINVDDDKVKIDEVVREVVAKWDYNKQNYFIW"
     CDS             complement(2107. .2382)
                     /note="ORF 3; Protein sequence is in conflict with the
                     conceptual translation"
                     /codon_start=1
                     /transl_table=11
                     /db_xref="PID:e115175"
                     /db_xref="PID:g1333699"
                     /translation="QHYSNLDTNRLAGTSRFTTCSACHTASGSEFERGTPLNYRVKLV
```

```
                     QLGLGSGRQRPRAWKLVDFDQFLLILLLQTTSFFWFCCKALTGRRL"
     CDS             join(3146. .3331,1. .78)
                     /note="ORF 4"
                     /codon_start=1
                     /transl_table=11
                     /db_xref="PID:g580666"
                     /translation="MIFYFQMNFFKVISVASHRPLFQISEIRFINRDFIFIHQGGFTS
                     NFFGIPWFDFIKLFFSICKLCLCLLFFGLYFFQFFPHSFVFGY"
     exon            3146. .3331
                     /number=1
BASE COUNT      921 a    739 c    562 g   1109 t
ORIGIN

  Query Match           9.9%;  Score 17;  DB 21;  Length 3331;
  Best Local Similarity 100.0%;  Pred. No. 2.56e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2715 CTTATTATTTATTTTAT 2731
         |||||||||||||||||
Cp    167 CTTATTATTTATTTTAT 151


RESULT   34
LOCUS       ECPUTA       4053 bp    DNA             BCT       06-JUL-1995
DEFINITION  E.coli (K12-CSH4) PUTA gene.
ACCESSION   X78340
NID         g467735
KEYWORDS    proline dehydrogenase; putA gene.
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Eubacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 4053)
  AUTHORS   Xia,M., Zhu,Y., Cao,X., You,L. and Chen,Z.
  TITLE     Cloning, sequencing and analysis of a gene encoding Escherichia
            coli proline dehydrogenase
  JOURNAL   FEMS Microbiol. Lett. 127 (3), 235-242 (1995)
  MEDLINE   95278725
REFERENCE   2  (bases 1 to 4053)
  AUTHORS   Zhu,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAR-1994) Y. Zhu, Peking University, The National
            Laboratory of Protein, Engineering & Plant Genetic Engineering,
            Beijing, China 100871, PROC
FEATURES                Location/Qualifiers
     source             1. .4053
                        /organism="Escherichia coli"
                        /strain="K12"
                        /sub_strain="CSH4"
                        /db_xref="taxon:562"
                        /clone="X"
```



```
                        /clone_lib="X"
     gene               1. .3933
                        /gene="PUTA"
     CDS                1. .3933
                        /gene="PUTA"
                        /EC_number="1.5.99.8"
                        /codon_start=1
                        /transl_table=11
                        /evidence=experimental
                        /product="proline dehydrogenase"
                        /db_xref="PID:g467736"
                        /translation="MGTTTMGVKLDDATPRTDQMAASRIDRPPHWLIKQAIFSYLDKL
                        ENSDTLPELPALFVGAANESEEPVAPQNGAASALSGVCRTNSSPIRLSAAISAWRRPE
                        TDAVSMLMEQARLPPPVAEQAHKLAYQLAEKLRNQKSASGRAGMVQGLLQEFSLSSQE
                        GVALMSGGSAAAYSRQSYARCVNSRQNQYWNWQSHIGRSPSLFVTAATWCLLFTGRLV
                        STHYEANLSRSLNRIIGKSGEPLIRKGVDMAMRLMGEQFVTGETIAQALANARKLEEK
                        GFRYSYDMLGERALTAADAQATWVSYQQAIHAIGKASNGRGIYEGPASRLSCPPCIHA
                        IARAIRSGMEELYPRLKSLTLLARQSTLGLNIDVRKRRIVWRSRSICWKNSASNPNWR
                        AGTGIGPVIQAYQKRCPLVIVYLVDLASRSRRPVMIRLVKGAYWDSEINARKWKGWRL
                        SGLYPQSVYRCLLSGLREKTARRPNLIYPQFASHNAHTLAAIYHLAGQNYYPGQYEPQ
                        CLHGMGEPLYEQVTGKVADGKLNRPCRIYAPVGTHETLLAYLVRRLLENGANTSFVNR
                        IADATLPLDELVADPVEAVEKLAQQEGQAGIPHPKLLPRDLYGEGRINSAGLDLAYEH
                        RLASLSSALLSNAMQKWQAKPVLEQPVADGEMTPVINPAEPKDIVGWGPRKQKARLNR
                        HCETRSIRRRFGLRRRRNAAILQRAVFNGKTEMQQVDLACWSVERETTFSMAIAEVRE
                        AVDFLHYYAGQVRDDFDNETHRPLGPVVCISPWNFPLAIFTGQNRRALAAGNSVLAKP
                        AEQTSLIAAQGIAILLEARTAGRRATVAGTGRNRRRPAYRRCACTRVMPTGSTEVATL
                        LQRNIATRLDAQGRPIPLIAETGGMNAMIVDSSALTEQVVVDVLASAFESAGQRCSAL
                        RVLCLQDDIAEHTLKMLRRPHGGVSVGESGRLTTDIGPVIDSEAKANIERHIQTMRAK
                        GRPVSQAARENSDDAQEWQTGTFVMPTLIELENFAELEKEVFGPVLHVVRYNRNQLAE
                        LIEQINASGYGVTLGVHTRIDETIAQVTGSPMSATCTLTVIWRARSSASKPFGAAACP
                        APGQKREDPLYLYRLLAHRPPNALNTTLTRQDARYPVDAQLKTTLLAPLTALTQWARI
                        ARRYRRSADNSPIWPRPARSACYRGRPASAIPGRCCRVNGYYAWLMMNRTRSQACRRS
                        RRRVRPMVRRRLPPRSGETSPRRRPRVQPAKRKRCCARFDAVIPHGDSEQAANRVEAV
                        AARKQAIVSVQGPARGESNILLERLYIERSLSVMTAAAGGNASLMTIG"
BASE COUNT      862 a   1175 c   1199 g    817 t
ORIGIN

  Query Match           9.9%;  Score 17;  DB 15;  Length 4053;
  Best Local Similarity 100.0%;  Pred. No. 2.56e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   3688 CGTGTCCAGTTTGCGAA 3704
         |||||||||||||||||
Qy     15 CGTGTCCAGTTTGCGAA 31


RESULT   35
LOCUS       DMBOSS       4082 bp    DNA             INV       05-FEB-1997
DEFINITION  D. melanogaster gene for the bride of sevenless protein.
ACCESSION   X55887
NID         g1834420
```



```
KEYWORDS    boss gene; bride of sevenless protein.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Arthropoda;
            Tracheata; Hexapoda; Insecta; Pterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 4082)
  AUTHORS   Kramer,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-SEP-1990) Kramer H., Molecular Biology Institute,
            UCLA, Dept of Biological Chemistry, The School of Medicine,
            University of California, Los Angeles, 10833 Le Conte Avenue, Los
            Angeles 90024, U S A
  REMARK    Revised by [3]
REFERENCE   2  (bases 187 to 478; 782 to 1252; 1312 to 1761; 1883 to 3342; 3402
            to 3869)
  AUTHORS   Hart,A.C., Kramer,H., Van Vactor,D.L. Jr., Paidhungat,M. and
            Zipursky,S.L.
  TITLE     Induction of cell fate in the Drosophila retina: the bride of
            sevenless protein is predicted to contain a large extracellular
            domain and seven transmembrane segments
  JOURNAL   Genes Dev. 4 (11), 1835-1847 (1990)
  MEDLINE   91115074
REFERENCE   3  (bases 1 to 4082)
  AUTHORS   Hart,A.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-FEB-1997) Hart A.C., MGH Cancer Center, CNY149-7202
            13th Street, Charlestown, MA 02129, U S A
COMMENT     On Feb 8, 1997 this sequence version replaced gi:7665.
FEATURES                Location/Qualifiers
     source             1. .4082
                        /organism="Drosophila melanogaster"
                        /strain="Oregon R"
                        /db_xref="taxon:7227"
                        /germline
                        /dev_stage="adult"
                        /clone_lib="genomic EMBL3 (Dr. M. Tanouye), head cDNA
                        lambda gt10 (dr. C. Zucker)"
                        /chromosome="3"
                        /map="96F"
     intron             1. .305
                        /number=1
     gene               305. .3661
                        /note="bride of sevenless"
                        /gene="boss"
                        /allele=""
                        /db_xref="FlyBase:FBgn0000206"
     exon               305. .478
                        /gene="boss"
                        /db_xref="FlyBase:FBgn0000206"
                        /number=2
```



```
     mRNA               join(305. .478,782. .1252,1312. .1761,1883. .3342,3404. .>3661)
                        /gene="boss"
                        /db_xref="FlyBase:FBgn0000206"
     CDS                join(427. .478,782. .1252,1312. .1761,1883. .3342,3404. .3661)
                        /gene="boss"
                        /codon_start=1
                        /product="bride of sevenless protein"
                        /db_xref="PID:e300455"
                        /db_xref="PID:g1834421"
                        /translation="MKVMDALQSGRRKPLPVALLCILVTVFCVLECHGADLTSPTKKS
                        APLRITKPQPTSQQAKPISITTRAPTTVASTTDDEVSSSVDGQLAPLISSTTEGPSSG
                        TTASLVPEICLNGLQLTVNSADEGTVIRKQEEPVKILEGDVVLSVLTKDPDSALPVIN
                        RVNQANLIMADPEIGIRAISIDNASIAENLLIQEVQPLQQCTTYSMGIPVDWELYKQL
                        ESVIRDLEYNIWPIPGTRAILPPKVAHLLHQWPWGEKIASYEIATETLEWTVNEPMEAA
                        RQEHMCLMHFKSDDNVYIMPGNKLASHFKENGTLFSVPTDRTDDEFLADLPNRAFVLM
                        ENEIDLSTAVELDATPTALDEILIGKSVLPSRVLSFAGSIIDLMNWLRGSLSKHCKRG
                        EEHDLYVLESCFNFLNFIEDWRTSEYRQAHDTAEILSLLLMRKLGTAMNFQMTQKKVL
                        TLDKITGESRTELREIASQNFVTNVTTYYHYNRDNHTSLELKTKFGQVPNCQYSAGDN
                        RRYPFLFDGESVMFWRIKMDTWVATGLTAAILGLIATLAILVFIVVRISLGDVFEGNP
                        TTSILLLLSLILVFCSFVPYSIEYVGEQRNSHVTFEDAQTLNTLCAVRVFIMTLVYCF
                        VFSLLLCRAVMLASIGSEGGFLSHVNGYIQAVICAPSVVAQVGMSVQLLVVMHVASET
                        VSCENIYYGRWLWGLLAYDPALLCCVGALIPSIYRSQRNYREGILIVIGSVLIMVIWV
                        AWIALSLFGDEWRDAAIPLGLQASGWAVLVGILIPRTFLIVRGIERSDIAQALPSLTS
                        LAFQNNQYFSEQSVTRCVVPAMRHCSQDEVNHQSPSEIPTLPLRGGGPRRQQPFANL
                        RQANANINPQRPPPRPQQSPSRSSVSSLPPSPDHNKITRF"
     intron             479. .781
                        /gene="boss"
                        /db_xref="FlyBase:FBgn0000206"
                        /number=2
     exon               782. .1252
                        /gene="boss"
                        /db_xref="FlyBase:FBgn0000206"
                        /number=3
     intron             1253. .1311
                        /gene="boss"
                        /db_xref="FlyBase:FBgn0000206"
                        /number=3
     exon               1312. .1761
                        /gene="boss"
                        /db_xref="FlyBase:FBgn0000206"
                        /number=4
     intron             1762. .1882
                        /gene="boss"
                        /db_xref="FlyBase:FBgn0000206"
                        /number=4
     exon               1883. .3342
                        /gene="boss"
                        /db_xref="FlyBase:FBgn0000206"
                        /number=5
     intron             3345. .3403
                        /gene="boss"
```

```
                     /db_xref="FlyBase:FBgn0000206"
                     /number=5
     exon            3404. .>3661
                     /gene="boss"
                     /db_xref="FlyBase:FBgn0000206"
                     /number=6
     polyA_signal    3776. .3781
     polyA_signal    3783. .3788
BASE COUNT     1078 a    965 c    939 g   1100 t
ORIGIN

  Query Match             9.9%;  Score 17;  DB 18;  Length 4082;
  Best Local Similarity 100.0%;  Pred. No. 2.56e+01;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   3771 GATAAAATAAATAATAA 3787
          |||||||||||||||||
Qy    150 GATAAAATAAATAATAA 166


RESULT   36
LOCUS       STPUTA       4102 bp    DNA             BCT       08-APR-1994
DEFINITION  S.typhimurium putA gene for proline dehydrogenase and
            pyrroline-5-carboxylic acid dehydrogenase.
ACCESSION   X70843
NID         g470179
KEYWORDS    proline dehydrogenase; putA gene; pyrroline-5-carboxylic acid
            dehydrogenase.
SOURCE      Salmonella typhimurium.
  ORGANISM  Salmonella typhimurium
            Eubacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Salmonella.
REFERENCE   1  (bases 1 to 4102)
  AUTHORS   Allen,S.W., Senti-Willis,A. and Maloy,S.R.
  TITLE     DNA sequence of the putA gene from Salmonella typhimurium: a
            bifunctional membrane-associated dehydrogenase that binds DNA
  JOURNAL   Nucleic Acids Res. 21 (7), 1676 (1993)
  MEDLINE   93241961
REFERENCE   2  (bases 1 to 3609)
  AUTHORS   Maloy,S.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-1993) S.R. Maloy, Univ of Illinois at
            Urbana-Champaign, Dept of Microbiology, 407 S. Goodwin Ave., 131
            Burrill Hall, Urbana, IL 61801, USA
COMMENT     On Apr 11, 1994 this sequence version replaced gi:49246.
FEATURES             Location/Qualifiers
     source          1. .4102
                     /organism="Salmonella typhimurium"
                     /strain="LT2"
                     /db_xref="taxon:602"
                     /map="22 minutes"
```



```
     gene            1. .3963
                     /gene="putA"
     CDS             1. .3963
                     /gene="putA"
                     /EC_number="1.5.99.8"
                     /codon_start=1
                     /transl_table=11
                     /product="proline dehydrogenase"
                     /db_xref="PID:g470180"
                     /db_xref="SWISS-PROT:P10503"
                     /translation="MGTTTMGVKLDDATRERIKMAASRIDRTPHWLIKQAIPSYLDKL
                     ENSDTLPELPALFVGAANESEEPVAPQDEPHQPFLEFAEQILPQSVSRAAITAAWRRP
                     ETDAVSMLMEQARLSPPVAEQAHKLAYQLAEKLRNQKSASGRAGMVQGLLQEFSLSSQ
                     EGVALMCLAEALLRIPDKATRDALIRDKISNGNWQSHIGRSPSLFVNAATWGLLFTGR
                     LVSTHNEANLSRSLNRIIGKSGEPLIRKGVDMAMRLMGEQFVTGETIAQALANARKLE
                     EKGFRYSYDMLGEAALTAADAQATMVSYQQAIHAIGKASNGRGIYEGPGISIKLSALH
                     PRYSRAQYDRVMEELYPRLKSLTLLARQYDIGLNIDAEEADRLEISLDLLEKLCFEPE
                     LAGWNGIGPVIQAYQKRCPLVIDYLVDLASRSRRRLMIRLVKGAYWDSEIKRAQMEGL
                     EGYPVYTRKVTTDVSYLACAKKLLAVPNLIYPQPATHNAHTLAAIYHLAGQNYYPGQY
                     EFQCLHGMGEPLYEQVTGKVADGKLNRPCRIYAPVGTHETLLAYLVRRLLENGANTSF
                     VNRIADATLPLDELVADPVEAVEKLAQQEGQAGIPHPKIPLPRDLYGEGRINSAGLDL
                     ANEHRLASLSSALLSNAMQKWQAKPVLEQPVADGEMTPVINPAEPKDIVGWGREATES
                     EVEQALQNAVNQAPVWFATPPQERAAILQRAAVLMEDQMQQLIGLLVREAGKTFSNAI
                     AEVREAVDFLHYYAGQVRDDFDNETHRPLGPVVCISPWNFPLAIFTGQIAAALAAGNS
                     VLAKPAEQTSLIAAQGIAILLEAGVPPGVVQLLPGRGETVGAQLTADARVRRVMFTGS
                     TEVATLLQRNIATRLDAQGRPIPLIAETGGMNAMIVDSSALTEQVVVDVLASAFDSAG
                     QRCSALRVLCLQDDIAETLKMLRGAMAECRMGNPGRLITTDIGPVIDSEAKANIERHI
                     QTMRAKGRPVFQAARENSDDAQEWQTGTFVMPTLIELENFAELEKEVFGPVLHVVRYN
                     RNQLAELIEQINASGYGLTLGVHTRIDETIAQVTGSAHVGNLYVNRNMVGAVVGVQPF
                     GGEGLSGTGPKAGGPLYLYRLLAHRPPNALNTTLTRQDARYPVDAQLKTTLLAPLTAL
                     TQWAADRPALQTLCRQFADLAQAGTQRLLPGPTGERNTWTLLPRERVLCLADDEQDAL
                     TQLAAVLAVGSQALWSDDAFHRDLAKRLPAAVAARVQFAKAETLMAQPFDAVIPHGDS
                     DKLRTVCEAVAAREGAIVSVQGFARGESNILLERLYIERSLSVNTAAAGGNASLMTIG
                     "

BASE COUNT      871 a   1188 c   1227 g    816 t
ORIGIN

  Query Match             9.9%;  Score 17;  DB 15;  Length 4102;
  Best Local Similarity 100.0%;  Pred. No. 2.56e+01;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   3715 CGTGTCCAGTTTGCGAA 3731
          |||||||||||||||||
Qy     15 CGTGTCCAGTTTGCGAA 31


RESULT   37
LOCUS       I35766       6295 bp    DNA             PAT       26-FEB-1997
DEFINITION  Sequence 4 from patent US 5602322.
ACCESSION   I35766
NID         g2087617
```



```
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 6295)
  AUTHORS   Ecker,J.R. and Kieber,J.J.
  TITLE     Constitutitive triple response gene and mutations
  JOURNAL   Patent: US 5602322-A 4 11-FEB-1997;
FEATURES             Location/Qualifiers
     source          1. .6295
                     /organism="unknown"
BASE COUNT     1815 a   1104 c   1295 g   2081 t
ORIGIN

  Query Match             9.9%;  Score 17;  DB 21;  Length 6295;
  Best Local Similarity 100.0%;  Pred. No. 2.56e+01;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    272 TTTTATCTTCCAATTTC 288
          |||||||||||||||||
Cp    156 TTTTATCTTCCAATTTC 140


RESULT   38
LOCUS       I14048       6295 bp    DNA             PAT       30-AUG-1995
DEFINITION  Sequence 4 from patent US 5444166.
ACCESSION   I14048
NID         g996471
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 6295)
  AUTHORS   Ecker,J.R. and Kieber,J.J.
  TITLE     Constitutive triple response gene and mutations
  JOURNAL   Patent: US 5444166-A 4 22-AUG-1995;
FEATURES             Location/Qualifiers
     source          1. .6295
                     /organism="unknown"
BASE COUNT     1815 a   1104 c   1295 g   2081 t
ORIGIN

  Query Match             9.9%;  Score 17;  DB 21;  Length 6295;
  Best Local Similarity 100.0%;  Pred. No. 2.56e+01;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    272 TTTTATCTTCCAATTTC 288
          |||||||||||||||||
Cp    156 TTTTATCTTCCAATTTC 140
```



```
RESULT   39
LOCUS       I35768       6312 bp    DNA             PAT       26-FEB-1997
DEFINITION  Sequence 6 from patent US 5602322.
ACCESSION   I35768
NID         g2087619
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 6312)
  AUTHORS   Ecker,J.R. and Kieber,J.J.
  TITLE     Constitutitive triple response gene and mutations
  JOURNAL   Patent: US 5602322-A 6 11-FEB-1997;
FEATURES             Location/Qualifiers
     source          1. .6312
                     /organism="unknown"
BASE COUNT     1818 a   1107 c   1297 g   2090 t
ORIGIN

  Query Match             9.9%;  Score 17;  DB 21;  Length 6312;
  Best Local Similarity 100.0%;  Pred. No. 2.56e+01;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    272 TTTTATCTTCCAATTTC 288
          |||||||||||||||||
Cp    156 TTTTATCTTCCAATTTC 140


RESULT   40
LOCUS       I35769       6312 bp    DNA             PAT       26-FEB-1997
DEFINITION  Sequence 7 from patent US 5602322.
ACCESSION   I35769
NID         g2087620
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 6312)
  AUTHORS   Ecker,J.R. and Kieber,J.J.
  TITLE     Constitutitive triple response gene and mutations
  JOURNAL   Patent: US 5602322-A 7 11-FEB-1997;
FEATURES             Location/Qualifiers
     source          1. .6312
                     /organism="unknown"
BASE COUNT     1818 a   1107 c   1296 g   2091 t
ORIGIN

  Query Match             9.9%;  Score 17;  DB 21;  Length 6312;
  Best Local Similarity 100.0%;  Pred. No. 2.56e+01;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```

```
Db    272 TTTTATCTTCCAATTTC 288
          ||||||||||||||||
Cp    156 TTTTATCTTCCAATTTC 140


RESULT   41
LOCUS       ATHCTR1B     6312 bp    DNA             PLN       02-MAR-1993
DEFINITION  Arabidopsis thaliana negative regulator of the ethylene response
            pathway (CTR1) gene, complete cds.
ACCESSION   L08790
NID         g166681
KEYWORDS    ethylene response pathway; negative regulator; protein kinase.
SOURCE      Arabidopsis thaliana (strain Columbia) DNA.
  ORGANISM  Arabidopsis thaliana
            Eukaryotae; mitochondrial eukaryotes; Viridiplantae;
            Charophyta/Embryophyta group; Embryophyta; Magnoliophyta;
            Magnoliopsida; Capparales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 6312)
  AUTHORS   Kieber,J.J., Rothenberg,M., Roman,G., Feldmann,K.A. and Ecker,J.R.
  TITLE     CTR1, a negative regulator of the ethylene response pathway in
            Arabidopsis, encodes a member of the Raf family of protein kinases
  JOURNAL   Cell 72, 427-441 (1993)
  MEDLINE   93161417
FEATURES             Location/Qualifiers
     source          1. .6312
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
     gene            join(776. .1423,1599. .1900,1998. .2139,2357. .2460,
                     2596. .2801,3159. .3434,3625. .3665,3942. .4010,4091. .4191,
                     4281. .4365,4460. .4558,4792. .4860,4964. .5095,5210. .5304,
                     5382. .5478)
                     /gene="CTR1"
     exon            <776. .1423
                     /gene="CTR1"
                     /number=1
     CDS             join(776. .1423,1599. .1900,1998. .2139,2357. .2460,
                     2596. .2801,3159. .3434,3625. .3665,3942. .4010,4091. .4191,
                     4281. .4365,4460. .4558,4792. .4860,4964. .5095,5210. .5304,
                     5382. .5478)
                     /gene="CTR1"
                     /codon_start=1
                     /product="protein kinase"
                     /db_xref="PID:g166682"
                     /translation="MEMPGRRSNYTLLSQFSDDQVSVSVTGAPPPHYDSLSSENRSNH
                     NSGNTGKAKAERGGFDWDPSGGGGGDHRLNMQPNRVGNNMYASSLGLQRQSSGSSPGE
                     SSLSGDYYMPTLSAAANEIESVGFPQDDGFRLGPGGGGGDLRIQMAADSAGGSSSGKS
                     WAQQTEESYQLQLALALRLSSEATCADDPNFLDPVPDESALRTSPSSAETVSHRPWVN
                     GCLSYYDKVPDGFYMMNGLDPYIWTLCIDLHESGRIPSIESLRAVDSGVDSSLEAIIV
                     DRRSDPAFKELHNRVHDISCSCITTKEVVDQLAKLICNRMGGPVIMGEDELVPMWKEC
                     IDGLKEIFKVVVPIGSLSVGLCRHRALLFKVLADIIDLPCRIAKGCKYCNRDDAASCL
```



```
                     VRFGLDREYLVDLVGKPGHLWEPDSLLNGPSSISISSPLRFPRPKPVEPAVDFRLLAK
                     QYFSDSQSLNLVFDPASDDMGFSMFHRQYDNPGGENDALAENGGGSLPPSANMPPQNM
                     MRASNQIEAAPMNAPPISQPVPNRANRELGLDGDDMDIPWCDLNIKEKIGAGSFGTVH
                     RAEWHGSDVAVKILMEQDFHAERVNEFLREVAIMKRLRHPNIVLFMGAVTQPPNLSIV
                     TEYLSRGSLYRLLHKSGAREQLDERRRLSMAYDVAKGMNYLHNRNPPIVHRDLKSPNL
                     LVDKKYTVKVCDFGLSRLKASTFLSSKSAAGTPEWMAPEVLRDEPSNEKSDVYSFGVI
                     LWELATLQQPWGNLNPAQVVAAVGFKCKRLEIPRNLNPQVAAIIEGCWTNEPWKRPSF
                     ATIMDLLRPLIKSAVPPPNRSDL"
     intron          1424. .1598
                     /gene="CTR1"
                     /number=1
     exon            1599. .1900
                     /gene="CTR1"
                     /number=2
     intron          1901. .1997
                     /gene="CTR1"
                     /number=2
     exon            1998. .2139
                     /gene="CTR1"
                     /number=3
     intron          2140. .2356
                     /gene="CTR1"
                     /number=3
     exon            2357. .2460
                     /gene="CTR1"
                     /number=4
     intron          2461. .2595
                     /gene="CTR1"
                     /number=4
     exon            2596. .2801
                     /gene="CTR1"
                     /number=5
     intron          2802. .3158
                     /gene="CTR1"
                     /number=5
     exon            3159. .3434
                     /gene="CTR1"
                     /number=6
     intron          3435. .3624
                     /gene="CTR1"
                     /number=6
     exon            3625. .3665
                     /gene="CTR1"
                     /number=7
     intron          3666. .3941
                     /gene="CTR1"
                     /number=7
     exon            3942. .4010
                     /gene="CTR1"
                     /number=8
     intron          4011. .4090
```



```
                     /gene="CTR1"
                     /number=8
     exon            4091. .4191
                     /gene="CTR1"
                     /number=9
     intron          4192. .4280
                     /gene="CTR1"
                     /number=9
     exon            4281. .4365
                     /gene="CTR1"
                     /number=10
     intron          4366. .4459
                     /gene="CTR1"
                     /number=10
     exon            4460. .4558
                     /gene="CTR1"
                     /number=11
     intron          4559. .4791
                     /gene="CTR1"
                     /number=11
     exon            4792. .4860
                     /gene="CTR1"
                     /number=12
     intron          4861. .4963
                     /gene="CTR1"
                     /number=12
     exon            4964. .5095
                     /gene="CTR1"
                     /number=13
     intron          5096. .5209
                     /gene="CTR1"
                     /number=13
     exon            5210. .5304
                     /gene="CTR1"
                     /number=14
     intron          5305. .5381
                     /gene="CTR1"
                     /number=14
     exon            5382. .>5478
                     /gene="CTR1"
                     /number=15
BASE COUNT     1816 a   1107 c   1298 g   2091 t
ORIGIN

  Query Match           9.9%;  Score 17;  DB 23;  Length 6312;
  Best Local Similarity 100.0%;  Pred. No. 2.56e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    272 TTTTATCTTCCAATTTC 288
          ||||||||||||||||
Cp    156 TTTTATCTTCCAATTTC 140
```



```
RESULT   42
LOCUS       I14047       6312 bp    DNA             PAT       30-AUG-1995
DEFINITION  Sequence 3 from patent US 5444166.
ACCESSION   I14047
NID         g996470
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 6312)
  AUTHORS   Ecker,J.R. and Kieber,J.J.
  TITLE     Constitutive triple response gene and mutations
  JOURNAL   Patent: US 5444166-A 3 22-AUG-1995;
FEATURES             Location/Qualifiers
     source          1. .6312
                     /organism="unknown"
BASE COUNT     1817 a   1107 c   1297 g   2091 t
ORIGIN

  Query Match           9.9%;  Score 17;  DB 21;  Length 6312;
  Best Local Similarity 100.0%;  Pred. No. 2.56e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    272 TTTTATCTTCCAATTTC 288
          ||||||||||||||||
Cp    156 TTTTATCTTCCAATTTC 140


RESULT   43
LOCUS       I14049       6312 bp    DNA             PAT       30-AUG-1995
DEFINITION  Sequence 5 from patent US 5444166.
ACCESSION   I14049
NID         g996472
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 6312)
  AUTHORS   Ecker,J.R. and Kieber,J.J.
  TITLE     Constitutive triple response gene and mutations
  JOURNAL   Patent: US 5444166-A 5 22-AUG-1995;
FEATURES             Location/Qualifiers
     source          1. .6312
                     /organism="unknown"
BASE COUNT     1817 a   1106 c   1297 g   2092 t
ORIGIN

  Query Match           9.9%;  Score 17;  DB 21;  Length 6312;
  Best Local Similarity 100.0%;  Pred. No. 2.56e+01;
```

```
    Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    272 TTTTATCTTCCAATTTC 288
          ||||||||||||||||
Cp    156 TTTTATCTTCCAATTTC 140


RESULT   44
LOCUS       I14051       6312 bp    DNA              PAT       30-AUG-1995
DEFINITION  Sequence 7 from patent US 5444166.
ACCESSION   I14051
NID         g996474
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 6312)
  AUTHORS   Ecker,J.R. and Kieber,J.J.
  TITLE     Constitutive triple response gene and mutations
  JOURNAL   Patent: US 5444166-A 7 22-AUG-1995;
FEATURES             Location/Qualifiers
     source          1. .6312
                     /organism="unknown"
BASE COUNT     1818 a   1107 c   1296 g   2091 t
ORIGIN

  Query Match           9.9%;  Score 17;  DB 21;  Length 6312;
  Best Local Similarity 100.0%;  Pred. No. 2.56e+01;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    272 TTTTATCTTCCAATTTC 288
          ||||||||||||||||
Cp    156 TTTTATCTTCCAATTTC 140


RESULT   45
LOCUS       I14050       6312 bp    DNA              PAT       30-AUG-1995
DEFINITION  Sequence 6 from patent US 5444166.
ACCESSION   I14050
NID         g996473
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 6312)
  AUTHORS   Ecker,J.R. and Kieber,J.J.
  TITLE     Constitutive triple response gene and mutations
  JOURNAL   Patent: US 5444166-A 6 22-AUG-1995;
FEATURES             Location/Qualifiers
     source          1. .6312
                     /organism="unknown"
```



```
BASE COUNT     1818 a   1107 c   1297 g   2090 t
ORIGIN

  Query Match           9.9%;  Score 17;  DB 21;  Length 6312;
  Best Local Similarity 100.0%;  Pred. No. 2.56e+01;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    272 TTTTATCTTCCAATTTC 288
          ||||||||||||||||
Cp    156 TTTTATCTTCCAATTTC 140


RESULT   46
LOCUS       I35765       6312 bp    DNA              PAT       26-FEB-1997
DEFINITION  Sequence 3 from patent US 5602322.
ACCESSION   I35765
NID         g2087616
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 6312)
  AUTHORS   Ecker,J.R. and Kieber,J.J.
  TITLE     Constitutitive triple response gene and mutations
  JOURNAL   Patent: US 5602322-A 3 11-FEB-1997;
FEATURES             Location/Qualifiers
     source          1. .6312
                     /organism="unknown"
BASE COUNT     1817 a   1107 c   1297 g   2091 t
ORIGIN

  Query Match           9.9%;  Score 17;  DB 21;  Length 6312;
  Best Local Similarity 100.0%;  Pred. No. 2.56e+01;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    272 TTTTATCTTCCAATTTC 288
          ||||||||||||||||
Cp    156 TTTTATCTTCCAATTTC 140


RESULT   47
LOCUS       I35767       6312 bp    DNA              PAT       26-FEB-1997
DEFINITION  Sequence 5 from patent US 5602322.
ACCESSION   I35767
NID         g2087618
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 6312)
  AUTHORS   Ecker,J.R. and Kieber,J.J.
```



```
  TITLE     Constitutitive triple response gene and mutations
  JOURNAL   Patent: US 5602322-A 5 11-FEB-1997;
FEATURES             Location/Qualifiers
     source          1. .6312
                     /organism="unknown"
BASE COUNT     1817 a   1106 c   1297 g   2092 t
ORIGIN

  Query Match           9.9%;  Score 17;  DB 21;  Length 6312;
  Best Local Similarity 100.0%;  Pred. No. 2.56e+01;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    272 TTTTATCTTCCAATTTC 288
          ||||||||||||||||
Cp    156 TTTTATCTTCCAATTTC 140


RESULT   48
LOCUS       DROBRSEVM    7553 bp    DNA              INV       29-AUG-1994
DEFINITION  Drosophila melanogaster bride of sevenless protein gene, complete
            cds.
ACCESSION   L08133
NID         g290217
KEYWORDS    bride of sevenless protein.
SOURCE      Drosophila melanogaster DNA.
  ORGANISM  Drosophila melanogaster
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Arthropoda;
            Tracheata; Insecta; Pterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 7553)
  AUTHORS   Hart,A.C., Harrison,S.D., Van Vactor,D.L. Jr., Rubin,G.M. and
            Zipursky,S.L.
  TITLE     The interaction of bride of sevenless with sevenless is conserved
            between Drosophila virilis and Drosophila melanogaster
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 90 (11), 5047-5051 (1993)
  MEDLINE   93281693
FEATURES             Location/Qualifiers
     source          1. .7553
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /tissue_lib="lambda"
     exon            862. .911
                     /gene="bride of sevenless"
                     /note="5' UTR"
                     /number=1
     intron          912. .2304
                     /gene="bride of sevenless"
                     /number=1
     gene            join(2427. .2477,2718. .3254,3314. .3760,3882. .5342,
                     5404. .5661)
                     /note="bride of sevenless"
```



```
                     /gene="boss"
                     /allele=""
                     /db_xref="FlyBase:FBgn0000206"
     CDS             join(2427. .2477,2718. .3254,3314. .3760,3882. .5342,
                     5404. .5661)
                     /gene="boss"
                     /note="bride of sevenless"
                     /codon_start=1
                     /db_xref="PID:g290218"
                     /translation="MKVMDALQSGRRKPLPVFKDDDIFHLIRIRFNSLLSSAALLCIL
                     VTVFCVLECHGADLTSPTKKSAPLRITKPQPTSQQAKPISITTRAPTTVASTTDDEVS
                     SSVDGQLAPLISSTTEGPSSGTTASLVPEICLNGLQLTVNSADEGTVIRKQEEFVKIL
                     EGDVVLSVLTKDPDSALFVINRVNQANLIMADFEIGIRAISIDNASLAENLLIQEVQF
                     LQQCTTTYSMGIFVDWELVKQLESVIKDLEYNIWPIPGTRAHLFPKVAHLLHQMPWGEK
                     IASVEIATETLEMYNEFMEAARQEHDCLMHFKSDDNVYIMFGNKLASHFKENGTLFSV
                     PTDRTDDEFLADLPNRAFVLMENEIDLSTAVELDATPTALDEILIGKSVLPSRVLSFA
                     GSIIDLMNWLRGSLSKHCKRGEEHDLYVLESCFNFLNFIEDWRTSETRQAHDYAEILS
                     LLLMRKLGTAMNPQMYQKKVLTLDKITGESRTELREIASQNFVTNVTTYTHYNRDNHT
                     SLELKTKFGQVFNCQYSAGDNRRYPFLFDGESVMFWRIKMDTWVATGLTAAILGLIAT
                     LAILVFIVVRISLGDVPEGNPTTSILLLLSLILVFCSFVPYSIEYVGEQRNSHVTFED
                     AQTLNTLCAVRVFIMTLVYCFVPSLLLCRAVMLASIGSEGGFLSHVNGYIQAVICAFS
                     VVAQVGMSVQLLVVMHVASETVSCENIYYGRWLWGLLAYDFALLCVGALIPSIYRSQ
                     RNYREGILIVIGSVLIMVTHVAWIALSLFGDEWRDAAIPLGLQASGWAVLVGILIPRT
                     FLIVRGIERSDIAQALPSLTSLAFAQNNQYSSEQSVYECVNPAMRHCSQDEVNHQSPS
                     EIPTLPLRGGGPRRQQFFANLRQANANINPQRPPPRPQQSPSRSSVSSLPPSPDHNKI
                     TRF"
     intron          2478. .2717
                     /gene="bride of sevenless"
                     /number=2
     exon            2718. .3254
                     /gene="boss"
                     /db_xref="FlyBase:FBgn0000206"
                     /number=3
     intron          3255. .3313
                     /gene="bride of sevenless"
                     /number=3
     exon            3314. .3760
                     /gene="boss"
                     /db_xref="FlyBase:FBgn0000206"
                     /number=4
     intron          3761. .3881
                     /gene="bride of sevenless"
                     /number=4
     exon            3882. .5342
                     /gene="boss"
                     /db_xref="FlyBase:FBgn0000206"
                     /number=5
     intron          5343. .5403
                     /gene="bride of sevenless"
                     /number=5
     exon            5404. .>5661
```

```
                     /gene="bcsS"
                     /note="5' UTR"
                     /db_xref="FlyBase:FBgn0000206"
                     /number=6
BASE COUNT     2123 a   1722 c   1703 g   2005 t
ORIGIN

  Query Match             9.9%;  Score 17;  DB 18;  Length 7553;
  Best Local Similarity 100.0%;  Pred. No. 2.56e+01;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   5771 GATAAAATAAATAATAA 5787
          |||||||||||||||||
Qy    150 GATAAAATAAATAATAA 166


RESULT   49
LOCUS       AF036961     7603 bp    DNA             BCT       03-SEP-1998
DEFINITION  Mycoplasma hominis glucose 1-phosphatase homolog gene, partial cds;
            topoisomerase IV subunit (parE) and topoisomerase IV subunit (parC)
            genes, complete cds; dipeptidase homolog gene, partial cds; and
            unknown gene.
ACCESSION   AF036961
NID         g3513504
KEYWORDS    .
SOURCE      Mycoplasma hominis.
  ORGANISM  Mycoplasma hominis
            Eubacteria; Firmicutes; clostridial firmicutes; Mycoplasmas and
            walled relatives; Mycoplasmatales; Mycoplasmataceae; Mycoplasma.
REFERENCE   1  (bases 1 to 7603)
  AUTHORS   Bebear,C.M., Charron,A., Bove,J.M., Bebear,C. and Renaudin,J.
  TITLE     Cloning and nucleotide sequences of the topoisomerase IV parC and
            parE genes of Mycoplasma hominis
  JOURNAL   Antimicrob. Agents Chemother. 42 (8), 2024-2031 (1998)
  MEDLINE   98353329
REFERENCE   2  (bases 1 to 7603)
  AUTHORS   Bebear,C.M., Charron,A., Bove,J.M., Bebear,C. and Renaudin,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-DEC-1997) Laboratoire de Bacteriologie, Universite
            Bordeaux2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .7603
                     /organism="Mycoplasma hominis"
                     /strain="PG21"
                     /db_xref="taxon:2098"
     CDS             <1. .261
                     /note="ORF1"
                     /codon_start=1
                     /transl_table=4
                     /product="glucose 1-phosphatase homolog"
                     /db_xref="PID:g3513505"
```



```
                     /translation="EFNNGKDLTLIVGHDTNIATTLAMLDVKMPDLKQLEKYPIGSKL
                     IFAINDDNSFDLYIAFFDFKDIRNFNLSNPQIIKIASKIFLK"
     gene            315. .2234
                     /gene="parE"
     CDS             315. .2234
                     /gene="parE"
                     /codon_start=1
                     /transl_table=4
                     /product="topoisomerase IV subunit"
                     /db_xref="PID:g3513506"
                     /translation="MSLNNYEASDLKVLKGLDAVRRRPGMYIGSTDSSGLHHLIWESF
                     DNAIDEALSGFANKIIVTLKKDSSVIVEDNGRGIPVNKHESGKSGVELVFTELHAGGK
                     FGDGAYKTSGGLHGVGSSVVNALSSKVEVTVWRDKKAYFTAFQQDTIIEKTHEISPTT
                     KQGTKVQFWADFSIFKKSKYSPEIIKEKLREASFLMPNLTIEFYNENTDEHVTFKTKD
                     GIKEYINFVGEGKKFISEIFYASGMCNYGIEIDLALVYTDGYSESIYSFVNNVKTRDG
                     GTHETALKTALTKTINDYASKNGLAKKITFEGSDVREGIIAVISLKVPEPILEFVGQT
                     KDKLGTPEAREVVEDFVSQKFYFFLNEKKIEAEKIISKIKKAYEAYVAARNAKNEARK
                     IKNKFENRKILSGKLTPAQSKNPAVKELFLVEGDSAGGSAKLGRNRVTQAILPLRGKV
                     INTDKAKLSDVLANEEIATIINTIGAGIGEDFNLKNAQYHKIIIMTDADTDGAHIQTL
                     LLTFFWRYMKKLVENGMVYIALPPLYKISIKSKNTQPLYAWDEDQLKEITSKYTNYEI
                     QRYKGLGEMNADQLWETTMNPETRSIIKVDVEDEILVERNVSTLMSDDVIARKQWINN
                     NVDPSSEDDFKINNK"
     gene            2240. .4840
                     /gene="parC"
     CDS             2240. .4840
                     /gene="parC"
                     /codon_start=1
                     /transl_table=4
                     /product="topoisomerase IV subunit"
                     /db_xref="PID:g3513507"
                     /translation="MKKDRKEEIQEVTENIIEKNMADIMSDRPGRYSKYIIQQRAIPD
                     ARDGLKPVQRRILYSMWNLHLKNSEPFKKSARIVGDVIGRYHPHGDSSIYEALVRMAQ
                     DWKSNFPLIEMHGNKGSIDDDPAAAMRYTESRLEKISELMLKDLDRKVVKMAPNFDDS
                     EYEPIVLPALFPNLLVNGAKGIAAGFATEIPPHNLGEVIDATIALIKNPTISIEELSE
                     IVKGPDFPTGAIINGINEIKKAYLSGQGRITISSKYHYVYDKKDFSKIIGIEIIEIPF
                     GVVKSKLVADIDAIAIDKKISGIKEVLDQTDRNGISIFIQLEDGANADAIIAYLMNKT
                     ELSISTSYKMVAIDNNRPVILNLYGALIAYLSHLKEVNINGINYDLKKFKLRLEIVEG
                     FIKVAEISDEVIHLIKESDNSKKGVILALMNKFKFSELQATAIAELRLYKLSRMDQIE
                     FQEEKKNLEIQIENCNKLLNDKWEFNQYLIKQLLEIRNQYSKPRLTEISDQKIDKEID
                     HKLLTKNEDFYLYITKDGYYKKISLKVTSNELSTFKLKEEDNVFYFDKVNSLSKILF
                     FTNLGNYFIIDCHLFKDCNWKDLGQHISSIVALESSEKIIRVIEITSPNNYANFILMS
                     KLGYAKKVNLRDFENKSSLKTKTCMSFKDDNDELIDAQISNDEKMLFILLNNGMTHLV
                     SENELKVGISLKARGIRLLLNLYKHPQLQVSGFITVSKYNNIIYLTQGGYIKCWDTSK
                     LESTTRNTPKMLPTPLKNNILGLQSLAVTLSNLKMLYTDNNGNLAEYDWKFILKDKTK
                     ESKLLKLDYSPTNPGYFITPIKINELIEVDEIBQEKIRQETQGYIDKILNWPLNMLWL
                     KNPIIKIFNI"
     CDS             5077. .5190
                     /note="ORF4"
                     /codon_start=1
                     /transl_table=4
                     /product="unknown"
```



```
                     /db_xref="PID:g3513508"
                     /translation="MRKDVRLANKVKAAKREHKLALEKARARRRLAKQQAE"
     misc_feature    5347. .6432
                     /note="similar to gene encoding ribonuclease precursor;
                     contains internal stop codon"
     CDS             6508. .>7603
                     /note="ORF6"
                     /codon_start=1
                     /transl_table=4
                     /product="dipeptidase homolog"
                     /db_xref="PID:g3513509"
                     /translation="MKKFITYKQSKEDFDEMIKNIANICAIPSISEEDFSSEFPFGKE
                     TDNALNYALNLARSPGFNIYKDPKNRYGFAEPGNGDKIIGILAHLDVVPAGDESQWRT
                     SAFVPVITNESIIGRGSLDDKGPAIINLYAMKYIHDHGLLSNEWKIRIIFCISEETTM
                     KSMKSYLADFKDPYISTTPDGEWPLIYAERMIVVTEIHPPKVPGVELNGGDVVNQIPD
                     STSLVLNEKPNIIKGIGGHGSTPEKGDNAIIKAINIASKENKEPAKSKLFKFIKENLN
                     SKSFNLSNIFKNYSDFSGELSANLGIIRTLKNETVLTFDFRVPVSHKIDDVTKSLNDY
                     LKDEPGDDVFSKLQGFKEPKYIEQDSPIVRI"
BASE COUNT     3059 a    956 c   1240 g   2348 t
ORIGIN

  Query Match             9.9%;  Score 17;  DB 16;  Length 7603;
  Best Local Similarity 100.0%;  Pred. No. 2.56e+01;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2216 TAAAATAAATAATAAGT 2232
          |||||||||||||||||
Qy    152 TAAAATAAATAATAAGT 168


RESULT   50
LOCUS       CHBLGLPS     7634 bp    DNA             MAM       04-JAN-1995
DEFINITION  C.hircus beta-lactoglobulin pseudogene.
ACCESSION   Z47079
NID         g619077
KEYWORDS    beta-lactoglobulin; pseudogene.
SOURCE      goat.
  ORGANISM  Capra hircus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Artiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Caprinae; Capra.
REFERENCE   1  (bases 1 to 7634)
  AUTHORS   Folch,J., Coll,A., Hayes,H. and Sanchez,A.
  TITLE     Characterization of a goat beta-lactoglobulin pseudogene
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 7634)
  AUTHORS   Folch,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-DEC-1994) Folch J., Facultat de Veterinaria,
            Universitat Autonoma de Barcelona, Patologia i Produccio Animals,
            Bellaterra, Barcelona, SPAIN, 08193
```



```
FEATURES             Location/Qualifiers
     source          1. .7634
                     /organism="Capra hircus"
                     /db_xref="taxon:9925"
                     /clone="LL"
                     /tissue_type="liver"
                     /clone_lib="lambda EMBL3 genomic library"
     TATA_signal     1143. .1149
     exon            1178. .1311
                     /number=1
                     /pseudo
                     /product="beta-lactoglobulin pseudogene"
     intron          1312. .1812
                     /number=1
     exon            1813. .1981
                     /number=2
                     /pseudo
                     /product="beta-lactoglobulin pseudogene"
     intron          1982. .2538
                     /cons_splice=(5'site:NO,3'site:YES)
                     /number=2
     exon            2539. .2612
                     /number=3
                     /pseudo
                     /product="beta-lactoglobulin pseudogene"
     intron          2613. .4792
                     /number=3
     repeat_region   3606. .3733
                     /rpt_type=TANDEM
                     /rpt_unit=3606. .3637
     repeat_unit     3971. .4053
                     /partial
                     /rpt_type=DISPERSED
                     /rpt_family="SINE Bov-tA"
     exon            4793. .4906
                     /number=4
                     /pseudo
                     /product="beta-lactoglobulin pseudogene"
     intron          4907. .5698
                     /cons_splice=(5'site:NO,3'site:YES)
                     /number=4
     exon            5699. .5804
                     /number=5
                     /pseudo
                     /product="beta-lactoglobulin pseudogene"
     intron          5805. .5952
                     /number=5
     exon            5953. .5994
                     /number=6
                     /pseudo
                     /product="beta-lactoglobulin pseudogene"
```

```
     intron          5995. .6385
                     /number=6
     exon            6386. .6563
                     /number=7
                     /pseudo
                     /product="beta-lactoglobulin pseudogene"
     polyA_signal    6545. .6550
BASE COUNT     1418 a   2383 c   2363 g   1470 t
ORIGIN
```

Query Match 9.9%; Score 17; DB 19; Length 7634;
Best Local Similarity 100.0%; Pred. No. 2.56e+01;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Db    305 AGAAGGTGGAAGAAGGC 321
          |||||||||||||||||
Qy     89 AGAAGGTGGAAGAAGGC 105
```

Search completed: Tue Dec 22 21:47:30 1998
Job time : 543 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Tue Dec 22 22:04:47 1998; MasPar time 163.60 Seconds
963.279 Million cell updates/sec

Tabular output not generated.

Title: >US-08-530-112A-3084
Description: (1-88) from US08530112A.seq
Perfect Score: 87
N.A. Sequence: 1 GATCAATCTTTGATGGTGAG.....AATCTATCTNTACATCTAAA 88
Comp: CTAGTTAGAAACTACCACTC.....TTAGATAGANATGTAGATTT

Scoring table: TABLE jmetric
Gap 60

Nmatch STD : Dbase 0; Query 0

Searched: 2275026 seqs, 895388244 bases x 2

Post-processing: Minimum Match 0%
Listing first 1000 summaries

Database: embl-est55
1:em_est1 2:em_gss1 3:em_gss2 4:em_gss3
Database: genbank-est107
5:gb_est1 6:gb_est10 7:gb_est11 8:gb_est12 9:gb_est13
10:gb_est14 11:gb_est15 12:gb_est16 13:gb_est17
14:gb_est18 15:gb_est19 16:gb_est2 17:gb_est20
18:gb_est21 19:gb_est3 20:gb_est4 21:gb_est5 22:gb_est6
23:gb_est7 24:gb_est8 25:gb_est9 26:gb_gss1 27:gb_gss2
28:gb_gss3 29:gb_gss4

Statistics: Mean 7.481; Variance 1.410; scale 5.307

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
	1	87	100.0	88	7	HUMGS03628	Human aortic endothel	2.90e-137
c	2	87	100.0	284	18	AI183912	qe23e10.x1 Soares_feta	2.90e-137
c	3	87	100.0	375	18	AI184115	qe24e12.x1 Soares_feta	2.90e-137
c	4	87	100.0	438	17	AI017590	ou23h06.x1 Soares_NFL_	2.90e-137
c	5	87	100.0	455	8	AA496035	zv72d06.s1 Soares tota	2.90e-137
c	6	87	100.0	501	6	AA032240	zk19a02.s1 Soares preg	2.90e-137
c	7	85	97.7	324	10	AA535684	nz33c01.s1 NCI_CGAP_Pr	4.81e-133
c	8	85	97.7	385	19	N27289	yw71b11.s1 Homo sapien	4.81e-133
c	9	84	96.6	416	6	AA043991	zk58h02.s1 Soares preg	6.15e-131
c	10	84	96.6	422	19	H62857	yr18b08.s1 Homo sapien	6.15e-131
c	11	84	96.6	562	11	AA700634	zi43b04.s1 Soares feta	6.15e-131
c	12	82	94.3	414	25	AA347704	EST54313 Fetal heart I	9.87e-127
c	13	76	87.4	307	5	T57478	yb56d02.s1 Homo sapien	3.54e-114
c	14	57	65.5	317	15	AA928824	on99c05.s1 Soares_NFL_	3.30e-75
c	15	52	59.8	452	17	AA862584	oh44e12.s1 NCI_CGAP_GC	3.20e-65
	16	21	24.1	303	19	N55979	J4995F Homo sapiens cD	6.76e-09
	17	19	21.8	772	26	B07885	YX87T7 pBluescript exp	6.04e-06
c	18	18	20.7	352	25	AA368830	EST80142 Placenta I Ho	1.56e-04
	19	18	20.7	402	26	B72729	RPCI11-10C13.TV RPCI11	1.56e-04
	20	18	20.7	565	10	AA415844	vd26g11.s1 Knowles Sol	1.56e-04
	21	18	20.7	612	11	C84145	Dictyostelium discoide	1.56e-04
	22	17	19.5	150	10	W14395	ma71h08.r1 Soares mous	3.57e-03
c	23	17	19.5	304	7	HUM163F05B	Human fetal brain cDNA	3.57e-03
	24	17	19.5	313	26	B72775	RPCI11-10E13.TV RPCI11	3.57e-03
c	25	17	19.5	317	11	H36017	14539 Lambda-PRL2 Arab	3.57e-03
	26	17	19.5	335	8	AA213510	zr92a04.r1 NCI_CGAP_GC	3.57e-03
	27	17	19.5	337	5	T92994	ye23c12.r1 Homo sapien	3.57e-03
c	28	17	19.5	341	9	AA589763	vl79a08.r1 Stratagene	3.57e-03
c	29	17	19.5	353	13	AA211119	zq87e10.s1 Stratagene	3.57e-03
c	30	17	19.5	367	15	AI048690	ub30g02.r1 Soares 2NbM	3.57e-03
	31	17	19.5	373	8	AA561954	vl40b10.r1 Stratagene	3.57e-03
c	32	17	19.5	379	26	B54002	CIT-HSP-2014D3.TR CIT-	3.57e-03
	33	17	19.5	383	20	W58969	md62b11.r1 Soares mous	3.57e-03
c	34	17	19.5	385	21	W02716	zc65b04.s1 Soares feta	3.57e-03
c	35	17	19.5	388	7	HUM070H01B	Human fetal brain cDNA	3.57e-03
c	36	17	19.5	397	28	AQ199311	RPCI11-58B19.TJ RPCI11	3.57e-03
	37	17	19.5	399	5	R32738	yh71g12.r1 Homo sapien	3.57e-03
c	38	17	19.5	404	18	AI145082	UI-R-BT0-pt-h-01-0-UI.	3.57e-03
c	39	17	19.5	417	6	AA043790	zk59f12.s1 Soares preg	3.57e-03
	40	17	19.5	423	16	H27973	yl62f12.r1 Homo sapien	3.57e-03
	41	17	19.5	434	10	AA648291	ns20e02.s1 NCI_CGAP_GC	3.57e-03
	42	17	19.5	443	27	AQ057587	CIT-HSP-2345K18.TR CIT	3.57e-03

	43	17	19.5	455	27	AQ052301	RPCI11-51C19.TJ RPCI11	3.57e-03
c	44	17	19.5	465	18	AI187151	qe42h06.x1 Soares_feta	3.57e-03
	45	17	19.5	466	12	AA760212	vv70e12.r1 Stratagene	3.57e-03
c	46	17	19.5	466	19	N68485	za14g04.s1 Homo sapien	3.57e-03
c	47	17	19.5	500	14	AA836628	od66a12.s1 NCI_CGAP_GC	3.57e-03
	48	17	19.5	504	28	AQ186050	HS_3070_A2_H05_MF CIT	3.57e-03
c	49	17	19.5	569	26	FR0014332	F.rubripes GSS sequenc	3.57e-03
	50	17	19.5	576	26	FR0014311	F.rubripes GSS sequenc	3.57e-03
	51	17	19.5	584	26	FR0010285	F.rubripes GSS sequenc	3.57e-03
c	52	17	19.5	605	26	AQ017518	CIT-HSP-2306H11.TF CIT	3.57e-03
c	53	16	18.4	107	7	AA480699	SWMFCA20468K Brugia ma	7.21e-02
	54	16	18.4	108	10	AA639719	nq93c03.s1 NCI_CGAP_Co	7.21e-02
c	55	16	18.4	119	10	AA600712	np43c01.s1 NCI_CGAP_Br	7.21e-02
	56	16	18.4	157	16	R80676	yi93h07.s1 Homo sapien	7.21e-02
	57	16	18.4	157	7	MUSGS01548	Mouse 3'-directed cDN	7.21e-02
	58	16	18.4	172	12	AA807787	nu88f02.s1 NCI_CGAP_Al	7.21e-02
	59	16	18.4	200	15	AI059775	UI-R-C1-lh-c-11-0-UI.s	7.21e-02
	60	16	18.4	206	22	AA150007	zo03g02.s1 Stratagene	7.21e-02
	61	16	18.4	207	19	N34502	yy55f04.s1 Homo sapien	7.21e-02
c	62	16	18.4	211	22	AA089475	ccp0910.seq.F Fetal he	7.21e-02
c	63	16	18.4	229	20	HSC0KH091	H. sapiens partial cDN	7.21e-02
c	64	16	18.4	252	25	AA362879	EST72654 Ovary II Homo	7.21e-02
	65	16	18.4	282	17	AA931727	oo77c05.s1 NCI_CGAP_Ki	7.21e-02
	66	16	18.4	284	14	AA887396	oj37b04.s1 NCI_CGAP_Lu	7.21e-02
c	67	16	18.4	285	7	AA443786	zw86e02.s1 Soares tota	7.21e-02
	68	16	18.4	287	19	HSC24H062	H. sapiens partial cDN	7.21e-02
	69	16	18.4	298	19	HSC2NE111	H. sapiens partial cDN	7.21e-02
	70	16	18.4	322	15	AI057734	TENU1824 T. cruzi epim	7.21e-02
	71	16	18.4	327	5	R37160	yf65f12.s1 Homo sapien	7.21e-02
c	72	16	18.4	330	15	AA933922	on91d12.s1 Soares_NFL_	7.21e-02
	73	16	18.4	343	15	AI059479	UI-R-C1-lg-d-07-0-UI.s	7.21e-02
c	74	16	18.4	348	7	AA460506	zx61h11.r1 Soares tota	7.21e-02
	75	16	18.4	349	13	AA883771	al59a07.s1 Soares NFL	7.21e-02
c	76	16	18.4	352	5	T98251	ye59b01.s1 Homo sapien	7.21e-02
	77	16	18.4	358	13	AA875219	UI-R-E0-cu-g-11-0-UI.s	7.21e-02
c	78	16	18.4	367	28	AQ110633	CIT-HSP-2373P13.TF CIT	7.21e-02
c	79	16	18.4	368	10	AA602936	np02g11.s1 NCI_CGAP_Pr	7.21e-02
c	80	16	18.4	370	16	R70372	yj82b04.s1 Homo sapien	7.21e-02
	81	16	18.4	377	15	AI027013	ow41f06.s1 Soares_para	7.21e-02
c	82	16	18.4	379	28	AQ164755	HS_3015_B1_A11_MR CIT	7.21e-02
c	83	16	18.4	384	21	W42734	zc24e01.s1 Soares sene	7.21e-02
	84	16	18.4	389	8	AA508674	ni22a09.s1 NCI_CGAP_Co	7.21e-02
c	85	16	18.4	392	15	AA945857	EST201356 Normalized r	7.21e-02
	86	16	18.4	393	13	AA874862	UI-R-E0-ci-b-10-0-UI.s	7.21e-02
c	87	16	18.4	397	28	AQ171330	HS_3071_B2_E07_MR CIT	7.21e-02
c	88	16	18.4	400	14	AA909368	oj92f06.s2 Soares_NFL_	7.21e-02
	89	16	18.4	402	18	AI185867	qe45a03.x1 Soares_feta	7.21e-02
	90	16	18.4	407	27	FR0034114	Fugu rubripes GSS sequ	7.21e-02
	91	16	18.4	407	9	AA514860	ni08h01.s1 NCI_CGAP_Br	7.21e-02
	92	16	18.4	407	9	AA579969	ng51c04.s1 NCI_CGAP_Co	7.21e-02
c	93	16	18.4	407	13	AA859490	UI-R-E0-bv-g-02-0-UI.s	7.21e-02

	94	16	18.4	411	10	AA428265	zv84g12.s1 Soares tota	7.21e-02
c	95	16	18.4	413	5	R28368	yh55f11.s1 Homo sapien	7.21e-02
c	96	16	18.4	419	14	AA977890	oq62a06.s1 NCI_CGAP_Ki	7.21e-02
c	97	16	18.4	419	16	H40304	yn86f01.s1 Homo sapien	7.21e-02
	98	16	18.4	422	6	AA410364	zv16a03.s1 Soares NhHM	7.21e-02
c	99	16	18.4	423	28	AQ178971	HS_3071_B1_D02_MR CIT	7.21e-02
c	100	16	18.4	430	17	AA410090	EST01942 Mouse 7.5 dpc	7.21e-02
	101	16	18.4	435	6	AA431212	zw76h08.s1 Soares test	7.21e-02
	102	16	18.4	436	8	AA445761	vc63b01.s1 Knowles Sol	7.21e-02
c	103	16	18.4	440	9	AA570861	SWMFCA2626SK Brugia ma	7.21e-02
	104	16	18.4	441	26	B96643	T30P13TR TAMU Arabidop	7.21e-02
	105	16	18.4	442	6	AA150135	zl42f02.s1 Soares preg	7.21e-02
	106	16	18.4	446	16	H23166	ym56f12.s1 Homo sapien	7.21e-02
c	107	16	18.4	458	27	AQ035346	CIT-HSP-2332D14.TF CIT	7.21e-02
	108	16	18.4	460	20	W78367	me85e11.r1 Soares mous	7.21e-02
c	109	16	18.4	460	6	AA406186	zv16a03.r1 Soares NhHM	7.21e-02
c	110	16	18.4	465	19	H97394	yw10b04.s1 Homo sapien	7.21e-02
c	111	16	18.4	469	15	AA917766	on38d06.s1 NCI_CGAP_Co	7.21e-02
c	112	16	18.4	471	19	N34495	yy13h08.s1 Homo sapien	7.21e-02
	113	16	18.4	477	16	H23181	ym56h12.s1 Homo sapien	7.21e-02
	114	16	18.4	478	14	AA900054	UI-R-E0-da-e-08-0-UI.s	7.21e-02
c	115	16	18.4	479	14	AA899905	UI-R-E0-dh-c-07-0-UI.s	7.21e-02
c	116	16	18.4	480	28	AQ186412	HS_3094_B1_F07_MF CIT	7.21e-02
c	117	16	18.4	483	28	AQ127712	HS_3091_B2_C10_MF CIT	7.21e-02
	118	16	18.4	494	5	AA992595	oc97g05.s1 Soares_tota	7.21e-02
c	119	16	18.4	494	16	R80470	yi93g01.r1 Homo sapien	7.21e-02
	120	16	18.4	495	16	H05730	yl73a08.s1 Homo sapien	7.21e-02
c	121	16	18.4	498	12	AA817950	UI-R-A0-ag-e-10-0-UI.s	7.21e-02
	122	16	18.4	494	29	AQ142775	HS_2230_B1_E05_MR CIT	7.21e-02
c	123	16	18.4	494	18	AI146115	UI-R-BT0-qh-b-06-0-UI.	7.21e-02
	124	16	18.4	498	13	AA485958	ab40h04.s1 Stratagene	7.21e-02
c	125	16	18.4	503	11	N21687	yx63h03.s1 Soares mela	7.21e-02
c	126	16	18.4	510	14	AA899502	UI-R-E0-bu-a-09-0-UI.s	7.21e-02
c	127	16	18.4	511	8	AA214529	zr92a04.s1 NCI_CGAP_GC	7.21e-02
c	128	16	18.4	514	5	T83918	yd67g08.r1 Homo sapien	7.21e-02
c	129	16	18.4	514	28	AQ209065	HS_3228_A1_A06_T7 CIT	7.21e-02
	130	16	18.4	516	14	AA892356	EST196159 Normalized r	7.21e-02
c	131	16	18.4	516	26	B14227	A-142A6.TP CIT978SKA1	7.21e-02
c	132	16	18.4	518	18	AI160460	qc08h10.x1 Soares_feta	7.21e-02
	133	16	18.4	531	17	AU023818	Mus musculus unfertili	7.21e-02
c	134	16	18.4	531	15	AA943816	EST199315 Normalized r	7.21e-02
	135	16	18.4	540	6	AA393368	zt75a07.r1 Soares test	7.21e-02
c	136	16	18.4	543	12	AA769319	nz39g04.s1 NCI_CGAP_GC	7.21e-02
c	137	16	18.4	545	28	AQ171608	HS_3168_B1_A09_T7 CIT	7.21e-02
c	138	16	18.4	551	18	AI092205	qa60b04.s1 Soares_feta	7.21e-02
	139	16	18.4	570	26	B72805	RPCI11-10F17.TV RPCI11	7.21e-02
c	140	16	18.4	577	19	N21426	yx58d01.s1 Homo sapien	7.21e-02
c	141	16	18.4	591	9	AA521439	aa69c10.s1 NCI_CGAP_GC	7.21e-02
	142	16	18.4	608	15	AI009141	EST203592 Normalized r	7.21e-02
	143	16	18.4	619	26	FR0004273	F.rubripes GSS sequenc	7.21e-02
c	144	16	18.4	619	26	FR0004291	F.rubripes GSS sequenc	7.21e-02

	145	16	18.4	621	19	N28769	yx69e10.r1 Homo sapien	7.21e-02
c	146	16	18.4	671	26	AQ022360	CIT-HSP-2313N24.TR CIT	7.21e-02
	147	16	18.4	695	28	AQ157362	nbxb0009G22r CUGI Rice	7.21e-02
c	148	16	18.4	697	15	AA932917	om83h04.s1 NCI_CGAP_Ki	7.21e-02
c	149	16	18.4	700	27	AG011813	Homo sapiens genomic D	7.21e-02
c	150	16	18.4	705	27	AG011814	Homo sapiens genomic D	7.21e-02
c	151	16	18.4	707	27	AG011812	Homo sapiens genomic D	7.21e-02
c	152	16	18.4	710	27	AG011815	Homo sapiens genomic D	7.21e-02
c	153	16	18.4	749	27	B69329	CIT-HSP-2053N15.TF CIT	7.21e-02
	154	16	18.4	772	29	AQ146201	HS_2231_A1_H08_MR CIT	7.21e-02
c	155	16	18.4	847	15	AA991126	BSBmMFSZ06B18SK Brugia	7.21e-02
c	156	16	18.4	864	15	AA991127	BSBmMFSZ06B12SK Brugia	7.21e-02
	157	16	18.4	1179	26	B09578	F21B24-Sp6 IGF Arabido	7.21e-02
	158	16	18.4	1187	26	B11467	F13J13-T7 IGF Arabidop	7.21e-02
c	159	15	17.2	64	13	AA843971	ai91a03.s1 Soares para	1.26e+00
c	160	15	17.2	158	20	N75195	yw33c08.r1 Homo sapien	1.26e+00
	161	15	17.2	182	27	B70508	CIT-HSP-2059H22.TR CIT	1.26e+00
c	162	15	17.2	182	10	AA587936	nm88e05.s1 NCI_CGAP_Co	1.26e+00
c	163	15	17.2	187	5	T10473	hbc330 Homo sapiens cD	1.26e+00
c	164	15	17.2	194	8	AA136464	zn94b07.r1 Stratagene	1.26e+00
	165	15	17.2	197	23	AA224584	zr17c03.s1 Stratagene	1.26e+00
c	166	15	17.2	208	19	N62544	za12b12.s1 Homo sapien	1.26e+00
	167	15	17.2	209	23	AA249505	j9773.seq.F Human feta	1.26e+00
c	168	15	17.2	220	8	AA255622	za31d07.s1 NCI_CGAP_GC	1.26e+00
c	169	15	17.2	223	27	AQ005391	CIT-HSP-2295N4.TF CIT-	1.26e+00
c	170	15	17.2	229	19	N71355	za30h08.s1 Homo sapien	1.26e+00
c	171	15	17.2	240	8	H05528	yl70g01.s1 Soares infa	1.26e+00
c	172	15	17.2	253	22	AA055047	zf19a08.s1 Soares feta	1.26e+00
	173	15	17.2	256	28	AQ096694	HS_3032_B1_G10_MF CIT	1.26e+00
c	174	15	17.2	256	15	AI063683	GH03633.5prime GH Dros	1.26e+00
c	175	15	17.2	258	10	T21342	3350 Lambda-PRL2 Arabi	1.26e+00
c	176	15	17.2	259	5	T09848	0424m7 Plasmodium falc	1.26e+00
	177	15	17.2	261	5	T06907	EST04796 Homo sapiens	1.26e+00
c	178	15	17.2	263	20	HSC36C072	H. sapiens partial cDN	1.26e+00
c	179	15	17.2	264	17	AI116283	uc17h10.r1 Soares mous	1.26e+00
c	180	15	17.2	264	14	AA932875	oo57d10.s1 NCI_CGAP_Lu	1.26e+00
c	181	15	17.2	267	8	AA507611	ng87g01.s1 NCI_CGAP_Pr	1.26e+00
c	182	15	17.2	268	16	H16106	yl28a04.r1 Homo sapien	1.26e+00
c	183	15	17.2	273	8	AA513283	nh85a03.s1 NCI_CGAP_Br	1.26e+00
c	184	15	17.2	276	19	N71405	za31h08.s1 Homo sapien	1.26e+00
	185	15	17.2	278	5	T09585	0178m7 Plasmodium falc	1.26e+00
c	186	15	17.2	280	14	AA872622	oi06g11.s1 NCI_CGAP_GC	1.26e+00
c	187	15	17.2	280	13	AA868895	ak55b04.s1 Soares test	1.26e+00
c	188	15	17.2	281	5	T75574	yd63d12.s1 Homo sapien	1.26e+00
c	189	15	17.2	282	26	B33914	HS-1023-B1-C04-MR.abi	1.26e+00
	190	15	17.2	285	27	AQ052152	RPCI11-50K4.TJ RPCI11	1.26e+00
c	191	15	17.2	287	5	R34083	yh79d12.r1 Homo sapien	1.26e+00
	192	15	17.2	288	25	AA360828	EST70041 T-cell lympho	1.26e+00
c	193	15	17.2	288	7	HUM225B12B	Human aorta cDNA 5'-en	1.26e+00
c	194	15	17.2	290	16	R43182	yg18d07.s1 Homo sapien	1.26e+00
c	195	15	17.2	292	14	AA991861	os67g09.s1 NCI_CGAP_GC	1.26e+00

c	196	15	17.2	293	13	AA644265	af63d01.s1 Soares NhHM	1.26e+00
c	197	15	17.2	297	12	AA771867	ai32h02.s1 Soares para	1.26e+00
	198	15	17.2	299	23	W35430	91 Mouse VM cDNA libra	1.26e+00
c	199	15	17.2	304	16	H22443	yn69c01.s1 Homo sapien	1.26e+00
	200	15	17.2	305	6	C01401	HUMGS0008383, Human Ge	1.26e+00
c	201	15	17.2	308	17	AI123985	qa99c11.s1 Soares_preg	1.26e+00
c	202	15	17.2	314	19	N29475	yy05h12.s1 Homo sapien	1.26e+00
	203	15	17.2	317	5	T80120	yd17c06.s1 Homo sapien	1.26e+00
c	204	15	17.2	317	11	AA716136	sg60c06.s1 Soares feta	1.26e+00
c	205	15	17.2	321	17	AI014718	ou28h01.x1 Soares_NFL_	1.26e+00
	206	15	17.2	323	16	T33470	EST57984 Homo sapiens	1.26e+00
c	207	15	17.2	325	14	AA832516	oe62g01.s1 NCI_CGAP_Lu	1.26e+00
c	208	15	17.2	326	5	T77866	yd20b02.s1 Homo sapien	1.26e+00
c	209	15	17.2	326	17	AI081602	ow77a04.s1 Soares_feta	1.26e+00
	210	15	17.2	327	25	AA332708	EST36690 Embryo, 8 wee	1.26e+00
	211	15	17.2	328	19	H62291	yz39h09.r1 Homo sapien	1.26e+00
	212	15	17.2	329	13	AA842265	MBAPCW5D01T3 Brugia ma	1.26e+00
	213	15	17.2	330	26	B39511	HS-1049-B2-E05-MR.abi	1.26e+00
	214	15	17.2	331	25	AA338630	EST43668 Fetal brain I	1.26e+00
c	215	15	17.2	332	23	T16824	NIB1904 Normalized inf	1.26e+00
	216	15	17.2	333	10	AA625013	SWOv3MCA3505K Onchocer	1.26e+00
	217	15	17.2	334	16	T29740	EST92658 Homo sapiens	1.26e+00
c	218	15	17.2	337	16	H40846	yp49e12.r1 Homo sapien	1.26e+00
c	219	15	17.2	338	16	T33679	EST58716 Homo sapiens	1.26e+00
	220	15	17.2	339	6	C18959	Human placenta cDNA 5'	1.26e+00
c	221	15	17.2	340	15	AA921980	om15g03.s1 Soares_NFL_	1.26e+00
	222	15	17.2	342	27	AQ034780	CIT-HSP-2319022.TR CIT	1.26e+00
c	223	15	17.2	343	10	AA428625	zw75h12.s1 Soares test	1.26e+00
c	224	15	17.2	343	9	AA576838	nm67f08.s1 NCI_CGAP_Br	1.26e+00
c	225	15	17.2	343	17	AI033659	ov22f12.x1 Soares_para	1.26e+00
c	226	15	17.2	344	21	W00752	T1230 MVAT4 bloodstrea	1.26e+00
c	227	15	17.2	344	16	R62210	yi17e07.r1 Homo sapien	1.26e+00
	228	15	17.2	345	5	R52920	yi02b04.s1 Homo sapien	1.26e+00
	229	15	17.2	345	6	C19097	Human placenta cDNA 5'	1.26e+00
	230	15	17.2	348	29	AQ211155	HS_2047_B2_A08_T7 CIT	1.26e+00
c	231	15	17.2	348	8	AA482173	aa53e02.r1 NCI_CGAP_GC	1.26e+00
c	232	15	17.2	348	15	AA910970	ok67b09.s1 NCI_CGAP_GC	1.26e+00
	233	15	17.2	349	23	AA152937	mr16a08.r1 Soares mous	1.26e+00
c	234	15	17.2	353	16	T29419	EST79555 Homo sapiens	1.26e+00
c	235	15	17.2	354	18	AI185431	qe53c09.x1 Soares_feta	1.26e+00
	236	15	17.2	355	16	H22488	yn69d08.r1 Homo sapien	1.26e+00
c	237	15	17.2	355	27	AQ027671	CIT-HSP-2310K24.TF CIT	1.26e+00
c	238	15	17.2	356	13	AA814435	nz08h08.s1 NCI_CGAP_GC	1.26e+00
c	239	15	17.2	356	16	R76839	yi63b11.r1 Homo sapien	1.26e+00
c	240	15	17.2	357	12	AA782157	ai48g09.s1 Soares para	1.26e+00
c	241	15	17.2	357	12	AA759126	ah78e01.s1 Soares test	1.26e+00
c	242	15	17.2	358	21	R93943	yt72c12.r1 Soares reti	1.26e+00
c	243	15	17.2	358	21	W42735	zc24e02.s1 Soares sene	1.26e+00
c	244	15	17.2	358	17	AI027204	ov83b06.x1 Soares_test	1.26e+00
	245	15	17.2	359	9	AA525344	ni53f08.s1 NCI_CGAP_Ov	1.26e+00
c	246	15	17.2	359	27	AQ087053	HS_2188_B2_C07_MR CIT	1.26e+00

	247	15	17.2	360	27	AQ0000046	CIT-HSP-2172N2.TR CIT-	1.26e+00
c	248	15	17.2	360	9	C40566	C.elegans cDNA clone y	1.26e+00
c	249	15	17.2	360	9	C44530	C.elegans cDNA clone y	1.26e+00
c	250	15	17.2	360	7	CELK075G5F	C.elegans cDNA clone y	1.26e+00
c	251	15	17.2	362	5	R50260	yj58c08.r1 Homo sapien	1.26e+00
c	252	15	17.2	362	28	AQ102332	HS_3040_A1_C06_MF CIT	1.26e+00
	253	15	17.2	363	14	AA447335	32.A10R PN001-Normal H	1.26e+00
	254	15	17.2	363	17	AI089351	qb05g06.x1 Soares_preg	1.26e+00
c	255	15	17.2	363	29	AQ137693	HS_3065_B2_C11_MR CIT	1.26e+00
c	256	15	17.2	365	22	AA121591	sl88g01.r1 Stratagene	1.26e+00
	257	15	17.2	367	5	T50614	yb30h05.r1 Homo sapien	1.26e+00
	258	15	17.2	368	26	B30985	HS-1004-A1-G06-MR.abi	1.26e+00
	259	15	17.2	369	11	AA691515	va12c08.r1 Barstead mo	1.26e+00
c	260	15	17.2	371	28	AQ103729	HS_3092_B1_D09_MF CIT	1.26e+00
c	261	15	17.2	372	27	AQ039512	CIT-HSP-2316P8.TR CIT-	1.26e+00
c	262	15	17.2	373	16	R68873	yi37e01.r1 Homo sapien	1.26e+00
	263	15	17.2	374	16	R91773	yp98d10.r1 Homo sapien	1.26e+00
c	264	15	17.2	374	14	AA973664	oo48f02.s1 NCI_CGAP_Lu	1.26e+00
c	265	15	17.2	374	21	W37312	sc10a03.s1 Soares para	1.26e+00
c	266	15	17.2	374	7	HUM110C06A	Human fetal brain cDNA	1.26e+00
	267	15	17.2	375	26	B35355	HS-1028-B1-D01-MR.abi	1.26e+00
c	268	15	17.2	375	21	AA047086	zk78e09.s1 Soares preg	1.26e+00
c	269	15	17.2	377	10	W09388	ma63e05.r1 Soares mous	1.26e+00
c	270	15	17.2	378	16	H03147	yj44d08.r1 Homo sapien	1.26e+00
	271	15	17.2	380	10	AA654478	nt62g05.s1 NCI_CGAP_Pr	1.26e+00
c	272	15	17.2	380	19	H70914	yu70c01.s1 Homo sapien	1.26e+00
c	273	15	17.2	380	5	R09395	yf22c12.s1 Homo sapien	1.26e+00
c	274	15	17.2	380	12	AA776371	ah16e05.s1 Gessler Wil	1.26e+00
c	275	15	17.2	381	8	AA232902	zr45h09.r1 Soares NhHM	1.26e+00
c	276	15	17.2	381	8	AA279815	zs92c03.s1 NCI_CGAP_GC	1.26e+00
c	277	15	17.2	382	17	AU023586	Mus musculus unfertili	1.26e+00
c	278	15	17.2	383	28	AQ191874	HS_3233_B1_H04_MR CIT	1.26e+00
c	279	15	17.2	383	28	AQ181096	HS_3218_A1_B12_MR CIT	1.26e+00
c	280	15	17.2	384	20	N95739	yy60f08.r1 Homo sapien	1.26e+00
c	281	15	17.2	385	14	AA829918	oe51h03.s1 NCI_CGAP_Lu	1.26e+00
c	282	15	17.2	386	8	AA504545	aa60c06.r1 NCI_CGAP_GC	1.26e+00
c	283	15	17.2	386	5	T36801	EST101807 Saccharomyce	1.26e+00
c	284	15	17.2	386	17	AA988657	os08d10.s1 NCI_CGAP_Lu	1.26e+00
	285	15	17.2	387	23	AA105944	ml80g04.r1 Stratagene	1.26e+00
	286	15	17.2	387	19	H69958	yu72e11.s1 Homo sapien	1.26e+00
c	287	15	17.2	388	11	AA074513	zm17f03.s1 Stratagene	1.26e+00
c	288	15	17.2	389	18	AI139689	qc20b12.x1 Soares_feta	1.26e+00
c	289	15	17.2	389	20	N89939	zb23f12.s1 Soares feta	1.26e+00
	290	15	17.2	390	27	AQ057487	CIT-HSP-2340J16.TF CIT	1.26e+00
c	291	15	17.2	390	16	H39711	yo52h03.r1 Homo sapien	1.26e+00
	292	15	17.2	391	27	AQ045649	RPCI11-34J20.TJ RPCI11	1.26e+00
c	293	15	17.2	392	20	W05318	za84e01.r1 Soares feta	1.26e+00
c	294	15	17.2	393	7	AA448667	zx10e07.s1 Soares tota	1.26e+00
	295	15	17.2	394	24	AA326510	EST29695 Cerebellum II	1.26e+00
	296	15	17.2	395	16	T32949	EST56026 Homo sapiens	1.26e+00
c	297	15	17.2	396	22	W92526	ze01b01.s1 Soares feta	1.26e+00

c	298	15	17.2	396	13	AA829389	od06c03.s1 NCI_CGAP_GC	1.26e+00
	299	15	17.2	397	6	C18651	Human placenta cDNA 5'	1.26e+00
	300	15	17.2	397	15	AA898337	EST198140 Normalized r	1.26e+00
c	301	15	17.2	397	17	AI128070	qc47d02.x1 Soares_preg	1.26e+00
c	302	15	17.2	397	10	AA612574	no46c10.s1 NCI_CGAP_Pr	1.26e+00
c	303	15	17.2	397	8	AA243819	zr67g05.r1 Soares NhHM	1.26e+00
c	304	15	17.2	399	20	ATTS2963	A. thaliana transcribe	1.26e+00
	305	15	17.2	400	21	W61216	zd31f05.r1 Soares feta	1.26e+00
c	306	15	17.2	400	26	B44334	HS-1059-A1-G08-MF.abi	1.26e+00
c	307	15	17.2	400	27	AQ071067	HS_3013_A2_E09_T7 CIT	1.26e+00
c	308	15	17.2	401	28	AQ134752	HS_3052_B1_G06_MF CIT	1.26e+00
c	309	15	17.2	402	6	AA007438	zh98b12.r1 Soares feta	1.26e+00
c	310	15	17.2	402	16	H20065	yn60f01.s1 Homo sapien	1.26e+00
	311	15	17.2	403	19	H81340	yu74d04.r1 Homo sapien	1.26e+00
	312	15	17.2	404	20	W19366	zb22f05.r1 Soares feta	1.26e+00
	313	15	17.2	404	7	HUMGS01730	Human HL60 3'directed	1.26e+00
	314	15	17.2	405	28	AQ191608	HS_3225_B1_A10_MR CIT	1.26e+00
c	315	15	17.2	405	5	R30884	yh59b01.r1 Homo sapien	1.26e+00
c	316	15	17.2	405	7	AA447594	zw82c02.s1 Soares test	1.26e+00
c	317	15	17.2	405	17	AI032613	ow17f03.x1 Soares_para	1.26e+00
c	318	15	17.2	406	14	AA829611	oe50b08.s1 NCI_CGAP_Lu	1.26e+00
	319	15	17.2	407	13	AA868330	ak40h03.s1 Soares test	1.26e+00
c	320	15	17.2	407	11	AA668567	ac44e08.s1 Stratagene	1.26e+00
c	321	15	17.2	408	29	AQ230599	HS_2033_A1_H11_MR CIT	1.26e+00
	322	15	17.2	410	6	C18245	Human placenta cDNA 5'	1.26e+00
	323	15	17.2	410	26	AQ022149	CIT-HSP-2308A18.TR CIT	1.26e+00
c	324	15	17.2	413	5	R52540	yg81f04.s1 Homo sapien	1.26e+00
	325	15	17.2	414	28	AQ220133	HS_3248_B1_F05_MR CIT	1.26e+00
c	326	15	17.2	415	21	W73365	zd53b08.s1 Soares feta	1.26e+00
c	327	15	17.2	415	5	R42139	yf99a02.s1 Homo sapien	1.26e+00
	328	15	17.2	416	16	H03956	yj44d08.s1 Homo sapien	1.26e+00
c	329	15	17.2	417	21	W73919	zd75b06.s1 Soares feta	1.26e+00
c	330	15	17.2	417	22	N62219	yz63b04.s1 Soares mult	1.26e+00
	331	15	17.2	418	13	AA827847	od01g09.s1 NCI_CGAP_GC	1.26e+00
	332	15	17.2	418	28	AQ112778	CIT-HSP-2375F15.TR CIT	1.26e+00
c	333	15	17.2	418	19	N73001	za07f06.s1 Homo sapien	1.26e+00
	334	15	17.2	419	6	AA135957	zl22h09.r1 Soares preg	1.26e+00
	335	15	17.2	420	17	T45433	8696 Lambda-PRL2 Arabi	1.26e+00
	336	15	17.2	420	11	AA682470	zj18c01.s1 Soares feta	1.26e+00
	337	15	17.2	420	27	AQ038443	CIT-HSP-2336G3.TF CIT-	1.26e+00
	338	15	17.2	421	6	C18553	Human placenta cDNA 5'	1.26e+00
c	339	15	17.2	421	17	AI090602	qa70g02.x1 Soares_feta	1.26e+00
c	340	15	17.2	421	7	HUM082A11A	Human fetal brain cDNA	1.26e+00
c	341	15	17.2	423	18	AI140940	qa83g11.s1 Soares_feta	1.26e+00
c	342	15	17.2	424	11	AA680133	zi39c11.s1 Soares feta	1.26e+00
c	343	15	17.2	424	15	AA987945	or93e08.s1 NCI_CGAP_Lu	1.26e+00
c	344	15	17.2	426	10	AA614830	nn90c05.s1 NCI_CGAP_Br	1.26e+00
c	345	15	17.2	426	7	AA509447	vy30g04.r1 Soares mous	1.26e+00
c	346	15	17.2	429	27	B65993	CIT-HSP-2023K24.TF CIT	1.26e+00
c	347	15	17.2	430	13	AA824280	aj29g04.s1 Soares test	1.26e+00
c	348	15	17.2	432	19	N58940	yy62f11.s1 Homo sapien	1.26e+00

	349	15	17.2	436	8	AA128426	zn60c10.r1 Stratagene	1.26e+00
	350	15	17.2	436	27	AQ065179	HS_2180_B2_E08_MR CIT	1.26e+00
	351	15	17.2	437	26	FR0012669	F.rubripes GSS sequenc	1.26e+00
c	352	15	17.2	437	10	AA640396	nt69h09.s1 NCI_CGAP_Pr	1.26e+00
	353	15	17.2	440	27	FR0034132	Fugu rubripes GSS sequ	1.26e+00
	354	15	17.2	441	22	AA164499	zo91g09.r1 Stratagene	1.26e+00
c	355	15	17.2	441	16	R40903	yf76h02.s1 Homo sapien	1.26e+00
c	356	15	17.2	441	20	N89752	zb14d03.s1 Soares feta	1.26e+00
c	357	15	17.2	441	6	C06268	similar to none, mRNA	1.26e+00
	358	15	17.2	442	8	AA484049	ne77c11.s1 NCI_CGAP_Ew	1.26e+00
	359	15	17.2	442	23	AA233782	zr44c11.r1 Soares NhHM	1.26e+00
c	360	15	17.2	442	20	N92595	zb27c12.s1 Homo sapien	1.26e+00
c	361	15	17.2	442	16	R80540	yi97h05.r1 Homo sapien	1.26e+00
	362	15	17.2	443	18	AI182048	ud73d07.y1 Sugano mous	1.26e+00
c	363	15	17.2	443	20	N89652	zb10h01.s1 Soares feta	1.26e+00
c	364	15	17.2	443	18	AI090593	qa70f04.x1 Soares_feta	1.26e+00
	365	15	17.2	444	29	AQ254143	HS_2243_A1_H05_MR CIT	1.26e+00
c	366	15	17.2	445	22	AA164675	zo93a10.s1 Stratagene	1.26e+00
c	367	15	17.2	445	12	AA767121	oa64e02.s1 NCI_CGAP_GC	1.26e+00
c	368	15	17.2	445	21	W60969	zd31f05.s1 Soares feta	1.26e+00
	369	15	17.2	448	18	AI167259	oo08h02.x1 Soares_NSF_	1.26e+00
c	370	15	17.2	448	29	AQ245627	HS_2064_A1_B09_T7 CIT	1.26e+00
c	371	15	17.2	450	17	AI126497	qd82h11.x1 Soares_test	1.26e+00
c	372	15	17.2	450	18	AI149725	qf32b07.x1 Soares_test	1.26e+00
	373	15	17.2	452	6	AA125893	zk92h09.r1 Soares preg	1.26e+00
c	374	15	17.2	452	20	N80035	za91d03.s1 Homo sapien	1.26e+00
c	375	15	17.2	453	15	AA972418	op78b03.s1 Soares_NFL_	1.26e+00
c	376	15	17.2	453	17	AI085309	qf18g11.x1 NCI_CGAP_Br	1.26e+00
c	377	15	17.2	454	6	AA099192	zk84b05.r1 Soares preg	1.26e+00
c	378	15	17.2	454	16	H26571	yl13d12.s1 Homo sapien	1.26e+00
c	379	15	17.2	454	5	R26848	yh52a08.s1 Homo sapien	1.26e+00
c	380	15	17.2	457	19	N25209	yy03b11.s1 Homo sapien	1.26e+00
c	381	15	17.2	457	19	N39243	yy50e03.s1 Homo sapien	1.26e+00
	382	15	17.2	461	22	W96518	ze11h04.r1 Soares feta	1.26e+00
	383	15	17.2	461	26	AQ017983	CIT-HSP-2302A4.TF CIT-	1.26e+00
c	384	15	17.2	461	9	AA553431	nk77c06.s1 NCI_CGAP_Sc	1.26e+00
	385	15	17.2	462	8	AA525895	ni57g12.s1 NCI_CGAP_Ov	1.26e+00
	386	15	17.2	462	26	B32980	HS-1016-A2-C03-MF.abi	1.26e+00
c	387	15	17.2	462	20	W44743	zb98a09.s1 Soares para	1.26e+00
c	388	15	17.2	462	26	AQ014487	CIT-HSP-2300I2.TR CIT-	1.26e+00
c	389	15	17.2	462	22	W96485	ze11h04.s1 Soares feta	1.26e+00
	390	15	17.2	463	15	AI049183	ub38g04.r1 Soares 2NbM	1.26e+00
c	391	15	17.2	463	17	AI025747	ow11e07.s1 Soares_para	1.26e+00
c	392	15	17.2	463	17	AI023814	ov70b04.s1 Soares_test	1.26e+00
c	393	15	17.2	465	9	AA515824	ng64d01.s1 NCI_CGAP_Li	1.26e+00
c	394	15	17.2	466	21	AA046813	zf14f09.s1 Soares feta	1.26e+00
c	395	15	17.2	466	29	AQ247574	HS_2055_B2_A06_T7 CIT	1.26e+00
	396	15	17.2	467	14	AA847332	oe18c09.s1 NCI_CGAP_Ov	1.26e+00
c	397	15	17.2	468	16	H15009	yl26d10.s1 Homo sapien	1.26e+00
	398	15	17.2	469	28	AQ167868	HS_2224_B2_D10_MR CIT	1.26e+00
	399	15	17.2	470	26	AQ015742	CIT-HSP-2307I6.TF CIT-	1.26e+00

c	400	15	17.2	470	22	AA053964	ze75h01.s1 Soares feta	1.26e+00
c	401	15	17.2	471	22	H96547	yw01b09.s1 Soares mela	1.26e+00
	402	15	17.2	472	5	T57623	yb59b10.s1 Homo sapien	1.26e+00
c	403	15	17.2	472	11	AA112180	zm64h08.s1 Stratagene	1.26e+00
	404	15	17.2	473	7	AA497456	vh28f04.r1 Soares mous	1.26e+00
c	405	15	17.2	473	5	T81926	yd29g11.s1 Homo sapien	1.26e+00
	406	15	17.2	475	6	AA099593	zk86e02.s1 Soares preg	1.26e+00
c	407	15	17.2	475	14	AA872922	oh71h09.s1 NCI_CGAP_K1	1.26e+00
c	408	15	17.2	475	26	B42127	HS-1055-A2-E03-MF-V000	1.26e+00
c	409	15	17.2	475	18	AI079401	oz64c12.x1 Soares_sene	1.26e+00
	410	15	17.2	476	12	AA762284	vw60h04.r1 Soares mous	1.26e+00
c	411	15	17.2	477	29	AQ229468	HS_2023_A2_F12_T7 CIT	1.26e+00
	412	15	17.2	478	26	B49948	CIT-HSP-553C2.TP CIT-H	1.26e+00
	413	15	17.2	478	27	B93933	CIT-HSP-2163M5.TR CIT-	1.26e+00
	414	15	17.2	479	27	AQ046353	RPCI11-34M4.TK RPCI11	1.26e+00
	415	15	17.2	479	21	W45210	zc24g12.r1 Soares sene	1.26e+00
c	416	15	17.2	480	16	R55324	yg88a08.s1 Homo sapien	1.26e+00
c	417	15	17.2	481	11	AA679013	zi37a07.r1 Soares feta	1.26e+00
c	418	15	17.2	481	21	W46945	zc40c04.s1 Soares sene	1.26e+00
c	419	15	17.2	482	8	AA528002	ne61d10.s1 NCI_CGAP_Co	1.26e+00
c	420	15	17.2	482	13	AA190836	zp85d04.s1 Stratagene	1.26e+00
	421	15	17.2	483	28	AQ176339	HS_3213_A1_C04_MR CIT	1.26e+00
	422	15	17.2	484	28	AQ134711	HS_3052_B1_C09_MF CIT	1.26e+00
c	423	15	17.2	485	17	AI033026	ow93b05.s1 Soares_feta	1.26e+00
	424	15	17.2	486	28	AQ155123	HS_3039_A2_C10_MF CIT	1.26e+00
	425	15	17.2	487	16	H02779	yj41a10.r1 Homo sapien	1.26e+00
c	426	15	17.2	487	6	AA437016	zv58g03.r1 Soares test	1.26e+00
c	427	15	17.2	490	15	AA997742	UI-R-C0-hn-d-07-0-UI.s	1.26e+00
c	428	15	17.2	492	12	AA782591	ai57e12.s1 Soares para	1.26e+00
	429	15	17.2	494	27	AQ034779	CIT-HSP-2320B22.TF CIT	1.26e+00
c	430	15	17.2	494	18	AI150467	qf41c10.x1 Soares_test	1.26e+00
	431	15	17.2	495	15	C94127	Dictyostelium discoide	1.26e+00
	432	15	17.2	495	29	AQ246324	HS_2059_B2_D11_MR CIT	1.26e+00
	433	15	17.2	497	23	AA217098	mu91e07.r1 Soares mous	1.26e+00
c	434	15	17.2	498	18	AI094757	qa13b01.x1 NCI_CGAP_Br	1.26e+00
c	435	15	17.2	502	6	W87453	zh67a08.s1 Soares feta	1.26e+00
c	436	15	17.2	502	13	AA881585	vx20e04.r1 Soares 2NbM	1.26e+00
	437	15	17.2	503	24	AA308574	EST179410 HCC cell lin	1.26e+00
c	438	15	17.2	505	19	N35714	yy26f02.s1 Homo sapien	1.26e+00
c	439	15	17.2	507	5	T62578	yc04h10.s1 Homo sapien	1.26e+00
c	440	15	17.2	507	19	N21568	yx60d06.s1 Homo sapien	1.26e+00
c	441	15	17.2	508	28	AQ186054	HS_3070_A2_H11_MF CIT	1.26e+00
c	442	15	17.2	509	16	R74270	yi57c04.r1 Homo sapien	1.26e+00
c	443	15	17.2	509	18	AI150872	qb68h03.x1 Soares_feta	1.26e+00
c	444	15	17.2	513	7	AA455911	aa14e09.s1 Soares NhHM	1.26e+00
	445	15	17.2	516	18	AI183407	qd24b11.x1 Soares_plac	1.26e+00
	446	15	17.2	518	8	AA237048	za01a04.r1 NCI_CGAP_GC	1.26e+00
	447	15	17.2	521	26	B76988	T29E2TR TAMU Arabidops	1.26e+00
c	448	15	17.2	522	7	AA529306	vi39a06.r1 Beddington	1.26e+00
	449	15	17.2	526	26	AQ017736	CIT-HSP-2310E21.TF CIT	1.26e+00
c	450	15	17.2	527	11	AA726594	vu93b12.r1 Stratagene	1.26e+00

c	451	15	17.2	528	9	AA535853	nf90a02.s1 NCI_CGAP_Co	1.26e+00
c	452	15	17.2	529	14	AA948052	oe60g09.s1 NCI_CGAP_Lu	1.26e+00
c	453	15	17.2	529	28	AQ109787	CIT-HSP-2372A12.TR CIT	1.26e+00
	454	15	17.2	530	11	N96567	21242 CD4-14 Arabidops	1.26e+00
c	455	15	17.2	530	14	C91282	Dictyostelium discoide	1.26e+00
c	456	15	17.2	535	8	AA130932	zo14c07.r1 Stratagene	1.26e+00
c	457	15	17.2	540	26	CI9G4	Ciona intestinalis gen	1.26e+00
	458	15	17.2	541	8	AA490650	aa01h06.r1 Soares NhHM	1.26e+00
	459	15	17.2	548	10	AA587014	nn73e06.s1 NCI_CGAP_La	1.26e+00
	460	15	17.2	549	15	AI050425	uc86b01.y1 Sugano mous	1.26e+00
	461	15	17.2	550	28	AQ171727	HS_3172_B2_F07_T7 CIT	1.26e+00
c	462	15	17.2	550	10	AA428065	zw57a05.s1 Soares tota	1.26e+00
c	463	15	17.2	555	6	AA150266	zl07e06.s1 Soares preg	1.26e+00
c	464	15	17.2	557	5	T80379	yd05c02.r1 Homo sapien	1.26e+00
	465	15	17.2	560	29	AQ238258	RPCI11-74I6.TK RPCI11	1.26e+00
	466	15	17.2	562	29	AQ231888	HS_2057_A2_C02_MR CIT	1.26e+00
c	467	15	17.2	564	17	AI089305	qb16c01.x1 Soares_preg	1.26e+00
	468	15	17.2	565	8	AA543125	vk36b11.r1 Soares mous	1.26e+00
c	469	15	17.2	566	16	R77200	yi65g05.r1 Homo sapien	1.26e+00
c	470	15	17.2	568	14	C91144	Dictyostelium discoide	1.26e+00
	471	15	17.2	573	29	AQ250701	T13M24-T7 TAMU Arabido	1.26e+00
c	472	15	17.2	574	17	AI091281	ow62c01.x1 Soares_NSF_	1.26e+00
c	473	15	17.2	579	22	AA098833	zn44c12.s1 Stratagene	1.26e+00
	474	15	17.2	584	8	AA543146	vk36d11.r1 Soares mous	1.26e+00
c	475	15	17.2	587	26	B39975	T20P21TFC TAMU Arabido	1.26e+00
c	476	15	17.2	589	21	W74445	zd77h06.s1 Soares feta	1.26e+00
	477	15	17.2	593	26	FR0014907	F.rubripes GSS sequenc	1.26e+00
	478	15	17.2	594	18	AI052376	oz23b07.x1 Soares_tota	1.26e+00
c	479	15	17.2	597	21	AA081359	zn34c11.s1 Stratagene	1.26e+00
	480	15	17.2	604	26	FR0025831	F.rubripes GSS sequenc	1.26e+00
c	481	15	17.2	608	27	AQ039527	CIT-HSP-2325D7.TV CIT-	1.26e+00
c	482	15	17.2	610	27	AQ077167	CIT-HSP-2359L15.TR CIT	1.26e+00
c	483	15	17.2	610	5	MUSB001A	Mus musculus expressed	1.26e+00
c	484	15	17.2	611	26	FR0025846	F.rubripes GSS sequenc	1.26e+00
c	485	15	17.2	614	26	FR0021002	F.rubripes GSS sequenc	1.26e+00
	486	15	17.2	617	9	HS39378	Homo sapiens mRNA; exp	1.26e+00
c	487	15	17.2	617	8	AA134833	zn90h04.r1 Stratagene	1.26e+00
	488	15	17.2	619	26	FR0005680	P.rubripes GSS sequenc	1.26e+00
c	489	15	17.2	620	22	AA015596	ze20g10.s1 Soares feta	1.26e+00
	490	15	17.2	622	13	AA825044	ETN359 Trypanosoma bru	1.26e+00
	491	15	17.2	631	27	B92876	CIT-HSP-2164N22.TR CIT	1.26e+00
	492	15	17.2	633	19	N36445	yx83c04.r1 Homo sapien	1.26e+00
c	493	15	17.2	637	21	W72457	zd67f01.s1 Soares feta	1.26e+00
c	494	15	17.2	638	26	AQ020441	CIT-HSP-2307C11.TF CIT	1.26e+00
c	495	15	17.2	642	28	AQ108235	CIT-HSP-2382P12.TR CIT	1.26e+00
	496	15	17.2	646	28	AQ110702	CIT-HSP-2378G16.TF CIT	1.26e+00
	497	15	17.2	650	23	AA066934	nm12e12.r1 Stratagene	1.26e+00
	498	15	17.2	650	27	AQ037429	CIT-HSP-2331M10.TF CIT	1.26e+00
c	499	15	17.2	656	12	AA739803	568 PtIPG2 Pinus taeda	1.26e+00
c	500	15	17.2	656	28	AQ109207	CIT-HSP-2383O14.TR CIT	1.26e+00
	501	15	17.2	657	17	AI073460	ov45a01.x1 Soares_test	1.26e+00

c	502	15	17.2	660	21	W60643	zd25f11.s1 Soares feta	1.26e+00
c	503	15	17.2	661	11	C83874	Dictyostelium discoide	1.26e+00
c	504	15	17.2	675	28	AQ163571	nbxb0007I10f CUGI Rice	1.26e+00
c	505	15	17.2	676	27	AQ035635	CIT-HSP-2320N24.TR CIT	1.26e+00
c	506	15	17.2	678	14	C91100	Dictyostelium discoide	1.26e+00
	507	15	17.2	681	26	B77619	T27E10TR TAMU Arabidop	1.26e+00
c	508	15	17.2	694	27	AQ048917	CLM-10h4-t cLM Giardia	1.26e+00
c	509	15	17.2	695	27	B59891	CIT-HSP-347A14.TPB CIT	1.26e+00
	510	15	17.2	696	27	AQ077426	CIT-HSP-2369F16.TR CIT	1.26e+00
c	511	15	17.2	702	27	AQ056056	CIT-HSP-2338L20.TR CIT	1.26e+00
c	512	15	17.2	703	29	AQ243334	HS_2060_B1_E02_MR CIT	1.26e+00
c	513	15	17.2	721	26	B76864	T27C8TR TAMU Arabidops	1.26e+00
c	514	15	17.2	725	14	C91106	Dictyostelium discoide	1.26e+00
c	515	15	17.2	730	18	AI084744	oz79f08.x1 Soares_sene	1.26e+00
	516	15	17.2	763	26	B06578	cSRL-82d5-u cSRL flow	1.26e+00
	517	15	17.2	764	7	C25569	Dictyostelium discoide	1.26e+00
c	518	15	17.2	766	27	AQ048180	cLM-6a8-t cLM Giardia	1.26e+00
c	519	15	17.2	848	20	W27863	39e8 Human retina cDNA	1.26e+00
	520	15	17.2	872	22	AA203316	zx54a08.r1 Soares feta	1.26e+00
c	521	15	17.2	888	26	B09564	F27G23-T7 IGF Arabidop	1.26e+00
c	522	15	17.2	908	26	B19143	F17F20-Sp6 IGF Arabido	1.26e+00
	523	15	17.2	911	26	B10963	T14N21-T7 TAMU Arabido	1.26e+00
	524	15	17.2	1014	7	C25704	Dictyostelium discoide	1.26e+00
c	525	15	17.2	1036	26	B12701	F28D3-T7.2 IGF Arabido	1.26e+00
c	526	15	17.2	1064	18	AU035503	Mus musculus cDNA, clo	1.26e+00
	527	15	17.2	1144	26	B09895	F24E23-T7.1 IGF Arabid	1.26e+00
	528	14	16.1	71	6	C21098	HUMGS0002619, Human Ge	1.88e+01
	529	14	16.1	73	26	B34603	HS-1025-A1-B03-MR.abi	1.88e+01
	530	14	16.1	78	15	AU007744	Schizosaccharomyces po	1.88e+01
	531	14	16.1	78	15	AU007745	Schizosaccharomyces po	1.88e+01
	532	14	16.1	79	10	AA637795	vr33c09.r1 Barstead mo	1.88e+01
c	533	14	16.1	122	17	AI086860	oz86c05.x1 Soares_sene	1.88e+01
c	534	14	16.1	123	15	AI028833	UI-R-C0-ij-h-11-0-UI.s	1.88e+01
	535	14	16.1	172	15	AA971759	op86b04.s1 Soares_NFL_	1.88e+01
c	536	14	16.1	180	20	R58641	G4494 Fetal heart Homo	1.88e+01
	537	14	16.1	183	15	AI139880	qd97h11.x1 Soares_test	1.88e+01
	538	14	16.1	196	17	AU012120	Schizosaccharomyces po	1.88e+01
	539	14	16.1	196	8	AA284039	za47f04.s1 NCI_CGAP_GC	1.88e+01
	540	14	16.1	203	23	AA123017	mp82e07.r1 Soares 2NbM	1.88e+01
	541	14	16.1	205	13	AA841757	MB3D6V6H03T3 Brugia ma	1.88e+01
c	542	14	16.1	207	19	H69712	yr93c02.s1 Homo sapien	1.88e+01
c	543	14	16.1	211	20	HSC0CC101	H. sapiens partial cDN	1.88e+01
	544	14	16.1	219	8	AA465141	aa33a04.s1 NCI_CGAP_GC	1.88e+01
	545	14	16.1	220	17	AI119994	uc25c10.r1 Soares mous	1.88e+01
c	546	14	16.1	222	19	H97121	yw13c10.r1 Homo sapien	1.88e+01
	547	14	16.1	223	7	HUM107E09B	Human fetal brain cDNA	1.88e+01
c	548	14	16.1	223	20	T26366	AB105E10F Homo sapiens	1.88e+01
	549	14	16.1	224	19	N45269	yv26g02.r1 Homo sapien	1.88e+01
	550	14	16.1	224	13	AA167282	zp14d11.s1 Stratagene	1.88e+01
	551	14	16.1	228	11	AA671215	vn80g02.r1 Knowles Sol	1.88e+01
	552	14	16.1	230	7	HUM125A06A	Human fetal brain cDNA	1.88e+01

	553	14	16.1	239	8	AA480318	ne32g02.s1 NCI_CGAP_Co	1.88e+01
c	554	14	16.1	243	10	AF029854	Entamoeba dispar clone	1.88e+01
c	555	14	16.1	244	20	HSBA4E031	H. sapiens partial cDN	1.88e+01
	556	14	16.1	247	13	AA689564	nx06g03.s1 NCI_CGAP_GC	1.88e+01
	557	14	16.1	248	7	HUM302D05A	Human aorta cDNA 3'-en	1.88e+01
c	558	14	16.1	251	20	R57315	F2677 Fetal heart Homo	1.88e+01
	559	14	16.1	252	9	AA559245	nl68f02.s1 NCI_CGAP_Pr	1.88e+01
	560	14	16.1	254	19	N47119	yy53a04.r1 Homo sapien	1.88e+01
	561	14	16.1	258	28	AQ099040	HS_3042_A2_A06_MR CIT	1.88e+01
	562	14	16.1	259	21	T17367	NIB802 Normalized infa	1.88e+01
c	563	14	16.1	259	23	AA250111	mx79f08.r1 Soares mous	1.88e+01
c	564	14	16.1	260	22	AA090893	yy0798.seq.F Fetal hea	1.88e+01
	565	14	16.1	261	9	AA577429	nm96g09.s1 NCI_CGAP_Co	1.88e+01
	566	14	16.1	262	6	C11253	C.elegans cDNA clone y	1.88e+01
c	567	14	16.1	264	6	C16380	Human aorta cDNA 5'-en	1.88e+01
	568	14	16.1	267	25	AA369929	EST81640 Prostate glan	1.88e+01
	569	14	16.1	267	10	AA583663	nn58b10.s1 NCI_CGAP_Ki	1.88e+01
	570	14	16.1	269	11	AA705397	zj90c03.s1 Soares feta	1.88e+01
	571	14	16.1	269	20	HSB88H071	H. sapiens partial cDN	1.88e+01
c	572	14	16.1	270	26	B07550	Mouse Igh variable reg	1.88e+01
c	573	14	16.1	276	22	N75899	yx63e09.r1 Soares mult	1.88e+01
	574	14	16.1	280	18	AI185559	qe54g09.x1 Soares_feta	1.88e+01
c	575	14	16.1	282	9	AA564950	nj01f10.s1 NCI_CGAP_Pr	1.88e+01
c	576	14	16.1	283	8	AA281781	zt07b02.s1 NCI_CGAP_GC	1.88e+01
	577	14	16.1	285	5	R41990	yg12f11.s1 Homo sapien	1.88e+01
c	578	14	16.1	285	6	AA383817	EST97249 Testis I Homo	1.88e+01
c	579	14	16.1	288	6	AA026114	zk01f08.s1 Soares preg	1.88e+01
	580	14	16.1	291	16	T33535	EST58179 Homo sapiens	1.88e+01
c	581	14	16.1	294	16	T34720	EST74203 Homo sapiens	1.88e+01
	582	14	16.1	295	19	H79253	yu48a06.s1 Homo sapien	1.88e+01
c	583	14	16.1	296	6	AA156818	zl20c05.s1 Soares preg	1.88e+01
	584	14	16.1	299	6	AA385805	EST99492 Thyroid Homo	1.88e+01
	585	14	16.1	300	15	C94728	Sus scrofa mRNA; expre	1.88e+01
	586	14	16.1	300	9	C33776	C.elegans cDNA clone y	1.88e+01
c	587	14	16.1	301	27	AQ067280	HS_2237_A2_G06_MF CIT	1.88e+01
	588	14	16.1	303	16	R69854	yi45h11.s1 Homo sapien	1.88e+01
c	589	14	16.1	303	24	AA302760	EST110439 Adipose tissu	1.88e+01
	590	14	16.1	305	13	AA841232	MB3D6AA5C01T3 Brugia m	1.88e+01
c	591	14	16.1	305	20	HSC2WE052	H. sapiens partial cDN	1.88e+01
	592	14	16.1	306	29	AQ236238	HS_2052_B1_D11_MR CIT	1.88e+01
	593	14	16.1	306	24	AA323100	EST25833 Cerebellum II	1.88e+01
c	594	14	16.1	307	5	R28495	yh54h09.s1 Homo sapien	1.88e+01
c	595	14	16.1	308	15	AI030405	UI-R-C0-jc-d-12-0-UI.s	1.88e+01
	596	14	16.1	312	17	AI085083	ow88h10.s1 Soares_feta	1.88e+01
c	597	14	16.1	313	11	AA678049	zi25e08.s1 Soares feta	1.88e+01
	598	14	16.1	314	19	N70370	za44f09.s1 Homo sapien	1.88e+01
c	599	14	16.1	314	23	AA089061	mm51c09.r1 Stratagene	1.88e+01
	600	14	16.1	315	19	HSC2XC052	H. sapiens partial cDN	1.88e+01
	601	14	16.1	316	27	AQ065678	HS_2238_B2_F07_MF CIT	1.88e+01
	602	14	16.1	317	21	R85791	yq27b01.s1 Soares feta	1.88e+01
	603	14	16.1	318	28	AQ178009	HS_2216_B2_B06_MF CIT	1.88e+01

	604	14	16.1	318	26	AQ019563	CIT-HSP-2309C9.TR CIT-	1.88e+01
	605	14	16.1	319	14	AA857565	of64e06.s1 NCI_CGAP_Co	1.88e+01
	606	14	16.1	320	22	N54841	yv73b09.s1 Soares feta	1.88e+01
c	607	14	16.1	322	24	AA322794	EST25568 Cerebellum II	1.88e+01
	608	14	16.1	323	27	AQ067632	HS_2238_B2_G05_MR CIT	1.88e+01
c	609	14	16.1	323	11	AA069806	zm15f11.s1 Stratagene	1.88e+01
c	610	14	16.1	323	29	AQ229615	HS_2028_A2_B03_MR CIT	1.88e+01
c	611	14	16.1	323	27	AQ027694	CIT-HSP-2323A17.TR CIT	1.88e+01
	612	14	16.1	328	19	H58725	yr25e03.s1 Homo sapien	1.88e+01
c	613	14	16.1	328	7	CELK090F6F	C.elegans cDNA clone y	1.88e+01
	614	14	16.1	329	15	AI034856	lmLv39p10/706C Leishma	1.88e+01
	615	14	16.1	329	27	AQ039563	CIT-HSP-2321F18.TF CIT	1.88e+01
c	616	14	16.1	330	27	AQ081023	CIT-HSP-2362G15.TF CIT	1.88e+01
c	617	14	16.1	331	16	R93125	yq15c02.s1 Homo sapien	1.88e+01
	618	14	16.1	332	7	HUM087G12B	Human fetal brain cDNA	1.88e+01
c	619	14	16.1	333	6	C03910	Human Heart cDNA, clon	1.88e+01
c	620	14	16.1	334	6	CELK061E8R	C.elegans cDNA clone y	1.88e+01
c	621	14	16.1	334	6	AA384946	EST98523 Thyroid Homo	1.88e+01
c	622	14	16.1	335	23	AA177707	mt20b11.r1 Soares mous	1.88e+01
c	623	14	16.1	335	27	AQ072836	HS_2266_B1_G08_MR CIT	1.88e+01
c	624	14	16.1	335	23	AA087319	mn95b09.r1 Stratagene	1.88e+01
c	625	14	16.1	335	12	AA749422	ny08g10.s1 NCI_CGAP_GC	1.88e+01
	626	14	16.1	336	20	N79936	yz86d06.r1 Homo sapien	1.88e+01
c	627	14	16.1	336	19	N66158	yy66f06.s1 Homo sapien	1.88e+01
	628	14	16.1	338	22	AA082339	ze88g04.r1 Soares feta	1.88e+01
	629	14	16.1	339	7	HUM094A05B	Human fetal brain cDNA	1.88e+01
	630	14	16.1	340	20	HSB10C082	H. sapiens partial cDN	1.88e+01
	631	14	16.1	340	8	AA410656	zt29g03.r1 Soares ovar	1.88e+01
c	632	14	16.1	341	19	N40495	yw74f12.r1 Homo sapien	1.88e+01
	633	14	16.1	342	5	R39089	yf50a10.s1 Homo sapien	1.88e+01
	634	14	16.1	343	7	HUM009G01A	Human fetal brain cDNA	1.88e+01
	635	14	16.1	344	8	AA280735	zs96h08.s1 NCI_CGAP_GC	1.88e+01
	636	14	16.1	344	7	HUM159F11B	Human fetal brain cDNA	1.88e+01
	637	14	16.1	344	5	T84182	yd45c09.r1 Homo sapien	1.88e+01
	638	14	16.1	346	8	AA278303	zs80b12.s1 NCI_CGAP_GC	1.88e+01
	639	14	16.1	346	15	AA918156	om37a01.s1 Soares_NFL_	1.88e+01
	640	14	16.1	347	7	HUM169A06B	Human fetal brain cDNA	1.88e+01
	641	14	16.1	347	27	AQ002505	CIT-HSP-2286O8.TF CIT-	1.88e+01
c	642	14	16.1	348	19	HSC2HH091	H. sapiens partial cDN	1.88e+01
	643	14	16.1	351	8	AA393446	zt75g01.r1 Soares test	1.88e+01
	644	14	16.1	356	6	C00060	HUMGS0004294, Human Ge	1.88e+01
c	645	14	16.1	356	10	AA599234	ag24a01.s1 Jia bone ma	1.88e+01
	646	14	16.1	360	9	C41151	C.elegans cDNA clone y	1.88e+01
c	647	14	16.1	360	9	C49378	C.elegans cDNA clone y	1.88e+01
c	648	14	16.1	360	9	C40592	C.elegans cDNA clone y	1.88e+01
c	649	14	16.1	360	9	C43489	C.elegans cDNA clone y	1.88e+01
c	650	14	16.1	360	9	C70783	C.elegans cDNA clone y	1.88e+01
c	651	14	16.1	360	6	CELK052E1F	C.elegans cDNA clone y	1.88e+01
c	652	14	16.1	360	9	C49235	C.elegans cDNA clone y	1.88e+01
c	653	14	16.1	360	6	CELK060F1F	C.elegans cDNA clone y	1.88e+01
c	654	14	16.1	360	9	C39516	C.elegans cDNA clone y	1.88e+01

c	655	14	16.1	360	9	C43971	C.elegans cDNA clone y	1.88e+01
c	656	14	16.1	360	14	AA938306	oc11a04.s1 NCI_CGAP_GC	1.88e+01
	657	14	16.1	361	8	AA505768	nh97c06.s1 NCI_CGAP_Br	1.88e+01
c	658	14	16.1	361	22	W95808	ze07g09.s1 Soares feta	1.88e+01
c	659	14	16.1	361	23	AA182069	mt59c05.r1 Soares 2NbM	1.88e+01
c	660	14	16.1	362	24	AA326557	EST29749 Cerebellum II	1.88e+01
c	661	14	16.1	363	5	R05860	ye93a05.r1 Homo sapien	1.88e+01
	662	14	16.1	364	28	AQ223025	HS_2017_A2_H08_T7 CIT	1.88e+01
	663	14	16.1	365	28	AQ134108	HS_3056_A1_G07_MR CIT	1.88e+01
	664	14	16.1	366	9	AA577796	nn24b07.s1 NCI_CGAP_Ga	1.88e+01
	665	14	16.1	367	27	AQ083412	RPCI11-55F19.TV RPCI11	1.88e+01
	666	14	16.1	372	27	AQ063420	CIT-HSP-2355O20.TF CIT	1.88e+01
	667	14	16.1	372	7	AA450023	zx33h06.s1 Soares tota	1.88e+01
c	668	14	16.1	372	8	AA211793	zr89a05.r1 NCI_CGAP_GC	1.88e+01
c	669	14	16.1	372	9	C47660	C.elegans cDNA clone y	1.88e+01
	670	14	16.1	373	19	N51633	yy87a06.s1 Homo sapien	1.88e+01
	671	14	16.1	373	21	AA018342	ze41d10.s1 Soares_reti	1.88e+01
	672	14	16.1	374	29	AQ146233	HS_2229_B1_C10_MF CIT	1.88e+01
	673	14	16.1	374	20	W38982	zb28g10.r1 Soares para	1.88e+01
c	674	14	16.1	374	7	AA455429	aa16g05.r1 Soares NhHM	1.88e+01
c	675	14	16.1	374	11	AA704166	zj77a12.s1 Soares feta	1.88e+01
	676	14	16.1	375	9	C39646	C.elegans cDNA clone y	1.88e+01
	677	14	16.1	376	11	T44510	7773 Lambda-PRL2 Arabi	1.88e+01
c	678	14	16.1	377	27	B83253	RPCI11-25K13.TP RPCI11	1.88e+01
c	679	14	16.1	377	16	R55954	yg89d09.r1 Homo sapien	1.88e+01
	680	14	16.1	378	20	N89787	zb34a01.s1 Soares para	1.88e+01
	681	14	16.1	378	23	T02909	FBL5C1 Fetal brain, St	1.88e+01
c	682	14	16.1	378	15	AA997438	UI-R-C0-hr-h-03-0-UI.s	1.88e+01
c	683	14	16.1	379	8	AA402514	zt66h06.r1 Soares test	1.88e+01
	684	14	16.1	381	15	AA972687	op91c02.s1 Soares_NFL_	1.88e+01
	685	14	16.1	381	13	AA875898	ob34g05.s1 NCI_CGAP_Ki	1.88e+01
	686	14	16.1	381	6	AA029046	zk09g09.s1 Soares preg	1.88e+01
c	687	14	16.1	381	28	AQ210069	HS_2225_B1_C07_MF CIT	1.88e+01
	688	14	16.1	382	7	AA448107	zw82d05.r1 Soares test	1.88e+01
c	689	14	16.1	383	6	CE2E6	C. elegans cDNA clone	1.88e+01
	690	14	16.1	384	9	AA597910	29159 Lambda-PRL2 Arab	1.88e+01
	691	14	16.1	385	27	AQ028039	CIT-HSP-2315K6.TF CIT-	1.88e+01
	692	14	16.1	385	12	AA769812	ah81a06.s1 Soares test	1.88e+01
	693	14	16.1	385	28	AQ105494	HS_3023_A2_G06_T7 CIT	1.88e+01
c	694	14	16.1	385	7	AA447608	zw82d05.s1 Soares test	1.88e+01
c	695	14	16.1	387	5	T06940	EST04829 Homo sapiens	1.88e+01
c	696	14	16.1	387	13	AA833397	ub58a09.r1 Soares mous	1.88e+01
	697	14	16.1	389	22	AA149287	zl25e06.s1 Soares preg	1.88e+01
	698	14	16.1	389	10	AA640380	nt69f12.s1 NCI_CGAP_Pr	1.88e+01
c	699	14	16.1	389	14	AA907849	ol24f06.s1 Soares_NFL_	1.88e+01
c	700	14	16.1	390	6	AA386321	EST185071 Fetal brain	1.88e+01
c	701	14	16.1	391	28	AQ190824	HS_3231_B2_B11_MR CIT	1.88e+01
	702	14	16.1	392	29	AQ148101	HS_3112_A2_H06_MR CIT	1.88e+01
c	703	14	16.1	395	18	AI181900	ud73e03.x1 Sugano mous	1.88e+01
	704	14	16.1	396	23	AA217442	mu94a12.r1 Soares mous	1.88e+01
c	705	14	16.1	396	19	H77434	yu08h10.s1 Homo sapien	1.88e+01

c	706	14	16.1	397	19	H70981	yr73h03.r1 Homo sapien	1.88e+01
c	707	14	16.1	397	14	AA984197	am82c06.s1 Stratagene	1.88e+01
	708	14	16.1	398	27	AQ066377	HS_2243_B1_D07_MF CIT	1.88e+01
	709	14	16.1	398	28	AQ167482	HS_3165_A1_B09_MR CIT	1.88e+01
c	710	14	16.1	399	19	N47869	yy95e10.s1 Homo sapien	1.88e+01
c	711	14	16.1	400	9	AA604767	no84f10.s1 NCI_CGAP_AA	1.88e+01
	712	14	16.1	401	13	AA841840	MB3D6V8A03T3 Brugia ma	1.88e+01
	713	14	16.1	402	16	H54516	yq91b05.r1 Homo sapien	1.88e+01
c	714	14	16.1	403	14	AA800972	EST190469 Normalized r	1.88e+01
c	715	14	16.1	405	27	B89866	CIT-HSP-2169P19.TR CIT	1.88e+01
c	716	14	16.1	406	16	R59183	yg97b02.s1 Homo sapien	1.88e+01
	717	14	16.1	407	14	AA857944	of69e09.s1 NCI_CGAP_Co	1.88e+01
	718	14	16.1	407	28	AQ124725	HS_2266_A1_G06_MF CIT	1.88e+01
c	719	14	16.1	407	6	C15793	Human fetal brain cDNA	1.88e+01
c	720	14	16.1	407	28	AQ181069	HS_3219_B2_H04_MR CIT	1.88e+01
	721	14	16.1	408	21	W69746	zd51f01.s1 Soares feta	1.88e+01
	722	14	16.1	409	10	AA110233	mp42c05.r1 Barstead MP	1.88e+01
c	723	14	16.1	409	7	AA456978	aa90h07.s1 Stratagene	1.88e+01
c	724	14	16.1	410	27	AQ071052	HS_3013_A2_D06_T7 CIT	1.88e+01
	725	14	16.1	411	6	CELK045G6R	C.elegans cDNA clone y	1.88e+01
	726	14	16.1	412	17	AI017156	ou28a02.x1 Soares_NFL_	1.88e+01
c	727	14	16.1	412	23	AA124260	mp98b05.r1 Soares 2NbM	1.88e+01
	728	14	16.1	413	10	AA613002	nq39c08.s1 NCI_CGAP_Co	1.88e+01
	729	14	16.1	413	16	R91881	yq03h12.r1 Homo sapien	1.88e+01
c	730	14	16.1	413	13	AA869489	vq08f11.r1 Barstead st	1.88e+01
c	731	14	16.1	413	27	B98994	CIT-HSP-2289N23.TR CIT	1.88e+01
c	732	14	16.1	414	27	AQ082337	RPCI11-55N8.TJ RPCI11	1.88e+01
	733	14	16.1	415	10	AA594432	nl97h06.s1 NCI_CGAP_Co	1.88e+01
	734	14	16.1	415	13	AA840840	MB3D6AA0D02T3 Brugia m	1.88e+01
	735	14	16.1	416	11	T43548	6811 Lambda-PRL2 Arabi	1.88e+01
c	736	14	16.1	416	27	B80550	CIT-HSP-2050D11.TF CIT	1.88e+01
c	737	14	16.1	416	17	AI016817	ov33b03.x1 Soares_test	1.88e+01
	738	14	16.1	417	19	N70445	za62d11.s1 Homo sapien	1.88e+01
	739	14	16.1	418	28	AQ217090	HS_2262_B2_H09_T7 CIT	1.88e+01
c	740	14	16.1	418	6	AA004733	zh93a04.r1 Soares feta	1.88e+01
	741	14	16.1	420	15	AI046775	uh57g04.r1 Soares mous	1.88e+01
c	742	14	16.1	420	6	AA127047	zl22c01.s1 Soares preg	1.88e+01
c	743	14	16.1	420	27	AQ032271	HS_2194_B2_H05_MF CIT	1.88e+01
c	744	14	16.1	421	22	AA172172	zp29b09.s1 Stratagene	1.88e+01
	745	14	16.1	422	12	AA752491	96MJ0418 Rice Immature	1.88e+01
	746	14	16.1	422	17	AI017049	ov01c04.x1 NCI_CGAP_Ki	1.88e+01
c	747	14	16.1	422	11	N95908	21317 Lambda-PRL1 Arab	1.88e+01
	748	14	16.1	423	16	H03970	yj44g02.s1 Homo sapien	1.88e+01
	749	14	16.1	423	17	AU021327	Mus musculus 8-cell em	1.88e+01
	750	14	16.1	424	16	R61359	yh15c02.s1 Homo sapien	1.88e+01
c	751	14	16.1	424	10	AA622551	np21d06.s1 NCI_CGAP_GC	1.88e+01
	752	14	16.1	425	16	H12117	ym16f11.s1 Homo sapien	1.88e+01
	753	14	16.1	425	9	AA564326	nk45a10.s1 NCI_CGAP_GC	1.88e+01
c	754	14	16.1	425	28	AQ110159	CIT-HSP-2370D10.TF CIT	1.88e+01
	755	14	16.1	427	6	CELK033GYR	C.elegans cDNA clone y	1.88e+01
	756	14	16.1	428	21	AA054946	zf16c12.s1 Soares feta	1.88e+01

c	757	14	16.1	428	16	R45160	yg40c03.s1 Homo sapien	1.88e+01
	758	14	16.1	429	11	AA669486	af74g09.r1 Soares NhHM	1.88e+01
	759	14	16.1	429	28	AQ108412	CIT-HSP-2374B19.TF CIT	1.88e+01
	760	14	16.1	430	15	AI036102	ub66d10.r1 Soares mous	1.88e+01
	761	14	16.1	430	21	W45146	zc20f10.s1 Soares sene	1.88e+01
c	762	14	16.1	431	21	W81133	zd85e12.s1 Soares feta	1.88e+01
	763	14	16.1	432	17	AI031574	ow06e08.x1 Soares_para	1.88e+01
	764	14	16.1	433	18	AI102299	EST211588 Normalized r	1.88e+01
	765	14	16.1	433	19	H70998	yu70c01.r1 Homo sapien	1.88e+01
c	766	14	16.1	433	28	AQ182673	HS_3103_B1_A07_MR CIT	1.88e+01
	767	14	16.1	434	5	T83874	yd66f02.s1 Homo sapien	1.88e+01
	768	14	16.1	434	10	AA625048	SWOv3MCA1522SK Onchoce	1.88e+01
c	769	14	16.1	434	10	AA593823	nn17b04.s1 NCI_CGAP_Co	1.88e+01
c	770	14	16.1	434	26	B14440	A-791F7.TF CIT978SKA2	1.88e+01
	771	14	16.1	435	8	AA252745	za26b03.r1 NCI_CGAP_GC	1.88e+01
c	772	14	16.1	435	17	AI087951	oo20g09.x1 Soares_NSF_	1.88e+01
	773	14	16.1	436	20	W03617	za02h12.r1 Soares_mela	1.88e+01
c	774	14	16.1	436	18	AI175998	EST219574 Normalized r	1.88e+01
	775	14	16.1	440	14	AA873507	oh70b03.s1 NCI_CGAP_Ki	1.88e+01
	776	14	16.1	440	19	N62996	yz84g08.s1 Homo sapien	1.88e+01
c	777	14	16.1	440	15	AA961613	op30h06.s1 Soares_NFL_	1.88e+01
c	778	14	16.1	440	29	AQ153128	HS_2219_A1_F07_MF CIT	1.88e+01
c	779	14	16.1	440	23	AA124991	mp74f03.r1 Soares 2NbM	1.88e+01
	780	14	16.1	441	18	AI094991	qa14b11.x1 NCI_CGAP_Br	1.88e+01
c	781	14	16.1	441	17	AI078736	ox51c02.x1 Soares_tota	1.88e+01
c	782	14	16.1	441	9	AA565613	nk47e09.s1 NCI_CGAP_GC	1.88e+01
	783	14	16.1	442	6	AA099743	zk86g11.s1 Soares preg	1.88e+01
c	784	14	16.1	442	5	T80542	yd08c10.r1 Homo sapien	1.88e+01
c	785	14	16.1	442	12	AA739052	vv66d12.r1 Stratagene	1.88e+01
	786	14	16.1	443	12	AA725200	ai09h09.s1 Soares para	1.88e+01
	787	14	16.1	443	11	AA678809	ah09a08.s1 Gessler Wil	1.88e+01
	788	14	16.1	443	18	AI093007	qa77e06.x1 Soares_feta	1.88e+01
	789	14	16.1	445	20	W47259	zc34a06.r1 Soares sene	1.88e+01
	790	14	16.1	445	26	AQ021393	CIT-HSP-2308O3.TF CIT-	1.88e+01
c	791	14	16.1	445	8	AA215313	zr94c07.r1 NCI_CGAP_GC	1.88e+01
c	792	14	16.1	445	22	N59464	yz31b10.s1 Soares mult	1.88e+01
	793	14	16.1	446	19	N25553	yx76f05.s1 Homo sapien	1.88e+01
	794	14	16.1	448	26	AQ018776	CIT-HSP-2308N14.TR CIT	1.88e+01
c	795	14	16.1	448	27	AQ062793	CIT-HSP-2349M22.TR CIT	1.88e+01
	796	14	16.1	451	17	AI015648	ou95b04.x1 NCI_CGAP_Ki	1.88e+01
	797	14	16.1	452	5	T98579	ye60b10.s1 Homo sapien	1.88e+01
	798	14	16.1	452	19	N20840	yx51c10.s1 Homo sapien	1.88e+01
c	799	14	16.1	452	6	AA436629	zw55e11.s1 Soares tota	1.88e+01
c	800	14	16.1	452	18	AI167869	ox91b03.x1 Soares_para	1.88e+01
	801	14	16.1	453	20	W00466	yz99f01.r1 Homo sapien	1.88e+01
	802	14	16.1	454	29	AQ236299	HS_2028_A1_A04_MR CIT	1.88e+01
c	803	14	16.1	454	23	AA170579	ms86a09.r1 Soares mous	1.88e+01
	804	14	16.1	455	11	AA688887	vs01h05.r1 Barstead mo	1.88e+01
	805	14	16.1	458	28	AQ113697	CIT-HSP-2371A24.TF CIT	1.88e+01
c	806	14	16.1	459	27	AQ053330	CIT-HSP-2339P18.TF CIT	1.88e+01
c	807	14	16.1	459	19	N73795	yz80e01.s1 Homo sapien	1.88e+01

c	808	14	16.1	460	7	AA442507	zv59c07.r1 Soares test	1.88e+01
	809	14	16.1	461	6	C19993	Rice cDNA, partial seq	1.88e+01
	810	14	16.1	461	29	AQ144757	HS_3092_A1_E11_MF CIT	1.88e+01
c	811	14	16.1	461	6	AA386634	vb54f01.r1 Ko mouse em	1.88e+01
c	812	14	16.1	461	17	AI004340	ot95a06.x1 Soares_tota	1.88e+01
	813	14	16.1	463	23	AA227826	zr56f11.r1 Soares NhHM	1.88e+01
	814	14	16.1	463	23	AA240083	mv31f03.r1 GuayWoodfor	1.88e+01
c	815	14	16.1	464	22	W25579	zc64g07.r1 Soares feta	1.88e+01
c	816	14	16.1	464	14	AA893808	EST197611 Normalized r	1.88e+01
	817	14	16.1	465	19	H57528	yr05b09.r1 Homo sapien	1.88e+01
c	818	14	16.1	465	29	AQ153011	HS_2231_B1_F10_MR CIT	1.88e+01
c	819	14	16.1	465	29	AQ251086	F18E7-8p6 IGF Arabidop	1.88e+01
	820	14	16.1	466	14	AA979104	LD33369.3prime LD Dros	1.88e+01
	821	14	16.1	466	23	AA154931	mr32g05.r1 Soares mous	1.88e+01
c	822	14	16.1	466	8	AA235153	za38c10.r1 Soares NhHM	1.88e+01
	823	14	16.1	467	19	N50647	yx18h11.s1 Homo sapien	1.88e+01
	824	14	16.1	467	16	R98760	yq68a02.r1 Homo sapien	1.88e+01
c	825	14	16.1	467	16	R96037	yr83d10.r1 Homo sapien	1.88e+01
c	826	14	16.1	467	26	B97624	F15P21TPC IGF Arabidop	1.88e+01
	827	14	16.1	468	8	AA481770	aa93b04.s1 Stratagene	1.88e+01
c	828	14	16.1	468	19	N73459	za05d02.s1 Homo sapien	1.88e+01
c	829	14	16.1	468	12	AA739453	vv53f10.r1 Soares 2NbM	1.88e+01
	830	14	16.1	469	21	W81089	zd84e08.s1 Soares feta	1.88e+01
	831	14	16.1	469	16	R46373	yg50b05.s1 Homo sapien	1.88e+01
	832	14	16.1	469	18	AI038912	ox96f06.x1 Soares_sene	1.88e+01
	833	14	16.1	469	10	AA109931	mp48c09.r1 Barstead MP	1.88e+01
c	834	14	16.1	469	18	AI173831	ud72f04.x1 Sugano mous	1.88e+01
c	835	14	16.1	470	5	R42501	yf91a09.s1 Homo sapien	1.88e+01
c	836	14	16.1	470	8	AA504599	aa60h04.r1 NCI_CGAP_GC	1.88e+01
	837	14	16.1	471	12	AA780346	af56a04.s1 Soares tota	1.88e+01
	838	14	16.1	471	13	AA884213	am37e04.s1 Soares NPL	1.88e+01
	839	14	16.1	471	18	AI144372	qb73c07.x1 Soares_feta	1.88e+01
c	840	14	16.1	471	27	B69520	CIT-HSP-2054K12.TF CIT	1.88e+01
	841	14	16.1	472	13	AA822436	vw13a03.r1 Soares 2NbM	1.88e+01
	842	14	16.1	473	22	AA009944	zs16d05.s1 Soares feta	1.88e+01
	843	14	16.1	474	5	T62928	yb99e12.s1 Homo sapien	1.88e+01
c	844	14	16.1	474	26	AQ021867	CIT-HSP-2310D10.TR CIT	1.88e+01
	845	14	16.1	475	20	W51610	AS4SLC884T3 Brugia mal	1.88e+01
c	846	14	16.1	475	15	AA969265	on57d09.s1 Soares_NFL_	1.88e+01
	847	14	16.1	476	10	T13079	1145 Lambda-PRL2 Arabi	1.88e+01
c	848	14	16.1	477	19	N73094	za08f03.s1 Homo sapien	1.88e+01
c	849	14	16.1	477	26	FR0004832	F.rubripes GSS sequenc	1.88e+01
	850	14	16.1	478	6	AA131946	zl39h12.s1 Soares preg	1.88e+01
c	851	14	16.1	478	21	W58532	zd20c02.r1 Soares feta	1.88e+01
	852	14	16.1	479	21	W30316	mc27f08.r1 Soares mous	1.88e+01
	853	14	16.1	481	27	B94659	CIT-HSP-216507.TF CIT-	1.88e+01
	854	14	16.1	481	29	AQ151756	HS_3080_B1_B01_T7 CIT	1.88e+01
	855	14	16.1	482	14	AA879463	oj91g12.s1 Soares_NFL_	1.88e+01
c	856	14	16.1	482	5	T83081	yd40e05.r1 Homo sapien	1.88e+01
c	857	14	16.1	482	11	H76001	17706 Lambda-PRL2 Arab	1.88e+01
c	858	14	16.1	484	5	R12047	yf53g10.r1 Homo sapien	1.88e+01

c	859	14	16.1	486	29	AQ235007	HS_2056_A2_F06_T7 CIT	1.88e+01
c	860	14	16.1	486	28	AQ194063	RPCI11-62F5.TK RPCI11	1.88e+01
	861	14	16.1	487	5	T65113	yc74b05.s1 Homo sapien	1.88e+01
c	862	14	16.1	487	5	R17782	yg04d09.r1 Homo sapien	1.88e+01
	863	14	16.1	489	5	R37391	yf77g03.s1 Homo sapien	1.88e+01
	864	14	16.1	489	18	AI139313	hc24d11.x1 Soares_preg	1.88e+01
c	865	14	16.1	489	19	N42851	yy13a05.r1 Homo sapien	1.88e+01
c	866	14	16.1	489	29	AQ137425	HS_3071_A2_E01_MF CIT	1.88e+01
	867	14	16.1	490	28	AQ171093	HS_3076_A2_D02_MF CIT	1.88e+01
	868	14	16.1	492	23	AA138884	mr09b09.r1 Soares mous	1.88e+01
c	869	14	16.1	492	15	AA938143	on28c10.s1 NCI_CGAP_Lu	1.88e+01
c	870	14	16.1	492	27	B88765	RPCI11-24L21.TPC RPCI1	1.88e+01
c	871	14	16.1	493	5	T71648	yd36h08.s1 Homo sapien	1.88e+01
c	872	14	16.1	494	15	C79592	Mus musculus 3.5-dpc b	1.88e+01
c	873	14	16.1	494	13	AA633585	ac15e12.s1 Stratagene	1.88e+01
c	874	14	16.1	495	28	AQ191037	HS_2212_B1_D02_MR CIT	1.88e+01
c	875	14	16.1	497	28	AQ121354	HS_3083_B1_E07_MF CIT	1.88e+01
	876	14	16.1	500	28	AQ180500	HS_3216_B2_H09_T7 CIT	1.88e+01
c	877	14	16.1	500	22	AA071517	ze88g04.s1 Soares feta	1.88e+01
c	878	14	16.1	500	19	N44643	yy33b09.r1 Homo sapien	1.88e+01
	879	14	16.1	502	12	AA753042	97AS1808 Rice Immature	1.88e+01
c	880	14	16.1	502	12	AA768411	oa63c11.s1 NCI_CGAP_GC	1.88e+01
c	881	14	16.1	502	28	AQ210413	HS_2197_B1_F10_MR CIT	1.88e+01
c	882	14	16.1	502	22	AA189799	mt91f03.r1 Soares mous	1.88e+01
c	883	14	16.1	503	18	AI148175	qb56d04.x1 NCI_CGAP_Br	1.88e+01
	884	14	16.1	504	20	MMBIEST1	M.musculus mRNA; EST s	1.88e+01
	885	14	16.1	505	6	AA433194	EST11 Drosophila melan	1.88e+01
	886	14	16.1	505	17	AI127979	qc37c02.x1 Soares_preg	1.88e+01
	887	14	16.1	507	19	N51691	yy72b08.s1 Homo sapien	1.88e+01
c	888	14	16.1	509	7	AA509411	vg20g03.r1 Soares mous	1.88e+01
	889	14	16.1	510	22	AA205345	zq68a06.r1 Stratagene	1.88e+01
	890	14	16.1	511	5	T86792	yd86e03.r1 Homo sapien	1.88e+01
c	891	14	16.1	512	26	B14773	342P18.TV CIT978SKA1 H	1.88e+01
	892	14	16.1	514	26	B07331	G396T7 MVAT4 sheared g	1.88e+01
c	893	14	16.1	514	26	B16346	A-791F7.TVB CIT978SKA2	1.88e+01
	894	14	16.1	515	12	AA777259	zl96s06.s1 Soares feta	1.88e+01
c	895	14	16.1	516	28	AQ118399	HS_3007_A2_C07_T7 CIT	1.88e+01
c	896	14	16.1	516	13	AA173555	zp04c01.s1 Stratagene	1.88e+01
c	897	14	16.1	524	29	AQ231097	HS_2026_B1_B10_T7 CIT	1.88e+01
c	898	14	16.1	524	29	AQ148667	HS_3137_B2_D10_T7 CIT	1.88e+01
c	899	14	16.1	525	13	AA858996	UI-R-A0-bi-c-02-0-UI.s	1.88e+01
	900	14	16.1	526	11	AA703249	zi70f01.s1 Soares feta	1.88e+01
c	901	14	16.1	527	27	FR0033960	Fugu rubripes GSS sequ	1.88e+01
c	902	14	16.1	529	19	N21387	yx54d04.s1 Homo sapien	1.88e+01
	903	14	16.1	530	10	AA631237	nq81a09.s1 NCI_CGAP_Co	1.88e+01
c	904	14	16.1	531	22	AA163085	mt74b10.r1 Soares mous	1.88e+01
c	905	14	16.1	535	26	B63413	RPCI11-3H5.TVB RPCI11	1.88e+01
c	906	14	16.1	536	28	AQ208262	HS_3219_A1_C04_T7 CIT	1.88e+01
c	907	14	16.1	539	17	AA947589	ok21c09.s1 Soares_NSF_	1.88e+01
	908	14	16.1	540	23	AA230687	mv76b12.r1 Soares mous	1.88e+01
	909	14	16.1	541	18	AI151569	SWOvAPCAP04D08SK Oncho	1.88e+01

c	910	14	16.1	541	10	AA602289	np25a01.s1 NCI_CGAP_Pr	1.88e+01
	911	14	16.1	545	21	W73157	zd53c09.s1 Soares feta	1.88e+01
c	912	14	16.1	545	29	AQ234600	HS_2054_A2_F06_T7 CIT	1.88e+01
c	913	14	16.1	546	6	AA001316	zh82h10.r1 Soares feta	1.88e+01
c	914	14	16.1	546	23	AA185231	mt61d11.r1 Soares 2NbM	1.88e+01
	915	14	16.1	547	22	AA198617	mv40e06.r1 Soares mous	1.88e+01
c	916	14	16.1	551	6	AA029879	zk10g05.r1 Soares preg	1.88e+01
c	917	14	16.1	551	12	AA777457	zi93b06.s1 Soares feta	1.88e+01
	918	14	16.1	553	23	AA145395	mr09d07.r1 Soares mous	1.88e+01
c	919	14	16.1	553	10	AA426499	zw02b06.r1 Soares NhHM	1.88e+01
c	920	14	16.1	553	14	AA851969	EST194738 Normalized r	1.88e+01
	921	14	16.1	554	27	AQ081633	RPCI11-55E5.TJ RPCI11	1.88e+01
c	922	14	16.1	557	10	AA604760	no84f02.s1 NCI_CGAP_AA	1.88e+01
	923	14	16.1	559	29	AQ231315	HS_2053_A1_B08_MR CIT	1.88e+01
c	924	14	16.1	561	13	AA732701	ns97d03.s1 NCI_CGAP_GC	1.88e+01
c	925	14	16.1	562	19	N38737	yy42b01.s1 Homo sapien	1.88e+01
c	926	14	16.1	563	27	FR0028566	Fugu rubripes GSS sequ	1.88e+01
c	927	14	16.1	563	6	AA131511	zl31d03.s1 Soares preg	1.88e+01
	928	14	16.1	565	21	W31037	zb86c06.r1 Soares sene	1.88e+01
	929	14	16.1	568	27	AQ048818	cLM-10b6-t cLM Giardia	1.88e+01
	930	14	16.1	571	26	B56459	CIT-HSP-2006P13.TR CIT	1.88e+01
	931	14	16.1	577	27	AQ061956	CIT-HSP-2357K4.TF CIT-	1.88e+01
c	932	14	16.1	577	13	AA843410	ak07e11.s1 Soares para	1.88e+01
	933	14	16.1	583	21	W55954	zc58g04.r1 Soares para	1.88e+01
	934	14	16.1	583	19	N25431	yx76b05.s1 Homo sapien	1.88e+01
	935	14	16.1	583	27	FR0028520	Fugu rubripes GSS sequ	1.88e+01
	936	14	16.1	584	19	N25104	yx18f11.r1 Homo sapien	1.88e+01
c	937	14	16.1	586	13	C85344	Mus musculus fertilize	1.88e+01
c	938	14	16.1	586	26	B16617	342O18.TVB CIT978SKA1	1.88e+01
c	939	14	16.1	586	12	AA752178	96BS0251 Rice Immature	1.88e+01
	940	14	16.1	590	7	AA512532	vg40a10.r1 Soares mous	1.88e+01
	941	14	16.1	590	17	AU016303	Mus musculus 2-cell em	1.88e+01
	942	14	16.1	593	27	B95729	CIT-HSP-2169K8.TR CIT-	1.88e+01
c	943	14	16.1	593	26	CI5G2	Ciona intestinalis gen	1.88e+01
	944	14	16.1	594	17	AU021017	Mus musculus 8-cell em	1.88e+01
c	945	14	16.1	595	27	AQ113108	CIT-HSP-2374C23.TR CIT	1.88e+01
	946	14	16.1	598	9	AA534459	nf80d09.s1 NCI_CGAP_Co	1.88e+01
	947	14	16.1	599	28	AQ194434	RPCI11-46P5.TJ RPCI11	1.88e+01
c	948	14	16.1	599	26	B77327	T32J8TF TAMU Arabidops	1.88e+01
c	949	14	16.1	600	9	AA521434	aa69c04.s1 NCI_CGAP_GC	1.88e+01
	950	14	16.1	602	26	FR0009157	F.rubripes GSS sequenc	1.88e+01
c	951	14	16.1	603	18	AU035959	Mus musculus cDNA, clo	1.88e+01
	952	14	16.1	612	28	AQ200864	RPCI11-61A22.TK RPCI11	1.88e+01
c	953	14	16.1	615	20	W20085	zb40f02.r1 Soares para	1.88e+01
c	954	14	16.1	616	28	AQ111244	CIT-HSP-2381D17.TR CIT	1.88e+01
	955	14	16.1	618	27	AQ014018	RPCI11-24K14.TKBR RPCI	1.88e+01
	956	14	16.1	619	26	FR0025833	F.rubripes GSS sequenc	1.88e+01
c	957	14	16.1	620	28	AQ111499	CIT-HSP-2375C19.TF CIT	1.88e+01
	958	14	16.1	625	27	AQ081923	RPCI11-54A23.TJ RPCI11	1.88e+01
c	959	14	16.1	630	14	AA985932	uc72e06.x1 Sugano mous	1.88e+01
c	960	14	16.1	630	6	AA194164	zr37h04.s1 Soares NhHM	1.88e+01

c	961	14	16.1	631	14	AA968357	uc71e08.x1 Sugano mous 1.88e+01
	962	14	16.1	635	15	C94444	Dictyostelium discoide 1.88e+01
c	963	14	16.1	639	22	AA117952	mn08h01.r1 Beddington 1.88e+01
c	964	14	16.1	641	7	AA536750	vj88d04.r1 Knowles Sol 1.88e+01
c	965	14	16.1	646	26	B18092	A-306G8.TP CIT978SK Ho 1.88e+01
c	966	14	16.1	648	26	B28868	T21D23TR TAMU Arabidop 1.88e+01
c	967	14	16.1	648	11	AA694645	ET2235 Trypanosoma bru 1.88e+01
c	968	14	16.1	650	18	AI172187	EST218182 Normalized r 1.88e+01
c	969	14	16.1	651	27	AQ014083	RPCI11-25K13.TKBR RPCI 1.88e+01
c	970	14	16.1	663	12	AA753507	96BS0661 Rice Immature 1.88e+01
c	971	14	16.1	673	26	B48072	RPCI11-4K7.TV RPCI11 H 1.88e+01
	972	14	16.1	674	26	AG004651	Homo sapiens genomic D 1.88e+01
c	973	14	16.1	676	28	AQ157412	nbxb0009M10r CUGI Rice 1.88e+01
c	974	14	16.1	676	26	B16731	343E11.TVB CIT978SKA1 1.88e+01
	975	14	16.1	680	27	AG010992	Homo sapiens genomic D 1.88e+01
	976	14	16.1	686	27	AG010991	Homo sapiens genomic D 1.88e+01
	977	14	16.1	693	15	coau0004016 Cotton Bol 1.88e+01	
	978	14	16.1	695	27	B92341	CIT-HSP-2173D17.TR CIT 1.88e+01
c	979	14	16.1	697	28	AQ116144	RPCI11-42C6.TK.1 RPCI1 1.88e+01
	980	14	16.1	699	15	AA932217	om84h07.s1 NCI_CGAP_Ki 1.88e+01
c	981	14	16.1	701	26	B24108	F18L10TF IGF Arabidops 1.88e+01
c	982	14	16.1	701	27	AQ039459	CIT-HSP-2327A21.TF CIT 1.88e+01
	983	14	16.1	725	26	AG007362	Homo sapiens genomic D 1.88e+01
c	984	14	16.1	725	21	W80986	zd84e08.r1 Soares feta 1.88e+01
	985	14	16.1	726	13	AA884109	am36e07.s1 Soares NFL 1.88e+01
c	986	14	16.1	734	15	AI046423	ud65f02.x1 Sugano mous 1.88e+01
c	987	14	16.1	734	15	AI046450	ud65h07.x1 Sugano mous 1.88e+01
	988	14	16.1	740	18	AI056274	oz02f03.x1 Soares_feta 1.88e+01
	989	14	16.1	742	27	AQ061960	CIT-HSP-2357K6.TF CIT- 1.88e+01
c	990	14	16.1	743	28	AQ160883	nbxb0006K05f CUGI Rice 1.88e+01
c	991	14	16.1	768	27	AQ050402	nbxb0003dH11f CUGI Ric 1.88e+01
	992	14	16.1	778	26	B05772	cSRL-71a5-u cSRL flow 1.88e+01
c	993	14	16.1	824	17	AI069212	mgae0005dG09f Magnapor 1.88e+01
c	994	14	16.1	839	26	B11474	F13P24-T7 IGF Arabidop 1.88e+01
	995	14	16.1	931	27	AQ072958	EP(3)3481 Drosophila m 1.88e+01
c	996	14	16.1	959	17	AI068534	mgae0002dF07f Magnapor 1.88e+01
	997	14	16.1	982	21	W30569	mc28a06.r1 Soares mous 1.88e+01
c	998	14	16.1	1115	26	B10557	F14M19-T7 IGF Arabidop 1.88e+01
	999	14	16.1	1241	26	B10118	T25P11-Sp6.1 TAMU Arab 1.88e+01
c	1000	14	16.1	2109	20	W18181	IMAGE:20099 Soares inf 1.88e+01

ALIGNMENTS

```
RESULT    1
LOCUS       HUMGS03628     88 bp    mRNA            EST       26-AUG-1994
DEFINITION  Human aortic endothel 3'directed MboI cDNA, HUMGS03628, clone
            ae0424, mRNA sequence.
ACCESSION   D26035
NID         g500365
KEYWORDS    EST; EST(expressed sequence tag); endothel; gene signature(GS).
```



```
SOURCE      Homo sapiens adult primary culture of aortic endothel cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 88)
  AUTHORS   Chae,K., Murakawa,K., Okubo,K. and Matsubara,K.
  JOURNAL   Unpublished (1993)
REFERENCE   2  (bases 1 to 88)
  AUTHORS   Okubo,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-NOV-1993) to the DDBJ/EMBL/GenBank databases. Kousaku
            Okubo, Osaka University, Institute for Molecular and Cellular Bio;
            1-3,Yamada-oka, Suita, Osaka Pref. 565, Japan
            (E-mail:kousaku@imcb.osaka-u.ac.jp, Tel:06-877-5111(ex.3315),
            Fax:06-877-1922)
COMMENT     Submitted (22-Nov-1993) to DDBJ by:
            Kousaku Okubo
            Institute for Molecular and Cellular Biology
            Osaka University
            3-1, Yamadaoka
            Suita, Osaka, 565
            Japan
            Phone: 06-877-5111
            Fax  : 06-875-1922.
FEATURES             Location/Qualifiers
     source          1..88
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_type="primary culture of aortic endothel"
                     /dev_stage="adult"
BASE COUNT       34 a      7 c     15 g     31 t      1 others
ORIGIN

  Query Match          100.0%;  Score 87;  DB 7;  Length 88;
  Best Local Similarity 100.0%;  Pred. No. 2.90e-137;
  Matches    88;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 GATCAATCTTTGATGGTGAGGGTTTTAGAAAGGAAAAATATAGTAAAATGTGTAATTTGT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      1 GATCAATCTTTGATGGTGAGGGTTTTAGAAAGGAAAAATATAGTAAAATGTGTAATTTGT 60

Db     61 CTTAATAAAATCTATCTNTACATCTAAA 88
          ||||||||||||||||||||||||||||
Qy     61 CTTAATAAAATCTATCTNTACATCTAAA 88


RESULT    2
LOCUS       AI183912      284 bp    mRNA            EST       09-OCT-1998
DEFINITION  qe23e10.x1 Soares_fetal_lung_NbHL19W Homo sapiens cDNA clone
            IMAGE:1739850 3', mRNA sequence.
```



```
ACCESSION   AI183912
NID         g3734550
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 284)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40UP from Gibco
            High quality sequence stop: 275.
FEATURES             Location/Qualifiers
     source          1. .284
                     /organism="Homo sapiens"
                     /note="Organ: lung; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCAATTTTTTTTTTTTTTTTT-3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by Bento
                     Soares and M.Fatima Bonaldo. This library was constructed
                     from the same fetus as the fetal heart library, Soares
                     fetal heart NbHH19W."
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1739850"
                     /clone_lib="Soares_fetal_lung_NbHL19W"
                     /dev_stage="19 weeks"
                     /lab_host="DH10B (ampicillin resistant)"
BASE COUNT       95 a     67 c     35 g     87 t
ORIGIN

  Query Match          100.0%;  Score 87;  DB 18;  Length 284;
  Best Local Similarity  98.9%;  Pred. No. 2.90e-137;
  Matches    87;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db      1 TTTAGATGTAGAGATAGATTTTATTAAGACAAATTACACATTTTACTATATTTTTCCTTT 60
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     88 TTTAGATGTANAGATAGATTTTATTAAGACAAATTACACATTTTACTATATTTTTCCTTT 29
```



```
Db     61 CTAAAACCCTCACCATCAAAGATTGATC 88
          ||||||||||||||||||||||||||||
Cp     28 CTAAAACCCTCACCATCAAAGATTGATC 1


RESULT    3
LOCUS       AI184115      375 bp    mRNA            EST       09-OCT-1998
DEFINITION  qe24e12.x1 Soares_fetal_lung_NbHL19W Homo sapiens cDNA clone
            IMAGE:1739950 3', mRNA sequence.
ACCESSION   AI184115
NID         g3734753
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 375)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40UP from Gibco
            High quality sequence stop: 357.
FEATURES             Location/Qualifiers
     source          1. .375
                     /organism="Homo sapiens"
                     /note="Organ: lung; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCAATTTTTTTTTTTTTTTTT-3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by Bento
                     Soares and M.Fatima Bonaldo. This library was constructed
                     from the same fetus as the fetal heart library, Soares
                     fetal heart NbHH19W."
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1739950"
                     /clone_lib="Soares_fetal_lung_NbHL19W"
                     /dev_stage="19 weeks"
                     /lab_host="DH10B (ampicillin resistant)"
BASE COUNT      115 a     94 c     60 g    106 t
ORIGIN
```

```
  Query Match          100.0%;  Score 87;  DB 18;  Length 375;
  Best Local Similarity  98.9%;  Pred. No. 2.90e-137;
  Matches   87;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db      1 TTTAGATGTAGAGATAGATTTTATTAAGACAAATTACACATTTTACTATATTTTTCCTTT 60
          ||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     88 TTTAGATGTANAGATAGATTTTATTAAGACAAATTACACATTTTACTATATTTTTCCTTT 29

Db     61 CTAAAACCCTCACCATCAAAGATTGATC 88
          ||||||||||||||||||||||||||||
Cp     28 CTAAAACCCTCACCATCAAAGATTGATC 1


RESULT    4
LOCUS       AI017590      438 bp    mRNA            EST       27-AUG-1998
DEFINITION  ou23h06.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
            IMAGE:1627163 3', mRNA sequence.
ACCESSION   AI017590
NID         g3231926
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 438)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1550   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 371.
FEATURES              Location/Qualifiers
     source           1. .438
                      /organism="Homo sapiens"
                      /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                      a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                      Equal amounts of plasmid DNA from three normalized
                      libraries (fetal lung NbHL19W, testis NHT, and B-cell
                      NCI_CGAP_GCB1) were mixed, and ss circles were made in
                      vitro. Following HAP purification, this DNA was used as
                      tracer in a subtractive hybridization reaction. The driver
                      was PCR-amplified cDNAs from pools of 5,000 clones made
                      from the same 3 libraries. The pools consisted of
```



```
                      I.M.A.G.E. clones 297480-302087, 682632-687239,
                      726408-728711, and 729096-731399. Subtraction by Bento
                      Soares and M. Fatima Bonaldo."
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1627163"
                      /clone_lib="Soares_NFL_T_GBC_S1"
                      /lab_host="DH10B"
BASE COUNT      132 a    107 c     79 g    120 t
ORIGIN

  Query Match          100.0%;  Score 87;  DB 17;  Length 438;
  Best Local Similarity  98.9%;  Pred. No. 2.90e-137;
  Matches   87;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db      1 TTTAGATGTAGAGATAGATTTTATTAAGACAAATTACACATTTTACTATATTTTTCCTTT 60
          ||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     88 TTTAGATGTANAGATAGATTTTATTAAGACAAATTACACATTTTACTATATTTTTCCTTT 29

Db     61 CTAAAACCCTCACCATCAAAGATTGATC 88
          ||||||||||||||||||||||||||||
Cp     28 CTAAAACCCTCACCATCAAAGATTGATC 1


RESULT    5
LOCUS       AA496035      455 bp    mRNA            EST       11-AUG-1997
DEFINITION  zv72d06.s1 Soares total fetus Nb2HF8 9w Homo sapiens cDNA clone
            759179 3', mRNA sequence.
ACCESSION   AA496035
NID         g2229356
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 455)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J.,
            Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-Merck EST Project 1997
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
```



```
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 420.
FEATURES              Location/Qualifiers
     source           1. .455
                      /organism="Homo sapiens"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                      was prepared from mRNA obtained from pooled 8-9 week
                      (total) fetus material with a Not I - oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCTTAATTTTTTTTTTTTTTTT 3'].
                      Double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Not I and cloned into the Not I
                      and Eco RI sites of the modified pT7T3 vector. Library
                      went through one round of normalization, and was
                      constructed  by Bento Soares and M. Fatima Bonaldo."
                      /db_xref="taxon:9606"
                      /clone="759179"
                      /clone_lib="Soares total fetus Nb2HF8 9w"
                      /dev_stage="8-9 weeks"
                      /lab_host="DH10B"
     mRNA             complement(<1. .>455)
BASE COUNT      134 a    109 c     83 g    129 t
ORIGIN

  Query Match          100.0%;  Score 87;  DB 8;  Length 455;
  Best Local Similarity  98.9%;  Pred. No. 2.90e-137;
  Matches   87;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db      1 TTTAGATGTAGAGATAGATTTTATTAAGACAAATTACACATTTTACTATATTTTTCCTTT 60
          ||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     88 TTTAGATGTANAGATAGATTTTATTAAGACAAATTACACATTTTACTATATTTTTCCTTT 29

Db     61 CTAAAACCCTCACCATCAAAGATTGATC 88
          ||||||||||||||||||||||||||||
Cp     28 CTAAAACCCTCACCATCAAAGATTGATC 1


RESULT    6
LOCUS       AA032240      501 bp    mRNA            EST       10-MAY-1997
DEFINITION  zk19a02.s1 Soares pregnant uterus NbHPU Homo sapiens cDNA clone
            470954 3', mRNA sequence.
ACCESSION   AA032240
NID         g1502202
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 501)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
```



```
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
            Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1148   Std Error: 0.00
            Seq primer: -40M13 fwd. from Amersham
            High quality sequence stop: 428.
FEATURES              Location/Qualifiers
     source           1. .501
                      /organism="Homo sapiens"
                      /note="Organ: uterus; Vector: pT7T3-Pac; Site_1: Not I;
                      Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                      oligo(dT) primer [5'
                      AACTGGAAGAATTCGCGGCCGCCTTTTTTTTTTTTTTTT 3'],
                      double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Not I and cloned into the Not I
                      and Eco RI sites of the modified pT7T3 vector. Library
                      went through one round of normalization.  Library
                      constructed by M. Fatima Bonaldo."
                      /db_xref="taxon:9606"
                      /clone="470954"
                      /clone_lib="Soares pregnant uterus NbHPU"
                      /sex="female"
                      /dev_stage="adult"
                      /lab_host="DH10B"
     mRNA             complement(<1. .>501)
                      /db_xref="GDB:3756956"
BASE COUNT      141 a    122 c     87 g    146 t      5 others
ORIGIN

  Query Match          100.0%;  Score 87;  DB 6;  Length 501;
  Best Local Similarity  98.9%;  Pred. No. 2.90e-137;
  Matches   87;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db      5 TTTAGATGTAGAGATAGATTTTATTAAGACAAATTACACATTTTACTATATTTTTCCTTT 64
          ||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     88 TTTAGATGTANAGATAGATTTTATTAAGACAAATTACACATTTTACTATATTTTTCCTTT 29

Db     65 CTAAAACCCTCACCATCAAAGATTGATC 92
          ||||||||||||||||||||||||||||
```

```
Cp     28 CTAAAACCCTCACCATCAAAGATTGATC 1


RESULT    7
LOCUS       AA635684      324 bp    mRNA            EST       31-OCT-1997
DEFINITION  nr33c01.s1 NCI_CGAP_Pr22 Homo sapiens cDNA clone IMAGE:1169760,
            mRNA sequence.
ACCESSION   AA635684
NID         g2559526
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 324)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 402   Std Error: 0.00
            Seq primer: -40ml3 fwd. ET from Amersham.
FEATURES              Location/Qualifiers
     source           1. .324
                      /organism="Homo sapiens"
                      /note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
                      with a modified polylinker; 1st strand cDNA was prepared
                      from normal prostate bulk tissue, and was then primed with
                      a Not I - oligo(dT) primer. Double-stranded cDNA was
                      ligated to Eco RI adaptors (Pharmacia), digested with Not
                      I and cloned into the Not I  and Eco RI sites of the
                      modified pT7T3 vector.  Library is normalized, and was
                      constructed by Bento Soares and M. Fatima Bonaldo."
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1169760"
                      /clone_lib="NCI_CGAP_Pr22"
                      /sex="male"
                      /tissue_type="normal prostate"
                      /lab_host="DH10B"
```



```
BASE COUNT      107 a     76 c     50 g     91 t
ORIGIN

  Query Match           97.7%;  Score 85;  DB 10;  Length 324;
  Best Local Similarity  98.8%;  Pred. No. 4.81e-133;
  Matches    85;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db      1 TAGATGTAGAGATAGATTTTATTAAGACAAATTACACATTTTACTATATTTTTCCTTTCT 60
          ||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     86 TAGATGTANAGATAGATTTTATTAAGACAAATTACACATTTTACTATATTTTTCCTTTCT 27

Db     61 AAAACCCTCACCATCAAAGATTGATC 86
          ||||||||||||||||||||||||||
Cp     26 AAAACCCTCACCATCAAAGATTGATC 1


RESULT    8
LOCUS       N27289        385 bp    mRNA            EST       29-DEC-1995
DEFINITION  yw71b11.s1 Homo sapiens cDNA clone 257661 3'.
ACCESSION   N27289
NID         g1141637
KEYWORDS    EST.
SOURCE      human clone=257661 primer=ml3 -40 forward library=Soares placenta
            8to9weeks 2NbHP8to9W vector=pT7T3D (Pharmacia) with a modified
            polylinker host=DH10B (ampicillin resistant) Rsitel=Not I
            Rsite2=Eco RI two placentae: one from 8 weeks and another from 9
            weeks post conception. 1st strand cDNA was primed with a Not I -
            oligo(dT) primer
            [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCGATTTTTTTTTTTTTTTTT-3'],
            double-stranded cDNA was size selected, ligated to Eco RI adapters
            (Pharmacia), digested with Not I and cloned into the Not I and Eco
            RI sites of a modified pT7T3 vector (Pharmacia). Library
            constructed by Bento Soares and M. Fatima Bonaldo.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 385)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
```



```
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 305
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES              Location/Qualifiers
     source           1. .385
                      /organism="Homo sapiens"
                      /clone="257661"
     mRNA             <1. .>385
BASE COUNT      115 a     96 c     65 g    109 t
ORIGIN

  Query Match           97.7%;  Score 85;  DB 19;  Length 385;
  Best Local Similarity  98.8%;  Pred. No. 4.81e-133;
  Matches    85;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db      1 TAGATGTAGAGATAGATTTTATTAAGACAAATTACACATTTTACTATATTTTTCCTTTCT 60
          ||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     86 TAGATGTANAGATAGATTTTATTAAGACAAATTACACATTTTACTATATTTTTCCTTTCT 27

Db     61 AAAACCCTCACCATCAAAGATTGATC 86
          ||||||||||||||||||||||||||
Cp     26 AAAACCCTCACCATCAAAGATTGATC 1


RESULT    9
LOCUS       AA043991      416 bp    mRNA            EST       10-MAY-1997
DEFINITION  zk58h02.s1 Soares pregnant uterus NbHPU Homo sapiens cDNA clone
            487059 3', mRNA sequence.
ACCESSION   AA043991
NID         g1521849
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 416)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
            Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
```



```
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 772   Std Error: 0.00
            Seq primer: -40M13 fwd. from Amersham
            High quality sequence stop: 370.
FEATURES              Location/Qualifiers
     source           1. .416
                      /organism="Homo sapiens"
                      /note="Organ: uterus; Vector: pT7T3-Pac; Site_1: Not I;
                      Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                      oligo(dT) primer [5'
                      AACTGGAAGAATTCGCGGCCGCCTTTTTTTTTTTTTTTT 3'],
                      double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Not I and cloned into the Not I
                      and Eco RI sites of the modified pT7T3 vector. Library
                      went through one round of normalization.  Library
                      constructed by M. Fatima Bonaldo."
                      /db_xref="taxon:9606"
                      /clone="487059"
                      /clone_lib="Soares pregnant uterus NbHPU"
                      /sex="female"
                      /dev_stage="adult"
                      /lab_host="DH10B"
     mRNA             complement(<1. .>416)
                      /db_xref="GDB:3760773"
BASE COUNT      124 a    103 c     71 g    117 t      1 others
ORIGIN

  Query Match           96.6%;  Score 84;  DB 6;  Length 416;
  Best Local Similarity  98.8%;  Pred. No. 6.15e-131;
  Matches    84;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db      1 AGATGTAGAGATAGATTTTATTAAGACAAATTACACATTTTACTATATTTTTCCTTTCTA 60
          ||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     85 AGATGTANAGATAGATTTTATTAAGACAAATTACACATTTTACTATATTTTTCCTTTCTA 26

Db     61 AAACCCTCACCATCAAAGATTGATC 85
          |||||||||||||||||||||||||
Cp     25 AAACCCTCACCATCAAAGATTGATC 1


RESULT   10
LOCUS       H62857        422 bp    mRNA            EST       10-OCT-1995
DEFINITION  yr18b08.s1 Homo sapiens cDNA clone 205623 3'.
ACCESSION   H62857
NID         g1017203
KEYWORDS    EST.
```

```
SOURCE      human clone=205623 primer=Promega -21m13 library=Soares fetal liver
            spleen 1NFLS vector=pT7T3D (Pharmacia) with a modified polylinker
            host=DH10B (ampicillin resistant) Rsitel=Pac I Rsite2=Eco RI Liver
            and spleen from a 20 week-post conception male fetus. 1st strand
            cDNA was primed with a Pac I - oligo(dT) primer (5'
            AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTT 3'), double-stranded
            cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac
            I and cloned into the Pac I and Eco RI sites of the modified pT7T3
            vector.  Library went through one round of normalization. Library
            constructed by Bento Soares and M.Fatima Bonaldo.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 422)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 279
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES              Location/Qualifiers
     source           1. .422
                      /organism="Homo sapiens"
                      /clone="205623"
     mRNA             <1. .>422
BASE COUNT      117 a     99 c     86 g    117 t      3 others
ORIGIN

  Query Match           96.6%;  Score 84;  DB 19;  Length 422;
  Best Local Similarity  98.8%;  Pred. No. 6.15e-131;
  Matches    84;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db      1 AGATGTAGAGATAGATTTTATTAAGACAAATTACACATTTTACTATATTTTTCCTTTCTA 60
          ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     85 AGATGTANAGATAGATTTTATTAAGACAAATTACACATTTTACTATATTTTTCCTTTCTA 26
```



```
Db     61 AAACCCTCACCATCAAAGATTGATC 85
          |||||||||||||||||||||||||
Cp     25 AAACCCTCACCATCAAAGATTGATC 1


RESULT   11
LOCUS       AA700634      562 bp    mRNA            EST       19-DEC-1997
DEFINITION  zi43b04.s1 Soares fetal liver spleen 1NFLS S1 Homo sapiens cDNA
            clone 433519 3', mRNA sequence.
ACCESSION   AA700634
NID         g2703597
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 562)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
            Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
            Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 461.
FEATURES              Location/Qualifiers
     source           1. .562
                      /organism="Homo sapiens"
                      /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                      with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                      This is a subtracted version of the original Soares fetal
                      liver spleen 1NFLS library.   1st strand cDNA was primed
                      with a Pac I - oligo(dT) primer (5'
                      AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTT 3'),
                      double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Pac I and cloned into the Pac I
                      and Eco RI sites of the modified pT7T3 vector.  Library
                      went through one round of normalization. Library
                      constructed by Bento Soares and M.Fatima Bonaldo."
                      /db_xref="GDB:1333288"
                      /db_xref="taxon:9606"
                      /clone="433519"
```



```
                      /clone_lib="Soares fetal liver spleen 1NFLS S1"
                      /sex="male"
                      /dev_stage="20 week-post conception fetus"
                      /lab_host="DH10B (ampicillin resistant)"
BASE COUNT      164 a    145 c     93 g    159 t      1 others
ORIGIN

  Query Match           96.6%;  Score 84;  DB 11;  Length 562;
  Best Local Similarity  98.8%;  Pred. No. 6.15e-131;
  Matches    84;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db      1 AGATGTAGAGATAGATTTTATTAAGACAAATTACACATTTTACTATATTTTTCCTTTCTA 60
          ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     85 AGATGTANAGATAGATTTTATTAAGACAAATTACACATTTTACTATATTTTTCCTTTCTA 26

Db     61 AAACCCTCACCATCAAAGATTGATC 85
          |||||||||||||||||||||||||
Cp     25 AAACCCTCACCATCAAAGATTGATC 1


RESULT   12
LOCUS       AA347704      414 bp    mRNA            EST       21-APR-1997
DEFINITION  EST54313 Fetal heart II Homo sapiens cDNA 3' end, mRNA sequence.
ACCESSION   AA347704
NID         g2000170
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 414)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
            Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
            White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
            Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
            Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
            Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
            Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,T.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
            He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
            Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H.,
            Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
            Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
            Fraser,C.M. and Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
```



```
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   96026280
COMMENT     Other_ESTs: EST54314 THC180445
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13-21.
FEATURES              Location/Qualifiers
     source           1. .414
                      /organism="Homo sapiens"
                      /note="Organ: heart; Vector: pBluescript SK-; Site_1:
                      EcoRI; Site_2: XhoI"
                      /db_xref="ATCC (inhost):149009"
                      /db_xref="taxon:9606"
                      /clone_lib="Fetal heart II"
                      /dev_stage="fetus"
     mRNA             complement(<1. .>414)
BASE COUNT      121 a    102 c     70 g    113 t      8 others
ORIGIN

  Query Match           94.3%;  Score 82;  DB 25;  Length 414;
  Best Local Similarity  96.5%;  Pred. No. 9.87e-127;
  Matches    82;  Conservative    0;  Mismatches   3;  Indels   0;  Gaps   0;

Db      1 AGATGTAGAGATAGATTTTATTAAGACAAATTACACATTTNACTATATTTTNCCTTTCTA 60
          ||||||| ||||||||||||||||||||||||||||||||| |||||||||| ||||||||
Cp     85 AGATGTANAGATAGATTTTATTAAGACAAATTACACATTTTACTATATTTTTCCTTTCTA 26

Db     61 AAACCCTCACCATCAAAGATTGATC 85
          |||||||||||||||||||||||||
Cp     25 AAACCCTCACCATCAAAGATTGATC 1


RESULT   13
LOCUS       T57478        307 bp    mRNA            EST       08-FEB-1995
DEFINITION  yb56d02.s1 Homo sapiens cDNA clone 75171 3'.
ACCESSION   T57478
NID         g659339
KEYWORDS    EST.
SOURCE      human clone=75171 library=Stratagene ovary (#937217)
            vector=Bluescript SK host=SOLR cells (kanamycin resistant)
            primer=-21m13 Rsite1=EcoRI Rsite2=XhoI Cloned unidirectionally.
            Primer: Oligo dT. Total ovary tissue, normal 49 year old caucasian
            female. Average insert size: 0.8 kb; Uni-ZAP XR Vector; 5' adaptor
```

```
            sequence: 5'-GAATTCGGCACGAG-3'; 3' adaptor sequence:
            5'-CTCGAGTTTTTTTTTTTTTTTTT-3'.
  ORGANISM  Homo sapiens
            Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 307)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
            Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: estewatson.wustl.edu
            High qality sequence stops: 280
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES              Location/Qualifiers
     source           1. .307
                      /organism="Homo sapiens"
                      /clone="75171"
BASE COUNT      102 a      75 c      45 g      84 t       1 others
ORIGIN

  Query Match           87.4%;  Score 76;  DB 5;  Length 307;
  Best Local Similarity  98.7%;  Pred. No. 3.54e-14;
  Matches    76;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db       1 AGATAGATNTTATTAAGACAAATTACACATTTTACTATATTTTTCCTTTCTAAAACCCTC 60
           |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp      77 AGATAGATTTTATTAAGACAAATTACACATTTTACTATATTTTTCCTTTCTAAAACCCTC 18

Db      61 ACCATCAAAGATTGATC 77
           |||||||||||||||||
Cp      17 ACCATCAAAGATTGATC 1


RESULT   14
LOCUS       AA928824      317 bp    mRNA            EST       23-JUN-1998
DEFINITION  on99c05.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
            IMAGE:1564808 3', mRNA sequence.
ACCESSION   AA928824
NID         g3078181
KEYWORDS    EST.
```



```
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 317)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 810   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 268.
FEATURES              Location/Qualifiers
     source           1. .317
                      /organism="Homo sapiens"
                      /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                      a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                      Equal amounts of plasmid DNA from three normalized
                      libraries (fetal lung NbHL19W, testis NHT, and B-cell
                      NCI_CGAP_GCB1) were mixed, and ss circles were made in
                      vitro. Following HAP purification, this DNA was used as
                      tracer in a subtractive hybridization reaction. The driver
                      was PCR-amplified cDNAs from pools of 5,000 clones made
                      from the same 3 libraries. The pools consisted of
                      I.M.A.G.E. clones 297480-302087, 682632-687239,
                      726408-728711, and 729096-731399. Subtraction by Bento
                      Soares and M. Fatima Bonaldo."
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1564808"
                      /clone_lib="Soares_NFL_T_GBC_S1"
                      /lab_host="DH10B"
BASE COUNT      105 a      77 c      45 g      89 t       1 others
ORIGIN

  Query Match           65.5%;  Score 57;  DB 15;  Length 317;
  Best Local Similarity 100.0%;  Pred. No. 3.30e-75;
  Matches    57;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      32 AATTACACATTTTACTATATTTTTCCTTTCTAAAACCCTCACCATCAAAGATTGATC 88
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp      57 AATTACACATTTTACTATATTTTTCCTTTCTAAAACCCTCACCATCAAAGATTGATC 1


RESULT   15
LOCUS       AA862584      452 bp    mRNA            EST       24-AUG-1998
```



```
DEFINITION  oh44e12.s1 NCI_CGAP_GC4 Homo sapiens cDNA clone IMAGE:1469518 3',
            mRNA sequence.
ACCESSION   AA862584
NID         g2955063
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 452)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 1964   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 431.
FEATURES              Location/Qualifiers
     source           1. .452
                      /organism="Homo sapiens"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; 1st strand cDNA was prepared from 3 pooled
                      germ cell tumors, and was then primed with a Not I -
                      oligo(dT) primer. Double-stranded cDNA was ligated to Eco
                      RI adaptors (Pharmacia), digested with Not I and cloned
                      into the Not I and Eco RI sites of the modified pT7T3
                      vector.  Library is normalized. Library was constructed by
                      Bento Soares and M. Fatima Bonaldo."
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1469518"
                      /clone_lib="NCI_CGAP_GC4"
                      /tissue_type="pooled germ cell tumors"
                      /lab_host="DH10B"
BASE COUNT      158 a      89 c      75 g     130 t
ORIGIN

  Query Match           59.8%;  Score 52;  DB 17;  Length 452;
  Best Local Similarity  98.1%;  Pred. No. 3.20e-65;
```



```
  Matches    52;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db     400 TAGATGTAGAGATAGATTTTATTAAGACAAATTACACATTTTACTATATTTTT 452
           |||||||| ||||||||||||||||||||||||||||||||||||||||||||
Cp      86 TAGATGTANAGATAGATTTTATTAAGACAAATTACACATTTTACTATATTTTT 34


RESULT   16
LOCUS       N55979        303 bp    mRNA            EST       20-FEB-1996
DEFINITION  J4995F Homo sapiens cDNA clone J4995 5'.
ACCESSION   N55979
NID         g1198827
KEYWORDS    EST.
SOURCE      human clone=J4995 primer=TCCAAAGAATTCGGCACGAG library=Fetal heart,
            Lambda ZAP Express vector=Lambda ZAP Express host=E. coli XL1-Blue
            Rsite1=EcoRI Rsite2=XhoI mRNA was purified from human fetal hearts
            (8-10 weeks).  cDNA was synthesized using a XhoI-Oligo dT
            adaptor-primer.  EcoRI adaptors were ligated, followed by digestion
            with XhoI, for directional cloning into predigested lambda ZAP
            Express.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 303)
  AUTHORS   Liew,C.C.
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Liew CC
            Molecular Cardiology
            University of Toronto
            Banting Institute, 100 College St., Toronto, Ontario, M5G1L5
            Tel: 4169788758
            Fax: 4169785650
            Email: liewcc@utcc.utoronto.ca.
FEATURES              Location/Qualifiers
     source           1. .303
                      /organism="Homo sapiens"
                      /clone="J4995"
     mRNA             <1. .>303
BASE COUNT       77 a      55 c      82 g      89 t
ORIGIN

  Query Match           24.1%;  Score 21;  DB 19;  Length 303;
  Best Local Similarity 100.0%;  Pred. No. 6.76e-09;
  Matches    21;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     283 GATCAATCTTTGATGGTGAGG 303
           |||||||||||||||||||||
Qy       1 GATCAATCTTTGATGGTGAGG 21
```

```
RESULT   17
LOCUS       B07885          772 bp    DNA             GSS       12-MAR-1998
DEFINITION  YX87T7 pBluescript expression library of genomic DNA from OG1RF
            Enterococcus faecalis genomic clone YX87 similar to LPLB protein of
            Bacillus subtilis (LL9163), genomic survey sequence.
ACCESSION   B07885
NID         g2058775
KEYWORDS    GSS.
SOURCE      Enterococcus faecalis.
  ORGANISM  Enterococcus faecalis
            Eubacteria; Firmicutes; clostridial firmicutes; Enterococcus.
REFERENCE   1  (bases 1 to 772)
  AUTHORS   Xu,Y., Jiang,L., Murray,B.E. and Weinstock,G.M.
  TITLE     Enterococcus faecalis antigens in human infections
  JOURNAL   Infect. Immun. 65 (10), 4207-4215 (1997)
  MEDLINE   97461347
COMMENT     Contact: Weinstock, GM
            Department of Biochemistry and Molecular Biology
            University of Texas Medical School
            6431 Fannin Street, Houston, TX 77030, USA
            Tel: 713 500 6083
            Fax: 713 500 0652
            Email: georgew@utmmg.med.uth.tmc.edu
            Seq primer: pBluescript SK (-) t7
            Class: shotgun
            High quality sequence stop: 700.
FEATURES             Location/Qualifiers
     source          1. .772
                     /organism="Enterococcus faecalis"
                     /strain="OG1RF"
                     /note="Vector: pBluescript SK(-); Site_1: HindIII; Site_2:
                     EcoRI; Genomic DNA from an Enterococcus faecalis strain
                     OG1RF was partially digested with Sau3AI, sized
                     fractionated and ligated into the BamHI site of cosmid
                     vectors pLAFRx and pBeloBAC11. After in vitro packaging
                     and plating, the cosmid clones were screened with antisera
                     from 4 enterococcal endocarditis patients and a rabbit
                     immunized with surface proteins from an E. faecalis
                     isolate. Cosmid DNA from immunopositive cosmid clones was
                     treated with DNase I, the ends were filled with the Klenow
                     fragment, and then ligated into the EcoRV site in
                     pBluescript SK(-). The ligation mixture was transformed
                     into XL1Blue MRF', and the transformants were screened
                     with antisera. The immunopositive clones were selected for
                     DNA sequencing."
                     /db_xref="taxon:1351"
                     /clone="YX87"
                     /clone_lib="pBluescript expression library of genomic DNA
```



```
                     from OG1RF"
                     /lab_host="XL1Blue MRF'"
BASE COUNT      266 a    142 c    118 g    230 t     16 others
ORIGIN

  Query Match           21.8%;  Score 19;  DB 26;  Length 772;
  Best Local Similarity 100.0%;  Pred. No. 6.04e-06;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    654 ATAGTAAAATGTGTAATTT 672
          |||||||||||||||||||
Qy     40 ATAGTAAAATGTGTAATTT 58


RESULT   18
LOCUS       AA368830        352 bp    mRNA            EST       21-APR-1997
DEFINITION  EST80142 Placenta I Homo sapiens cDNA similar to insulin-like
            growth factor binding protein 3, mRNA sequence.
ACCESSION   AA368830
NID         g2021148
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 352)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
            Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
            White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
            Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
            Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
            Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
            Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
            He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
            Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H.,
            Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
            Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
            Fraser,C.M. and Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   96026280
COMMENT     Other_ESTs: THC168514
            Contact: Kerlavage, AR
            Bioinformatics
```



```
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .352
                     /organism="Homo sapiens"
                     /note="Organ: placenta; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: EcoRI"
                     /db_xref="ATCC (inhost):173527"
                     /db_xref="taxon:9606"
                     /clone_lib="Placenta I"
                     /tissue_type="placenta"
                     /dev_stage="fetus"
     mRNA            <1. .>352
BASE COUNT      116 a     39 c     61 g    134 t      2 others
ORIGIN

  Query Match           20.7%;  Score 18;  DB 25;  Length 352;
  Best Local Similarity 100.0%;  Pred. No. 1.56e-04;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    172 ATTTTATTAAGACAAATT 189
          ||||||||||||||||||
Cp     71 ATTTTATTAAGACAAATT 54


RESULT   19
LOCUS       B72729          402 bp    DNA             GSS       06-JUN-1998
DEFINITION  RPCI11-10C13.TV RPCI11 Homo sapiens genomic clone R-10C13, genomic
            survey sequence.
ACCESSION   B72729
NID         g2711880
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 402)
  AUTHORS   Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K.,
            Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences for Sequence-Ready Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: RPCI11-10C13.TP
            Contact: Mark Adams
```



```
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            For clone availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .402
                     /organism="Homo sapiens"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
                     /db_xref="taxon:9606"
                     /clone="R-10C13"
                     /clone_lib="RPCI11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
BASE COUNT      157 a     39 c     94 g    112 t
ORIGIN

  Query Match           20.7%;  Score 18;  DB 26;  Length 402;
  Best Local Similarity 100.0%;  Pred. No. 1.56e-04;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    130 TTTTAGAAAGGAAAAATA 147
          ||||||||||||||||||
Qy     23 TTTTAGAAAGGAAAAATA 40


RESULT   20
LOCUS       AA415844        565 bp    mRNA            EST       16-OCT-1997
DEFINITION  vd26g11.s1 Knowles Solter mouse 2 cell Mus musculus cDNA clone
            793700 5' similar to gb:L22338 Mouse MHC class I cell surface
            glycoprotein mRNA, 3' end (MOUSE);, mRNA sequence.
ACCESSION   AA415844
NID         g2076091
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 565)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
```

```
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:478108
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 494.
FEATURES              Location/Qualifiers
     source           1. .565
                      /organism="Mus musculus"
                      /strain="B6D2 F1/J"
                      /note="Organ: embryo; Vector: pBluescribe (modified);
                      Site_1: MluI; Site_2: SalI; Cloned unidirectionally from
                      mRNA prepared from 13,500 2-cell stage  embryos.  Primer:
                      SalI(dT): 5'-CGGTCGACCGTCGACCGTTTTTTTTTTTT-3'.  cDNAs
                      were cloned into the MluI/SalI sites of a modified
                      pBluescribe  vector using commercial linkers (NEB).
                      Average insert size:  1.2 kb."
                      /db_xref="taxon:10090"
                      /clone="793700"
                      /clone_lib="Knowles Solter mouse 2 cell"
                      /tissue_type="embryo"
                      /dev_stage="2-cell"
                      /lab_host="DH10B"
BASE COUNT      186 a      95 c     150 g     134 t
ORIGIN

  Query Match           20.7%;  Score 18;  DB 10;  Length 565;
  Best Local Similarity 100.0%;  Pred. No. 1.56e-04;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db       8 TTAGAAAGGAAAAATATA 25
           ||||||||||||||||||
Qy      25 TTAGAAAGGAAAAATATA 42


RESULT   21
LOCUS       C84145         612 bp    mRNA            EST       27-DEC-1997
DEFINITION  Dictyostelium discoideum slug cDNA, clone SSC330, mRNA sequence.
ACCESSION   C84145
NID         g2723751
KEYWORDS    EST; EST(expressed sequence tag).
```



```
SOURCE      Dictyostelium discoideum (strain:AX4) slug cDNA to mRNA,
            clone_lib:SS.
  ORGANISM  Dictyostelium discoideum
            Eukaryotae; Dictyosteliida; Dictyostelium.
REFERENCE   1  (sites)
  AUTHORS   Yoshida,M., Morio,T. and Urushihara,H.
  TITLE     Developmental cDNA in Dictyostelium discoideum(81)
  JOURNAL   Unpublished (1997)
REFERENCE   2  (bases 1 to 612)
  AUTHORS   Urushihara,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-DEC-1997) to the DDBJ/EMBL/GenBank databases. Hideko
            Urushihara, University of Tsukuba, Insstitute of Biological
            Sciences; 1-1-1 Ten-noudai, Tsukuba-shi 305-0006, Japan
            (E-mail:qxurushi@bank.dna.affrc.go.jp, Tel:+81-0298-53-4664,
            Fax:+81-0298-53-6614)
COMMENT     PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES              Location/Qualifiers
     source           1. .612
                      /organism="Dictyostelium discoideum"
                      /strain="AX4"
                      /db_xref="taxon:44689"
                      /clone_lib="SS"
                      /dev_stage="slug"
BASE COUNT      262 a      86 c      73 g     189 t       2 others
ORIGIN

  Query Match           20.7%;  Score 18;  DB 11;  Length 612;
  Best Local Similarity 94.7%;  Pred. No. 1.56e-04;
  Matches   18;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db     163 AATAAAATCTATCTTTACA 181
           |||||||||||||| ||||
Qy      64 AATAAAATCTATCTNTACA 82


RESULT   22
LOCUS       W14395         150 bp    mRNA            EST       02-OCT-1997
DEFINITION  ma71h08.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone 316191
            5', mRNA sequence.
ACCESSION   W14395
NID         g1283140
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 150)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
```



```
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI: 206807
            Seq primer: mob.REGA+ET
            High quality sequence stop: 139.
FEATURES              Location/Qualifiers
     source           1. .150
                      /organism="Mus musculus"
                      /note="Vector: pT7T3D (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                      was primed with a Not I - oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCATTTTTTTTTTTTTTTTTT 3'],
                      double-stranded cDNA was size selected, ligated to Eco RI
                      adapters (Pharmacia), digested with Not I and cloned into
                      the Not I and Eco RI sites of a modified pT7T3 vector
                      (Pharmacia). Library went through one round of
                      normalization to a Cot = 5. Library constructed by Bento
                      Soares and M.Fatima Bonaldo. RNA was kindly provided by
                      Dr. Minoru Ko (Wayne State University)."
                      /db_xref="taxon:10090"
                      /clone="316191"
                      /clone_lib="Soares mouse p3NMF19.5"
                      /dev_stage="19.5 dpc total fetus"
                      /lab_host="DH10B (ampicillin resistant)"
BASE COUNT       34 a      46 c      30 g      40 t
ORIGIN

  Query Match           19.5%;  Score 17;  DB 10;  Length 150;
  Best Local Similarity 100.0%;  Pred. No. 3.57e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      15 GTTTTAGAAAGGAAAAA 31
           |||||||||||||||||
Qy      22 GTTTTAGAAAGGAAAAA 38


RESULT   23
LOCUS       HUM163F05B     304 bp    mRNA            EST       19-DEC-1995
DEFINITION  Human fetal brain cDNA 5'-end GEN-163F05, mRNA sequence.
```



```
ACCESSION   D81426
NID         g1179303
KEYWORDS    EST; EST(expressed sequence tag); Human fetal brain.
SOURCE      Homo sapiens Fetus brain cDNA to mRNA, clone_lib:human fetal brain.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 304)
  AUTHORS   Fujiwara,T., Hirano,H., Katagiri,T., Kawai,A., Kuga,Y., Nagata,M.,
            Okuno,S., Osaki,K., Shimizu,F., Shimada,Y., Shinomiya,H.,
            Takaichi,A., Takeda,S., Watanabe,T., Takahashi,E., Hirai,Y.,
            Maekawa,H., Shin,S. and Nakamura,Y.
  TITLE     Large-scale sequencing project at Otsuka GEN Research Institute
  JOURNAL   Unpublished (1995)
REFERENCE   2  (bases 1 to 304)
  AUTHORS   Fujiwara,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-NOV-1995) to the DDBJ/EMBL/GenBank databases. Tsutomu
            Fujiwara, Otsuka GEN Research Institute,Otsuka Pharmaceutical
            Co.,Ltd, 463-10 Kagasuno Kawauchi-cho, Tokushima, Tokushima 771-01,
            Japan
COMMENT     Submitted (7-Nov-1995) to DDBJ by:
            Tsutomu Fujiwara
            Otsuka GEN Research Institute
            Otsuka Pharmaceutical CO.,Ltd
            463-10 Kagasuno Kawauchi-cho
            Tokushima, Tokushima
            771-01
            Japan
            Phone:0886-65-2888
            Fax  :0886-37-1035.
FEATURES              Location/Qualifiers
     source           1. .304
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone_lib="human fetal brain"
                      /dev_stage="Fetus"
                      /tissue_type="brain"
BASE COUNT      101 a      59 c      40 g     102 t       2 others
ORIGIN

  Query Match           19.5%;  Score 17;  DB 7;  Length 304;
  Best Local Similarity 100.0%;  Pred. No. 3.57e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      90 AATTACACATTTTACTA 106
           |||||||||||||||||
Cp      57 AATTACACATTTTACTA 41
```

```
RESULT   24
LOCUS       B72775         313 bp    DNA             GSS       06-JUN-1998
DEFINITION  RPCI11-10E13.TV RPCI11 Homo sapiens genomic clone R-10E13, genomic
            survey sequence.
ACCESSION   B72775
NID         g2711926
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 313)
  AUTHORS   Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K.,
            Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences for Sequence-Ready Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: RPCI11-10E13.TP
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            For clone availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .313
                     /organism="Homo sapiens"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
                     /db_xref="taxon:9606"
                     /clone="R-10E13"
                     /clone_lib="RPCI11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
BASE COUNT      105 a      35 c      78 g      95 t
ORIGIN

  Query Match           19.5%;  Score 17;  DB 26;  Length 313;
  Best Local Similarity 100.0%;  Pred. No. 3.57e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    109 TTTAGAAAGGAAAAATA 125
          |||||||||||||||||
Qy     24 TTTAGAAAGGAAAAATA 40
```



```
RESULT   25
LOCUS       H36017         317 bp    mRNA            EST       30-DEC-1997
DEFINITION  14539 Lambda-PRL2 Arabidopsis thaliana cDNA clone 172K8T7, mRNA
            sequence.
ACCESSION   H36017
NID         g2733426
KEYWORDS    EST.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryotae; Viridiplantae; Charophyta/Embryophyta group;
            Embryophyta; Tracheophyta; seed plants; Magnoliophyta;
            Magnoliopsida; Capparales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 317)
  AUTHORS   Newman,T., deBruijn,F.J., Green,P., Keegstra,K., Kende,H.,
            McIntosh,L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomashow,M.,
            Retzel,E. and Somerville,C.
  TITLE     Genes galore: a summary of methods for accessing results from
            large-scale partial sequencing of anonymous Arabidopsis cDNA clones
  JOURNAL   Plant Physiol. 106, 1241-1255 (1994)
  MEDLINE   95148729
COMMENT     On Dec 31, 1997 this sequence version replaced gi:905516.

            Contact: Thomas Newman
            MSU-DOE Plant Research Laboratory
            Michigan State University
            MSU-DOE-PRL, Michigan State University,Plant Biology Bldg.,E.
            Lansing,MI
            Tel: 517-353-0854
            Fax: 517-353-9168
            Email: 22313tcn@ibm.cl.msu.edu
            Seq primer: T7 dye primer.
FEATURES             Location/Qualifiers
     source          1. .317
                     /organism="Arabidopsis thaliana"
                     /strain="var columbia"
                     /note="Vector: lambda Zip-Lox; Site_1: Sal; Site_2: Not;
                     Lambda PRL2 is a cDNA library derived from equal
                     quantities of 4 pools of mRNA.  The mRNA sources were 1) 7
                     day germinated etiolated seedlings; 2) tissue culture
                     grown roots; 3) staged plants half with 24 hour light
                     cycle, half on 16 hr light, 8 hour dark- rosettes;  4)
                     same plants as 3 but aerial tissue (stems, flowers and
                     siliques.  The vector is BRL's lambda Zip-Lox.  The cDNA
                     inserts were directionally cloned with Sal-Not arms using
                     oligo dT primed cDNA."
                     /db_xref="taxon:3702"
                     /clone="172K8T7"
                     /clone_lib="Lambda-PRL2"
BASE COUNT       88 a      68 c      60 g      89 t      12 others
```



```
ORIGIN

  Query Match           19.5%;  Score 17;  DB 11;  Length 317;
  Best Local Similarity 100.0%;  Pred. No. 3.57e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    276 TTTCCTTTCTAAAACCC 292
          |||||||||||||||||
Cp     36 TTTCCTTTCTAAAACCC 20


RESULT   26
LOCUS       AA213510       335 bp    mRNA            EST       13-AUG-1997
DEFINITION  zr92a04.r1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:683118 5',
            mRNA sequence.
ACCESSION   AA213510
NID         g1812127
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 335)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -28m13 rev2 ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .335
                     /organism="Homo sapiens"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from human tonsillar cells enriched for
                     germinal center B cells by flow sorting (CD20+, IgD-),
                     provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                     (NCI) and Dr. Gerald Marti (CBER).  cDNA synthesis was
                     primed with a Not I - oligo(dT) primer
                     (5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCCTCATTTTTTTTTTTTTTTTT-
                     3'). Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo."
```



```
                     /db_xref="taxon:9606"
                     /clone="IMAGE:683118"
                     /clone_lib="NCI_CGAP_GCB1"
                     /tissue_type="germinal center B cell"
                     /lab_host="DH10B"
     mRNA            <1. .>335
                     /db_xref="GDB:5586438"
BASE COUNT      101 a      74 c      56 g     101 t       3 others
ORIGIN

  Query Match           19.5%;  Score 17;  DB 8;  Length 335;
  Best Local Similarity 100.0%;  Pred. No. 3.57e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    189 AGAAAGGAAAAATATAG 205
          |||||||||||||||||
Qy     27 AGAAAGGAAAAATATAG 43


RESULT   27
LOCUS       T92994         337 bp    mRNA            EST       22-MAR-1995
DEFINITION  ye23c12.r1 Homo sapiens cDNA clone 118582 5' similar to contains
            MER29 repetitive element ;.
ACCESSION   T92994
NID         g724907
KEYWORDS    EST.
SOURCE      human clone=118582 library=Stratagene lung (#937210)
            vector=pBluescript SK- host=SOLR cells (kanamycin resistant)
            primer=M13RP1 RsiteI=EcoRI RsiteII=XhoI Normal lung tissue from a 72
            year old male. Cloned unidirectionally. Primer: Oligo dT. Average
            insert size: 1.0 kb; Uni-ZAP XR Vector; 5' adaptor sequence:
            5'-GAATTCGGCACGAG-3'; 3' adaptor sequence:
            5'-CTCGAGTTTTTTTTTTTTTTTTT-3'.
  ORGANISM  Homo sapiens
            Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 337)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
            Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
```

```
            High quality sequence stops: 255
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES              Location/Qualifiers
     source           1. .337
                      /organism="Homo sapiens"
                      /clone="118582"
BASE COUNT      101 a      50 c      64 g     122 t
ORIGIN

  Query Match           19.5%;  Score 17;  DB 5;  Length 337;
  Best Local Similarity 100.0%;  Pred. No. 3.57e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    107 AGGAAAAATATAGTAAA 123
          |||||||||||||||||
Qy     31 AGGAAAAATATAGTAAA 47


RESULT   28
LOCUS       AA589763      341 bp    mRNA            EST       16-SEP-1997
DEFINITION  vl79a08.rl Stratagene mouse diaphragm (#937303) Mus musculus cDNA
            clone 986870 5', mRNA sequence.
ACCESSION   AA589763
NID         g2403143
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 341)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:559150
            Seq primer: -28ml3 rev1 ET from Amersham
```



```
            High quality sequence stop: 291.
FEATURES              Location/Qualifiers
     source           1. .341
                      /organism="Mus musculus"
                      /note="Organ: diaphragm; Vector: pBluescript SK-; Site_1:
                      EcoRI; Site_2: XhoI; Cloned unidirectionally from mRNA
                      prepared from diaphragm muscle.  Primer: Oligo dT. Average
                      insert size: 1.5 kb. Uni-ZAP XR Vector; -5' adaptor
                      sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor sequence: 5'
                      CTCGAGTTTTTTTTTTTTTTTTT 3'"
                      /db_xref="taxon:10090"
                      /clone="986870"
                      /clone_lib="Stratagene mouse diaphragm (#937303)"
                      /tissue_type="diaphragm"
                      /dev_stage="adult"
                      /lab_host="SOLR (kanamycin resistant)"
     mRNA             <1. .>341
BASE COUNT       78 a      87 c      74 g     102 t
ORIGIN

  Query Match           19.5%;  Score 17;  DB 9;  Length 341;
  Best Local Similarity 100.0%;  Pred. No. 3.57e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    252 TTATTAAGACAAATTAC 268
          |||||||||||||||||
Cp     68 TTATTAAGACAAATTAC 52


RESULT   29
LOCUS       AA211119      353 bp    mRNA            EST       12-MAR-1998
DEFINITION  zq87e10.s1 Stratagene hNT neuron (#937233) Homo sapiens cDNA clone
            649002 3' similar to contains Alu repetitive element;, mRNA
            sequence.
ACCESSION   AA211119
NID         g1809774
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 353)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
            Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
            Theising,B., White,T., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Wilson RK
            Washington University School of Medicine
```



```
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 3802  Std Error: 0.00
            Seq primer: -41ml3 fwd. ET from Amersham
            High quality sequence stop: 185.
FEATURES              Location/Qualifiers
     source           1. .353
                      /organism="Homo sapiens"
                      /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                      XhoI; Cloned unidirectionally.  Primer: Oligo dT.
                      Differentiated, post mitotic hNT neurons. Average insert
                      size: 1.5 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5'
                      GAATTCGGCACGAG 3' ~3' adaptor sequence: 5'
                      CTCGAGTTTTTTTTTTTTTTTTT 3'"
                      /db_xref="GDB:5278281"
                      /db_xref="taxon:9606"
                      /clone="649002"
                      /clone_lib="Stratagene hNT neuron (#937233)"
                      /dev_stage="hNT neurons"
                      /lab_host="SOLR (kanamycin resistant)"
BASE COUNT       86 a      91 c      80 g      95 t       1 others
ORIGIN

  Query Match           19.5%;  Score 17;  DB 13;  Length 353;
  Best Local Similarity 100.0%;  Pred. No. 3.57e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    290 ATTTTTCCTTTCTAAAA 306
          |||||||||||||||||
Cp     39 ATTTTTCCTTTCTAAAA 23


RESULT   30
LOCUS       AI048690      367 bp    mRNA            EST       08-JUL-1998
DEFINITION  ub30g02.r1 Soares 2NbMT Mus musculus cDNA clone 1379282 5', mRNA
            sequence.
ACCESSION   AI048690
NID         g3296977
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 367)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
```



```
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:901750
            Seq primer: -28ml3 rev2 ET from Amersham
            High quality sequence stop: 183.
FEATURES              Location/Qualifiers
     source           1. .367
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                      was primed with a Not I - oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCGTTTTTTTTTTTTTTTTTTTTTTT
                      3']; double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Not I and cloned into the Not I
                      and Eco RI sites of the modified pT7T3 vector. RNA
                      provided by Dr. Bertrand Jordan. Library went through two
                      rounds of normalization, and was constructed by Bento
                      Soares and M.Fatima Bonaldo."
                      /db_xref="taxon:10090"
                      /clone="1379282"
                      /clone_lib="Soares 2NbMT"
                      /sex="male"
                      /tissue_type="Thymus"
                      /dev_stage="4 weeks"
                      /lab_host="DH10B"
BASE COUNT      110 a      43 c      73 g     141 t
ORIGIN

  Query Match           19.5%;  Score 17;  DB 15;  Length 367;
  Best Local Similarity 100.0%;  Pred. No. 3.57e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    117 TATATTTTCCTTTCTA 133
          ||||||||||||||||
Cp     42 TATATTTTCCTTTCTA 26


RESULT   31
```

```
LOCUS       AA561954      373 bp    mRNA            EST       18-AUG-1997
DEFINITION  vl40b10.rl Stratagene mouse skin (#937313) Mus musculus cDNA clone
            974683 5', mRNA sequence.
ACCESSION   AA561954
NID         g2333419
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae;
            Murinae; Mus.
REFERENCE   1  (bases 1 to 373)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:555411
            Seq primer: -28ml3 rev1 ET from Amersham
            High quality sequence stop: 317.
FEATURES              Location/Qualifiers
     source           1. .373
                      /organism="Mus musculus"
                      /strain="C57BL/6"
                      /note="Organ: skin; Vector: pBluescript SK-; Site_1:
                      EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                      Oligo dT. Whole skin from 11 week old C57BL/6 female mice.
                      Average insert size: 1.0 kb; Uni-ZAP XR Vector; ~5'
                      adaptor sequence: 5' GAATTCGGCACGAG 3' -3' adaptor
                      sequence: 5' CTCGAGTTTTTTTTTTTTTTTTT 3'"
                      /db_xref="taxon:10090"
                      /clone="974683"
                      /clone_lib="Stratagene mouse skin (#937313)"
                      /sex="females"
                      /tissue_type="whole skin"
                      /dev_stage="11 weeks old"
                      /lab_host="SOLR (kanamycin resistant)"
     mRNA             <1. .>373
BASE COUNT      112 a     79 c     93 g     89 t
```



```
ORIGIN

  Query Match           19.5%;  Score 17;  DB 8;  Length 373;
  Best Local Similarity 100.0%;  Pred. No. 3.57e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    108 GTAATTTGTCTTAATAA 124
          |||||||||||||||||
Qy     52 GTAATTTGTCTTAATAA 68


RESULT   32
LOCUS       B54002        379 bp    DNA             GSS       19-JUN-1998
DEFINITION  CIT-HSP-2014D3.TR CIT-HSP Homo sapiens genomic clone 2014D3,
            genomic survey sequence.
ACCESSION   B54002
NID         g2608336
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 379)
  AUTHORS   Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K.,
            Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
            Simon,M. and Venter,J.C.
  TITLE     Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: CIT-HSP-2014D3.TF
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: M13 Reverse
            Class: BAC ends.
FEATURES              Location/Qualifiers
     source           1. .379
                      /organism="Homo sapiens"
                      /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                      HindIII"
                      /db_xref="taxon:9606"
                      /clone="2014D3"
                      /clone_lib="CIT-HSP"
                      /sex="Male"
```



```
                      /cell_type="Sperm"
BASE COUNT      108 a     58 c     66 g    147 t
ORIGIN

  Query Match           19.5%;  Score 17;  DB 26;  Length 379;
  Best Local Similarity 100.0%;  Pred. No. 3.57e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    314 ATTTTTCCTTTCTAAAA 330
          |||||||||||||||||
Cp     39 ATTTTTCCTTTCTAAAA 23


RESULT   33
LOCUS       W58969        383 bp    mRNA            EST       06-JUN-1996
DEFINITION  md62b11.rl Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
            clone 372957 5', mRNA sequence.
ACCESSION   W58969
NID         g1365733
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;
            Murinae; Mus.
REFERENCE   1  (bases 1 to 383)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:234389
            Seq primer: ETPrimer
            High quality sequence stop: 347.
FEATURES              Location/Qualifiers
     source           1. .383
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
```



```
                      polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                      was primed with a Not I - oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGGGAAATTTTTTTTTTTTTTTTTTTTT
                      T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
                      14.5dpc embryos [total RNA provided by  Minoru Ko, Wayne
                      State Univ., from 2 ]; double-stranded cDNA was ligated to
                      Eco RI adaptors (Pharmacia), digested with Not I and
                      cloned into the Not I and Eco RI sites of the modified
                      pT7T3 vector. Library went through one round of
                      normalization, and was constructed by Bento Soares and
                      M.Fatima Bonaldo."
                      /db_xref="taxon:10090"
                      /clone="372957"
                      /clone_lib="Soares mouse embryo NbME13.5 14.5"
                      /sex="unknown"
                      /tissue_type="embryo"
                      /dev_stage="13.5-14.5dpc total fetus"
                      /lab_host="DH10B"
     mRNA             <1. .>383
BASE COUNT      108 a    102 c     68 g    105 t
ORIGIN

  Query Match           19.5%;  Score 17;  DB 20;  Length 383;
  Best Local Similarity 100.0%;  Pred. No. 3.57e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    231 GTTTTAGAAAGGAAAAA 247
          |||||||||||||||||
Qy     22 GTTTTAGAAAGGAAAAA 38


RESULT   34
LOCUS       W02716        385 bp    mRNA            EST       11-OCT-1996
DEFINITION  zc65b04.s1 Soares fetal heart NbHH19W Homo sapiens cDNA clone
            327151 3', mRNA sequence.
ACCESSION   W02716
NID         g1274714
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 385)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
```

```
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 472  Std Error: 0.00
            Seq primer: mob.REGA+ET
            High quality sequence stop: 237.
FEATURES              Location/Qualifiers
     source           1. .385
                      /organism="Homo sapiens"
                      /note="Organ: heart; Vector: pT7T3D (Pharmacia) with a
                      modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                      strand cDNA was primed with a Not I - oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCAGTCTTTTTTTTTTTTTTTT 3'],
                      double-stranded cDNA was size selected, ligated to Eco RI
                      adapters (Pharmacia), digested with Not I and cloned into
                      the Not I and Eco RI sites of a modified pT7T3 vector
                      (Pharmacia). Library went through one round of
                      normalization to a Cot = 5. Library constructed by
                      M.Fatima Bonaldo. This library was constructed from the
                      same fetus as the fetal lung library, Soares fetal lung
                      NbHL19W."
                      /db_xref="taxon:9606"
                      /clone="327151"
                      /clone_lib="Soares fetal heart NbHH19W"
                      /sex="unknown"
                      /dev_stage="19 weeks"
                      /lab_host="DH10B (ampicillin resistant)"
     mRNA             complement(<1. .>385)
BASE COUNT      116 a      64 c      89 g     114 t       2 others
ORIGIN

  Query Match           19.5%;  Score 17;  DB 21;  Length 385;
  Best Local Similarity 100.0%;  Pred. No. 3.57e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    123 CTATATTTTCCTTTCT 139
          ||||||||||||||||
Cp     43 CTATATTTTCCTTTCT 27


RESULT   35
LOCUS       HUM070H01B    388 bp    mRNA            EST       21-MAY-1996
DEFINITION  Human fetal brain cDNA 5'-end GEN-070H01, mRNA sequence.
ACCESSION   D59850
NID         g960956
```



```
KEYWORDS    EST; EST(expressed sequence tag); Human fetal brain; similar to
            none(May 29,1995).
SOURCE      Homo sapiens cDNA to mRNA, clone_lib:Clontech human fetal brain
            polyA+ mRNA (#6535).
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 388)
  AUTHORS   Fujiwara,T., Hirano,H., Katagiri,T., Kawai,A., Kuga,Y., Nagata,M.,
            Okuno,S., Ozaki,K., Shimizu,F., Shimada,Y., Shinomiya,H.,
            Takaichi,A., Takeda,S., Watanabe,T., Takahashi,E., Hirai,Y.,
            Maekawa,H., Shin,S. and Nakamura,Y.
  TITLE     Unpublished(101)
  JOURNAL   Unpublished (1995)
REFERENCE   2  (bases 1 to 388)
  AUTHORS   Fujiwara,T
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAY-1995) to the DDBJ/EMBL/GenBank databases. Tsutomu
            Fujiwara, Otsuka GEN Research Institute,Otsuka Pharmaceutical
            Co.,Ltd; 463-10 Kagasuno Kawauchi-cho, Tokushima, Tokushima 771-01,
            Japan (Tel:0886-65-2888, Fax:0886-37-1035)
COMMENT     Submitted (30-May-1995) to DDBJ by:
            Tsutomu Fujiwara
            Otsuka GEN Research Institute
            Otsuka Pharmaceutical Co.,Ltd
            463-10 kagasuno Kawauchi-cho
            Tokushima, Tokushima
            771-01
            Japan
            Phone: 0886-65-2888
            Fax  : 0886-37-1035.
FEATURES              Location/Qualifiers
     source           1. .388
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone_lib="Clontech human fetal brain polyA+ mRNA
                      (#6535)"
BASE COUNT      130 a      74 c      53 g     131 t
ORIGIN

  Query Match           19.5%;  Score 17;  DB 7;  Length 388;
  Best Local Similarity 100.0%;  Pred. No. 3.57e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     89 AATTACACATTTTACTA 105
          |||||||||||||||||
Cp     57 AATTACACATTTTACTA 41


RESULT   36
```



```
LOCUS       AQ199311      397 bp    DNA             GSS       16-SEP-1998
DEFINITION  RPCI11-58B19.TJ RPCI11 Homo sapiens genomic clone R-58B19, genomic
            survey sequence.
ACCESSION   AQ199311
NID         g3611510
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 397)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
            Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
  TITLE     Use of human BAC End Sequences for Sequence-Ready Map Building
  JOURNAL   Unpublished (1998)
COMMENT     Other_GSSs: RPCI11-58B19.TK
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            For clone availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Class: BAC ends.
FEATURES              Location/Qualifiers
     source           1. .397
                      /organism="Homo sapiens"
                      /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                      RPCI11 Human Male BAC Library"
                      /db_xref="taxon:9606"
                      /clone="R-58B19"
                      /clone_lib="RPCI11"
                      /sex="Male"
                      /cell_type="Lymphocytes"
BASE COUNT      146 a      74 c      43 g     134 t
ORIGIN

  Query Match           19.5%;  Score 17;  DB 28;  Length 397;
  Best Local Similarity 100.0%;  Pred. No. 3.57e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     86 ACAAATTACACATTTTA 102
          |||||||||||||||||
Cp     60 ACAAATTACACATTTTA 44


RESULT   37
```



```
LOCUS       R32738        399 bp    mRNA            EST       28-APR-1995
DEFINITION  yh71g12.r1 Homo sapiens cDNA clone 135238 5'.
ACCESSION   R32738
NID         g788581
KEYWORDS    EST.
SOURCE      human clone=135238 library=Soares placenta Nb2HP vector=pT7T3D
            (Pharmacia) with a modified polylinker host=DH10B (ampicillin
            resistant) primer=M13RP1 Rsite1=Not I Rsite2=Eco RI Female placenta
            obtained at birth (full term). 1st strand cDNA was primed with a
            Not I - oligo(dT) primer [5'
            AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTTT 3'], double-stranded
            cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not
            I and cloned into the Not I and Eco RI sites of the modified pT7T3
            vector. Library went through one round of normalization. Library
            constructed by Bento Soares and M.Fatima Bonaldo.
  ORGANISM  Homo sapiens
            Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 399)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 318
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES              Location/Qualifiers
     source           1. .399
                      /organism="Homo sapiens"
                      /clone="135238"
BASE COUNT      123 a      70 c      81 g     119 t       6 others
ORIGIN

  Query Match           19.5%;  Score 17;  DB 5;  Length 399;
  Best Local Similarity 100.0%;  Pred. No. 3.57e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     96 GTTTTAGAAAGGAAAAA 112
          |||||||||||||||||
```

```
Qy      22 GTTTTAGAAAGGAAAAA 38

RESULT   38
LOCUS       AI145082      404 bp    mRNA            EST       28-SEP-1998
DEFINITION  UI-R-BT0-pt-h-01-0-UI.s1 UI-R-BT0 Rattus norvegicus cDNA clone
            UI-R-BT0-pt-h-01-0-UI 3', mRNA sequence.
ACCESSION   AI145082
NID         g3666881
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 404)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT
            Contact: Soares, MB
            Program for Rat Gene Diacovery and Mapping
            University of Iowa
            451 Eckestein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence tag present in the cDNA between the NotI site and the
            oligo-dT track served to identify it as a clone from the normalized
            adult NRBmedulla library. cDNA Library Preparation: M. Fatima
            Bonaldo, Ph.D. Clone distribution: clones will be available through
            Research Genetics The following repetitive elements were found in
            this cDNA sequence: 1-64, >AT_rich#Low_complexity 186-290,
            >B1_MM#SINE/Alu
            Seq primer: M13 Forward.
FEATURES             Location/Qualifiers
     source          1. .404
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; This library
                     (UI-R-BT0) consists of a mixture of individually tagged
                     normalized libraries constructed from rat hippocampus,
                     thalamus, mid-brain, medulla, corpus striatum, cerebral
                     cortex and testis. The tag used to  identify the source
                     tissue is a string of 3-6 nucleotides present  between the
                     Not I site and the oligo-dT track which allows
                     identification  of the library of origin of a clone within
                     the mixture. This library was then subtracted using a
                     driver consisting of a mixture of all clones from UI-R-A0,
```



```
                     UI-R-A1, UI-R-E0, UI-R-E1, UI-R-C0, UI-R-C1, UI-R-C2 and
                     UI-R-C2p."
                     /db_xref="taxon:10116"
                     /clone="UI-R-BT0-pt-h-01-0-UI"
                     /clone_lib="UI-R-BT0"
                     /dev_stage="adult"
                     /lab_host="DH10B (Life Technologies)"
BASE COUNT      116 a      76 c      54 g     158 t
ORIGIN

  Query Match           19.5%;  Score 17;  DB 18;  Length 404;
  Best Local Similarity 100.0%;  Pred. No. 3.57e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      25 ACATTTTACTATATTTT 41
           |||||||||||||||||
Cp      51 ACATTTTACTATATTTT 35


RESULT   39
LOCUS       AA043790      417 bp    mRNA            EST       10-MAY-1997
DEFINITION  zx59f12.s1 Soares pregnant uterus NbHPU Homo sapiens cDNA clone
            487151 3', mRNA sequence.
ACCESSION   AA043790
NID         g1521675
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 417)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
            Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 730   Std Error: 0.00
            Seq primer: -40M13 fwd. from Amersham
            High quality sequence stop: 332.
```



```
FEATURES             Location/Qualifiers
     source          1. .417
                     /organism="Homo sapiens"
                     /note="Organ: uterus; Vector: pT7T3-Pac; Site_1: Not I;
                     Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                     oligo(dT) primer [5'
                     AACTGGAAGAATTCGCGGCCGCCTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization.  Library
                     constructed by M. Fatima Bonaldo."
                     /db_xref="taxon:9606"
                     /clone="487151"
                     /clone_lib="Soares pregnant uterus NbHPU"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
     mRNA            complement(<1. .>417)
                     /db_xref="GDB:3760865"
BASE COUNT      122 a      76 c      95 g     124 t
ORIGIN

  Query Match           19.5%;  Score 17;  DB 6;  Length 417;
  Best Local Similarity 100.0%;  Pred. No. 3.57e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     122 CTATATTTTCCTTTCT 138
           ||||||||||||||||
Cp      43 CTATATTTTCCTTTCT 27


RESULT   40
LOCUS       H27973        423 bp    mRNA            EST       13-JUL-1995
DEFINITION  y162f12.r1 Homo sapiens cDNA clone 162863 5'.
ACCESSION   H27973
NID         g898326
KEYWORDS    EST.
SOURCE      human clone=162863 library=Soares breast 3NbHBst vector=pT7T3D
            (Pharmacia) with a modified polylinker host=DH10B (ampicillin
            resistant) primer=M13RP1 RsiteI=Not I RsiteII=Eco RI Adult human.
            1st strand cDNA was primed with a Not I - oligo(dT) primer [5'
            TGTTACCAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTT 3'],
            double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia),
            digested with Not I and cloned into the Not I and Eco RI sites of a
            modified pT7T3 vector (Pharmacia). Library went through one round
            of normalization to a Cot = 20. Library constructed by Bento Soares
            and M.Fatima Bonaldo.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
```



```
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 423)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 371
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES             Location/Qualifiers
     source          1. .423
                     /organism="Homo sapiens"
                     /clone="162863"
BASE COUNT      112 a      90 c      92 g     128 t      1 others
ORIGIN

  Query Match           19.5%;  Score 17;  DB 16;  Length 423;
  Best Local Similarity 100.0%;  Pred. No. 3.57e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     390 AGAAAGGAAAAATATAG 406
           |||||||||||||||||
Qy      27 AGAAAGGAAAAATATAG 43


RESULT   41
LOCUS       AA648291      434 bp    mRNA            EST       13-NOV-1997
DEFINITION  ns20e02.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1184186
            similar to contains element MSR1 repetitive element ;, mRNA
            sequence.
ACCESSION   AA648291
NID         g2574720
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 434)
```

```
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            unknown library type
            Insert Length: 1311    Std Error: 0.00
            Seq primer: -40ml3 fwd. ET from Amersham
            High quality sequence stop: 384.
FEATURES             Location/Qualifiers
     source          1..434
                     /organism="Homo sapiens"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from human tonsillar cells enriched for
                     germinal center B cells by flow sorting (CD20+, IgD-),
                     provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                     (NCI) and Dr. Gerald Marti (CBER).  cDNA synthesis was
                     primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTT-
                     3']. Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo."
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1184186"
                     /clone_lib="NCI_CGAP_GCB1"
                     /tissue_type="germinal center B cell"
                     /lab_host="DH10B"
BASE COUNT       90 a     87 c    135 g    122 t
ORIGIN

  Query Match           19.5%;  Score 17;  DB 10;  Length 434;
  Best Local Similarity 100.0%;  Pred. No. 3.57e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      24 AAAATATAGTAAAATGT 40
           |||||||||||||||||
Qy      35 AAAATATAGTAAAATGT 51


RESULT   42
LOCUS       AQ057587      443 bp    DNA              GSS       29-JUL-1998
DEFINITION  CIT-HSP-2345K18.TR CIT-HSP Homo sapiens genomic clone 2345K18,
            genomic survey sequence.
ACCESSION   AQ057587
NID         g3354113
```



```
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 443)
  AUTHORS   Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K.,
            Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
            Simon,M. and Venter,J.C.
  TITLE     Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: CIT-HSP-2345K18.TF
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13 Reverse
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..443
                     /organism="Homo sapiens"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII"
                     /db_xref="taxon:9606"
                     /clone="2345K18"
                     /clone_lib="CIT-HSP"
                     /sex="Male"
                     /cell_type="Sperm"
BASE COUNT      142 a     89 c     67 g    145 t
ORIGIN

  Query Match           19.5%;  Score 17;  DB 27;  Length 443;
  Best Local Similarity 100.0%;  Pred. No. 3.57e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     282 TTAGAAAGGAAAAATAT 298
           |||||||||||||||||
Qy      25 TTAGAAAGGAAAAATAT 41


RESULT   43
LOCUS       AQ052301      455 bp    DNA              GSS       29-JUL-1998
DEFINITION  RPCI11-51C19.TJ RPCI11 Homo sapiens genomic clone R-51C19, genomic
            survey sequence.
```



```
ACCESSION   AQ052301
NID         g3349338
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 455)
  AUTHORS   Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K.,
            Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences for Sequence-Ready Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: RPCI11-51C19.TK
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            For clone availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..455
                     /organism="Homo sapiens"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
                     /db_xref="taxon:9606"
                     /clone="R-51C19"
                     /clone_lib="RPCI11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
BASE COUNT      145 a     58 c     70 g    182 t
ORIGIN

  Query Match           19.5%;  Score 17;  DB 27;  Length 455;
  Best Local Similarity 100.0%;  Pred. No. 3.57e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     310 TAAAATGTGTAATTTGT 326
           |||||||||||||||||
Qy      44 TAAAATGTGTAATTTGT 60


RESULT   44
LOCUS       AI187151      465 bp    mRNA             EST       09-OCT-1998
DEFINITION  qe42h06.x1 Soares_fetal_lung_NbHL19W Homo sapiens cDNA clone
```



```
            IMAGE:1741691 3', mRNA sequence.
ACCESSION   AI187151
NID         g3737789
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 465)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40UP from Gibco
            High quality sequence stop: 403.
FEATURES             Location/Qualifiers
     source          1..465
                     /organism="Homo sapiens"
                     /note="Organ: lung; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCAATTTTTTTTTTTTTTTTT-3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5.  Library constructed by Bento
                     Soares and M.Fatima Bonaldo. This library was constructed
                     from the same fetus as the fetal heart library, Soares
                     fetal heart NbHH19W."
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1741691"
                     /clone_lib="Soares_fetal_lung_NbHL19W"
                     /dev_stage="19 weeks"
                     /lab_host="DH10B (ampicillin resistant)"
BASE COUNT      132 a     87 c    106 g    140 t
ORIGIN

  Query Match           19.5%;  Score 17;  DB 18;  Length 465;
  Best Local Similarity 100.0%;  Pred. No. 3.57e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     122 CTATATTTTTCCTTTCT 138
           |||||||||||||||||
Cp      43 CTATATTTTTCCTTTCT 27
```

```
RESULT   45
LOCUS       AA760212      466 bp    mRNA            EST       23-JAN-1998
DEFINITION  vv70e12.r1 Stratagene mouse skin (#937313) Mus musculus cDNA clone
            1227790 5' similar to WP:P35G2.2 CE05809 E.COLI YCAC LIKE ;, mRNA
            sequence.
ACCESSION   AA760212
NID         g2808006
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 466)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,H., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:655382
            Possible reversed clone: similarity on wrong strand
            Seq primer: -28ml3 rev1 ET from Amersham
            High quality sequence stop: 403.
FEATURES             Location/Qualifiers
     source          1. .466
                     /organism="Mus musculus"
                     /strain="C57BL/6"
                     /note="Organ: skin; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. Whole skin from 11 week old C57BL/6 female mice.
                     Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'
                     adaptor sequence: 5' GAATTCGGCACGAG 3' -3' adaptor
                     sequence: 5' CTCGAGTTTTTTTTTTTTTTTTT 3'"
                     /db_xref="taxon:10090"
                     /clone="1227790"
                     /clone_lib="Stratagene mouse skin (#937313)"
                     /sex="females"
                     /tissue_type="whole skin"
```



```
                     /dev_stage="11 weeks old"
                     /lab_host="SOLR (kanamycin resistant)"
BASE COUNT      130 a     106 c     123 g     107 t
ORIGIN

  Query Match           19.5%;  Score 17;  DB 12;  Length 466;
  Best Local Similarity 100.0%;  Pred. No. 3.57e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      70 GTAATTTGTCTTAATAA 86
           ||||||||||||||||
Qy      52 GTAATTTGTCTTAATAA 68


RESULT   46
LOCUS       N68485        466 bp    mRNA            EST       13-MAR-1996
DEFINITION  za14g04.s1 Homo sapiens cDNA clone 292566 3' similar to contains
            Alu repetitive element;contains element PTR7 repetitive element ;.
ACCESSION   N68485
NID         g1224646
KEYWORDS    EST.
SOURCE      human clone=292566 primer=ml3 -40 forward library=Soares fetal
            liver spleen 1NFLS vector=pT7T3D (Pharmacia) with a modified
            polylinker host=DH10B (ampicillin resistant) Rsite1=Pac I
            Rsite2=Eco RI Liver and spleen from a 20 week-post conception male
            fetus. 1st strand cDNA was primed with a Pac I - oligo(dT) primer
            [5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTT 3'],
            double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia),
            digested with Pac I and cloned into the Pac I and Eco RI sites of
            the modified pT7T3 vector.  Library went through one round of
            normalization. Library constructed by Bento Soares and M.Fatima
            Bonaldo.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 466)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
```



```
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 435
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES             Location/Qualifiers
     source          1. .466
                     /organism="Homo sapiens"
                     /clone="292566"
     mRNA            <1. .>466
BASE COUNT      161 a     74 c      76 g     152 t      3 others
ORIGIN

  Query Match           19.5%;  Score 17;  DB 19;  Length 466;
  Best Local Similarity 100.0%;  Pred. No. 3.57e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      72 TTTATTAAGACAAATTA 88
           |||||||||||||||||
Cp      69 TTTATTAAGACAAATTA 53


RESULT   47
LOCUS       AA836628      500 bp    mRNA            EST       07-APR-1998
DEFINITION  od66a12.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1372894,
            mRNA sequence.
ACCESSION   AA836628
NID         g2910947
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 500)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
            Ph.D., Gerald Marti, M.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
```



```
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 1162   Std Error: 0.00
            Seq primer: -40ml3 fwd. ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .500
                     /organism="Homo sapiens"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from human tonsillar cells enriched for
                     germinal center B cells by flow sorting (CD20+, IgD-),
                     provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                     (NCI) and Dr. Gerald Marti (CBER).  cDNA synthesis was
                     primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCTCATTTTTTTTTTTTTTTTT-
                     3']. Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo."
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1372894"
                     /clone_lib="NCI_CGAP_GCB1"
                     /tissue_type="germinal center B cell"
                     /lab_host="DH10B"
BASE COUNT      146 a     95 c     101 g     157 t      1 others
ORIGIN

  Query Match           19.5%;  Score 17;  DB 14;  Length 500;
  Best Local Similarity 100.0%;  Pred. No. 3.57e-03;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     450 CTATATTTTTCCTTTCT 466
           |||||||||||||||||
Cp      43 CTATATTTTTCCTTTCT 27


RESULT   48
LOCUS       AQ186050      504 bp    DNA             GSS       11-SEP-1998
DEFINITION  HS_3070_A2_H05_MF CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3070 Col=10 Row=O, genomic survey
            sequence.
ACCESSION   AQ186050
NID         g3585492
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 504)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Furlong,J. and
```

```
            Shaker,R.
  TITLE     Construction of a Characterized Clone Resource for Genomic
            Sequencing
  JOURNAL   Unpublished (1998)
COMMENT
            Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3070  row: O  column: 10
            Class: BAC ends
            High quality sequence stop: 504.
FEATURES              Location/Qualifiers
     source           1. .504
                      /organism="Homo sapiens"
                      /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                      E-Coli DH10B"
                      /db_xref="taxon:9606"
                      /clone="Plate=3070 Col=10 Row=O"
                      /clone_lib="CIT Approved Human Genomic Sperm Library D"
                      /sex="male"
BASE COUNT      178 a    101 c     75 g    146 t      4 others
ORIGIN

  Query Match           19.5%;  Score 17;  DB 28;  Length 504;
  Best Local Similarity  89.5%;  Pred. No. 3.57e-03;
  Matches    17;  Conservative    0;  Mismatches   2;  Indels   0;  Gaps   0;

Db      54 TTAATANAATCTATCTCTA 72
           |||||| ||||||||| ||
Qy      62 TTAATAAAATCTATCTNTA 80


RESULT   49
LOCUS       FR0014332     569 bp    DNA             GSS       18-SEP-1997
DEFINITION  F.rubripes GSS sequence, clone 143J05bG12, genomic survey sequence.
ACCESSION   AL005575
NID         g2451145
KEYWORDS    GSS; genome survey sequence.
SOURCE      Fugu rubripes.
  ORGANISM  Fugu rubripes
            Eukaryotae; Metazoa; Chordata; Vertebrata; Actinopterygii;
            Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Percomorpha;
            Tetraodontiformes; Tetraodontoidei; Tetraodontidae; Fugu.
REFERENCE   1  (bases 1 to 569)
  AUTHORS   Elgar,G., Clark,M., Smith,S., Meek,S., Warner,S., Umrania,Y.,
            Williams,G. and Brenner,S.
```



```
  TITLE     Direct Submission
  JOURNAL   Submitted (09-SEP-1997) MRC Human Genome Mapping Project Resource
            Centre Hinxton, Cambridge, CB10 1SB. Email: biohelp@hgmp.mrc.ac.uk
COMMENT     Vector: pBluescript II KS
            V_type: phagemid
            PRIMER: KS
            DESCR:
            One pass dye-terminator sequencing of cosmid cloned genomic
            sequence.
FEATURES              Location/Qualifiers
     source           1. .569
                      /organism="Fugu rubripes"
                      /db_xref="taxon:31033"
                      /clone_lib="cosmid 143J05"
                      /clone="143J05bG12"
BASE COUNT      180 a    115 c     95 g    152 t     27 others
ORIGIN

  Query Match           19.5%;  Score 17;  DB 26;  Length 569;
  Best Local Similarity 100.0%;  Pred. No. 3.57e-03;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     318 GACAAATTACACATTTT 334
           |||||||||||||||||
Cp      61 GACAAATTACACATTTT 45


RESULT   50
LOCUS       FR0014311     576 bp    DNA             GSS       18-SEP-1997
DEFINITION  F.rubripes GSS sequence, clone 143J05aG2, genomic survey sequence.
ACCESSION   AL005554
NID         g2451124
KEYWORDS    GSS; genome survey sequence.
SOURCE      Fugu rubripes.
  ORGANISM  Fugu rubripes
            Eukaryotae; Metazoa; Chordata; Vertebrata; Actinopterygii;
            Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Percomorpha;
            Tetraodontiformes; Tetraodontoidei; Tetraodontidae; Fugu.
REFERENCE   1  (bases 1 to 576)
  AUTHORS   Elgar,G., Clark,M., Smith,S., Meek,S., Warner,S., Umrania,Y.,
            Williams,G. and Brenner,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-SEP-1997) MRC Human Genome Mapping Project Resource
            Centre Hinxton, Cambridge, CB10 1SB. Email: biohelp@hgmp.mrc.ac.uk
COMMENT     Vector: pBluescript II KS
            V_type: phagemid
            PRIMER: KS
            DESCR:
            One pass dye-terminator sequencing of cosmid cloned genomic
            sequence.
FEATURES              Location/Qualifiers
```



```
     source           1. .576
                      /organism="Fugu rubripes"
                      /db_xref="taxon:31033"
                      /clone_lib="cosmid 143J05"
                      /clone="143J05aG2"
BASE COUNT      137 a    122 c    126 g    166 t     25 others
ORIGIN

  Query Match           19.5%;  Score 17;  DB 26;  Length 576;
  Best Local Similarity 100.0%;  Pred. No. 3.57e-03;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     102 AAAATGTGTAATTTGTC 118
           |||||||||||||||||
Qy      45 AAAATGTGTAATTTGTC 61


Search completed: Tue Dec 22 22:09:20 1998
Job time : 273 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Tue Dec 22 22:09:40 1998; MasPar time 27.52 Seconds
435.072 Million cell updates/sec

Tabular output not generated.

Title: >US-08-530-112A-3084
Description: (1-88) from US08530112A.seq
Perfect Score: 87
N.A. Sequence: 1 GATCAATCTTTGATGGTGAG.....AATCTATCTNTACATCTAAA 88
Comp: CTAGTTAGAAACTACCACTC.....TTAGATAGANATGTAGATTT

Scoring table: TABLE jmetric
Gap 60

Kmatch STD : Dbase 0; Query 0

Searched: 188442 seqs, 68026449 bases x 2

Post-processing: Minimum Match 0%
Listing first 1000 summaries

Database: n-geneseq32
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39 40:part40

Statistics: Mean 5.796; Variance 2.997; scale 1.934

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
	1	87	100.0	88	19	T22084	Human gene signature	3.08e-52
c	2	17	19.5	1667	8	Q48589	Alcohol oxidase gene	1.46e+00
c	3	16	18.4	69	6	Q41037	pFP-Zsp11 constructio	5.57e+00
	4	16	18.4	266	25	T30632	Probe nucleic acid #1	5.57e+00
	5	16	18.4	1095	40	V18817	Human papillomavirus-	5.57e+00
c	6	16	18.4	1376	5	Q32251	BDV structural glycop	5.57e+00
c	7	16	18.4	2210	5	Q32252	Encodes BDV structura	5.57e+00
c	8	16	18.4	2406	28	T59675	DNA encoding hRIP, a	5.57e+00
c	9	16	18.4	2487	5	Q32250	Encodes BDV structura	5.57e+00
c	10	16	18.4	2583	34	T79881	cDNA encoding HIV Rev	5.57e+00
c	11	16	18.4	2625	28	T61925	Full length human cad	5.57e+00
c	12	16	18.4	2625	33	T85403	Human cadherin-11 cod	5.57e+00
c	13	16	18.4	3224	5	Q32253	Encodes fusion of BDV	5.57e+00
c	14	16	18.4	3712	10	Q44392	Sequence of human OSF	5.57e+00
c	15	16	18.4	3914	10	Q44393	Sequence of human OSF	5.57e+00
	16	16	18.4	4057	23	T36899	Mouse neuron restrict	5.57e+00
	17	16	18.4	7902	24	T09847	Human papilloma virus	5.57e+00
	18	16	18.4	7904	35	T94723	Human papillomavirus	5.57e+00
	19	16	18.4	7904	35	T94724	Human papillomavirus	5.57e+00
c	20	16	18.4	12492	1	N70543	Complete sequence of	5.57e+00
	21	15	17.2	280	21	T11901	Mycobacterium ITG 872	2.05e+01
	22	15	17.2	325	20	T22640	Human gene signature	2.05e+01
	23	15	17.2	404	19	T20519	Human gene signature	2.05e+01
	24	15	17.2	2726	12	Q70916	Multidrug-resistance	2.05e+01
	25	15	17.2	2726	31	T43322	Multidrug resistance	2.05e+01
	26	15	17.2	2726	12	Q70907	Multidrug-resistance	2.05e+01
	27	15	17.2	4378	1	Q04522	Multidrug Resistance	2.05e+01
	28	15	17.2	4646	12	Q72872	Human multidrug resis	2.05e+01
	29	15	17.2	4669	2	N70752	Sequence of human mul	2.05e+01
	30	15	17.2	4669	9	Q52726	Sequence of human mul	2.05e+01
c	31	15	17.2	9048	25	T43225	Brassica napus PCA ge	2.05e+01
c	32	15	17.2	9763	25	T43224	Arabidopsis PCA gene.	2.05e+01
	33	15	17.2	14042	20	T28774	Human placental calci	2.05e+01
	34	15	17.2	14042	39	V05995	Human placenta calciu	2.05e+01
	35	15	17.2	14044	39	V05997	Human parathyroid cal	2.05e+01
	36	15	17.2	14044	20	T28776	Human parathyroid cal	2.05e+01
	37	15	17.2	14080	39	V05996	Human kidney calcium	2.05e+01
	38	15	17.2	14080	20	T28775	Human kidney calcium	2.05e+01
	39	15	17.2	14086	39	V05994	Human calcium protein	2.05e+01
	40	15	17.2	14086	20	T28773	Human calcium sensor	2.05e+01
c	41	14	16.1	30	17	T04088	Trypsin inhibitory pr	7.26e+01
	42	14	16.1	39	21	T35614	M.leprae pre-rRNA pro	7.26e+01

	43	14	16.1	39	39	V12977	Mycobacterium leprae	7.26e+01
	44	14	16.1	43	10	Q58024	Sequence of PCR prime	7.26e+01
	45	14	16.1	63	20	T23965	Human gene signature	7.26e+01
	46	14	16.1	73	10	Q58025	Sequence of PCR prime	7.26e+01
c	47	14	16.1	190	20	T22257	Human gene signature	7.26e+01
c	48	14	16.1	195	21	T26382	Human gene signature	7.26e+01
c	49	14	16.1	227	19	T21383	Human gene signature	7.26e+01
c	50	14	16.1	275	23	T26396	Human gene signature	7.26e+01
	51	14	16.1	280	21	T11902	Mycobacterium ITG 872	7.26e+01
	52	14	16.1	281	21	T11795	Mycobacterium strain	7.26e+01
	53	14	16.1	281	21	T11802	Mycobacterium strain	7.26e+01
	54	14	16.1	282	21	T11798	Mycobacterium strain	7.26e+01
	55	14	16.1	282	21	T11900	Mycobacterium strain	7.26e+01
	56	14	16.1	286	21	T11823	M. genavense ITG 8777	7.26e+01
	57	14	16.1	286	21	T11825	M. genavense ITG 9500	7.26e+01
	58	14	16.1	286	21	T11824	M. genavense ITG 92-7	7.26e+01
	59	14	16.1	288	21	T11826	M. simiae-like ITG 73	7.26e+01
	60	14	16.1	289	21	T11827	M. simiae-like ITG 97	7.26e+01
	61	14	16.1	299	21	T11838	M. haemophilum ITG 30	7.26e+01
	62	14	16.1	299	21	T11836	M. haemophilum ITG 77	7.26e+01
	63	14	16.1	299	21	T11837	M. haemophilum ITG 77	7.26e+01
	64	14	16.1	341	8	Q59482	Human brain Expressed	7.26e+01
	65	14	16.1	567	17	T04085	Human trypsin inhibit	7.26e+01
	66	14	16.1	594	17	T04077	Trypsin inhibitory pr	7.26e+01
	67	14	16.1	597	17	T04092	Human T98G cells tryp	7.26e+01
	68	14	16.1	642	38	T84942	Human prostate protei	7.26e+01
	69	14	16.1	816	8	Q43054	Maize HSP70 introns.	7.26e+01
c	70	14	16.1	983	5	Q29860	Odorant receptor clon	7.26e+01
	71	14	16.1	987	34	T90555	Clostridium perfringe	7.26e+01
	72	14	16.1	987	34	T90554	Clostridium perfringe	7.26e+01
	73	14	16.1	1060	19	T11196	Legionella pneumophil	7.26e+01
c	74	14	16.1	1241	3	Q20253	Wound-inducible potat	7.26e+01
	75	14	16.1	1537	20	T26895	Bacteriophage resista	7.26e+01
	76	14	16.1	1618	35	T89285	Tobacco microspore-sp	7.26e+01
	77	14	16.1	1747	2	Q13798	Zea mays 15 kD seed s	7.26e+01
	78	14	16.1	1776	10	Q58018	Sequence of altered c	7.26e+01
	79	14	16.1	1776	10	Q58017	Sequence of altered c	7.26e+01
	80	14	16.1	1867	7	Q40424	Bacillus protease 164	7.26e+01
	81	14	16.1	1901	1	Q06356	Sequence encoding sal	7.26e+01
	82	14	16.1	2031	40	V24737	H. pylori ORF hp3e103	7.26e+01
	83	14	16.1	2048	20	T07273	Hsp150 gene (Heat Sho	7.26e+01
	84	14	16.1	2048	8	Q48872	hsp150 gene.	7.26e+01
	85	14	16.1	2110	7	Q41228	Clone GP22 encoding m	7.26e+01
c	86	14	16.1	2270	1	N90704	Rhoptry membrane anti	7.26e+01
	87	14	16.1	2409	2	Q12084	C-terminal sequence e	7.26e+01
	88	14	16.1	2600	28	T43624	Chromatin regulator p	7.26e+01
	89	14	16.1	2880	20	T05158	res2+ gene.	7.26e+01
c	90	14	16.1	2927	12	Q69008	2-5A-dependent RNA-as	7.26e+01
c	91	14	16.1	2928	16	T03900	Human 2-5A-dependent	7.26e+01
c	92	14	16.1	2928	29	T59648	Human 2-5A-dependent	7.26e+01
c	93	14	16.1	2928	16	T03901	Human 2-5A-dependent	7.26e+01

c	94	14	16.1	2969	16	Q96405	Yeast NLT1 gene.	7.26e+01
	95	14	16.1	3182	10	Q55940	Human leukaemia inhib	7.26e+01
	96	14	16.1	3182	9	Q58425	Human LIF-R clone 65.	7.26e+01
	97	14	16.1	3182	7	Q42590	Human LIF-R N-termina	7.26e+01
	98	14	16.1	3182	4	Q25799	hLIF-R.	7.26e+01
	99	14	16.1	3182	16	Q92270	Human leukaemia inhib	7.26e+01
	100	14	16.1	3266	17	T12476	Tumour suppressor HTS	7.26e+01
c	101	14	16.1	3467	19	T12324	DNA encoding Sulfolob	7.26e+01
	102	14	16.1	3591	10	Q55942	Human leukaemia inhib	7.26e+01
	103	14	16.1	3591	9	Q58427	Human LIF-R.	7.26e+01
	104	14	16.1	3591	16	Q92272	Human leukaemia inhib	7.26e+01
c	105	14	16.1	3695	23	T36391	Lactococcus lactis su	7.26e+01
	106	14	16.1	3720	13	Q80513	Genetic locus bcl-6.	7.26e+01
c	107	14	16.1	3980	10	Q62696	Sequence of human p-3	7.26e+01
	108	14	16.1	4291	18	T04702	ALVAC C3 flanking arm	7.26e+01
	109	14	16.1	4464	30	T67059	AUX1 gene involved in	7.26e+01
	110	14	16.1	4857	30	T69962	FIPV nucleocapsid (N)	7.26e+01
	111	14	16.1	5176	39	V04017	Human multiple myelom	7.26e+01
c	112	14	16.1	5278	37	V04446	Rat programmed cell d	7.26e+01
	113	14	16.1	5883	20	T12234	BmNPVCP genomic seque	7.26e+01
	114	14	16.1	7091	33	T47530	H6 promoted cytomegal	7.26e+01
	115	14	16.1	7291	12	Q67860	ALVAC subgenomic frag	7.26e+01
	116	14	16.1	7350	30	T69982	Canarypox virus DNA c	7.26e+01
	117	14	16.1	7351	3	Q22770	Canarypox virus BglII	7.26e+01
	118	14	16.1	7351	33	T47529	Canarypox virus DNA c	7.26e+01
c	119	14	16.1	7587	40	V15703	Zucchini ACC synthase	7.26e+01
c	120	14	16.1	7587	3	Q15133	Zucchini ACC synthase	7.26e+01
	121	14	16.1	7753	4	Q26727	Encodes exons XV to X	7.26e+01
c	122	14	16.1	10614	15	Q89555	Hamster cholesterol 7	7.26e+01
	123	14	16.1	134525	1	Q04525	Total base sequence o	7.26e+01
	124	14	16.1	580073	27	T58840	Mycoplasma genitalium	7.26e+01
	125	13	14.9	34	21	T35606	M.intracellulare pre-	2.45e+02
	126	13	14.9	34	39	V12970	Mycobacterium intrace	2.45e+02
	127	13	14.9	34	39	V12957	Mycobacterium avium p	2.45e+02
	128	13	14.9	34	21	T35590	M.avium pre-rRNA prob	2.45e+02
	129	13	14.9	35	2	Q12995	Cytotoxic lymphocyte	2.45e+02
	130	13	14.9	50	17	Q98891	Anti-human IL-4 MAb h	2.45e+02
c	131	13	14.9	58	17	Q98890	Anti-human IL-4 MAb h	2.45e+02
	132	13	14.9	72	6	Q35430	PCR primer H6P24.	2.45e+02
c	133	13	14.9	117	20	T21526	Human gene signature	2.45e+02
c	134	13	14.9	134	22	T25352	Human gene signature	2.45e+02
c	135	13	14.9	177	26	T40839	Serine protease nfSP1	2.45e+02
c	136	13	14.9	218	19	T20145	Human gene signature	2.45e+02
c	137	13	14.9	235	18	T21009	Human gene signature	2.45e+02
	138	13	14.9	269	19	T20837	Human gene signature	2.45e+02
c	139	13	14.9	269	15	Q90459	ChIFN cDNA probe pCh2	2.45e+02
	140	13	14.9	278	21	T11792	M. avium ATCC 151.769	2.45e+02
	141	13	14.9	278	21	T11793	M. paratuberculosis s	2.45e+02
	142	13	14.9	280	21	T11794	Mycobacterium strain	2.45e+02
	143	13	14.9	280	21	T11812	Mycobacterium strain	2.45e+02
	144	13	14.9	280	21	T11809	Mycobacterium strain	2.45e+02

	145	13	14.9	280	21	T11806	Mycobacterium strain 2.45e+02
	146	13	14.9	280	21	T11800	Mycobacterium strain 2.45e+02
	147	13	14.9	280	21	T11814	Mycobacterium strain 2.45e+02
	148	13	14.9	280	21	T11808	Mycobacterium strain 2.45e+02
	149	13	14.9	280	21	T11796	Mycobacterium strain 2.45e+02
	150	13	14.9	280	21	T11799	Mycobacterium strain 2.45e+02
	151	13	14.9	280	21	T11813	Mycobacterium strain 2.45e+02
	152	13	14.9	281	21	T11811	Mycobacterium strain 2.45e+02
	153	13	14.9	281	21	T11810	Mycobacterium strain 2.45e+02
	154	13	14.9	281	21	T11803	Mycobacterium strain 2.45e+02
	155	13	14.9	281	21	T11815	M. scrofulaceum ITG 4 2.45e+02
	156	13	14.9	282	21	T11801	Mycobacterium strain 2.45e+02
	157	13	14.9	282	21	T11797	Mycobacterium strain 2.45e+02
	158	13	14.9	282	21	T11805	M. lufu ITG 4755 16S- 2.45e+02
	159	13	14.9	293	20	T22183	Human gene signature 2.45e+02
	160	13	14.9	315	8	Q60405	Human brain Expressed 2.45e+02
c	161	13	14.9	318	40	V19751	Antibody LG2-5-VL cha 2.45e+02
c	162	13	14.9	318	29	T61238	Human anti-RSV monocl 2.45e+02
c	163	13	14.9	321	8	Q60524	Human brain Expressed 2.45e+02
c	164	13	14.9	328	8	Q61058	Human brain Expressed 2.45e+02
c	165	13	14.9	336	8	Q59913	Human brain Expressed 2.45e+02
c	166	13	14.9	338	2	Q10975	Bovine bone specific 2.45e+02
	167	13	14.9	343	35	T69305	Murine metastatic nuc 2.45e+02
c	168	13	14.9	383	8	Q59194	Human brain Expressed 2.45e+02
c	169	13	14.9	383	6	Q39782	Expressed Sequence Ta 2.45e+02
c	170	13	14.9	384	37	V01888	Human OTK27 gene. 2.45e+02
c	171	13	14.9	390	14	Q87237	Anti-interleukin-1-al 2.45e+02
c	172	13	14.9	393	17	Q98824	Anti-human IL-4 human 2.45e+02
	173	13	14.9	396	8	Q59612	Human brain Expressed 2.45e+02
	174	13	14.9	403	8	Q60838	Human brain Expressed 2.45e+02
	175	13	14.9	404	8	Q59426	Human brain Expressed 2.45e+02
c	176	13	14.9	452	9	Q53217	Sequence encoding gas 2.45e+02
	177	13	14.9	474	31	T43493	ATM gene exon 61. 2.45e+02
	178	13	14.9	484	18	T13967	Eimeria gametocyte an 2.45e+02
c	179	13	14.9	510	2	N70546	Sequence encoding N-t 2.45e+02
	180	13	14.9	528	18	T19025	Human gene signature 2.45e+02
c	181	13	14.9	549	11	Q66815	AmEPV thymidine-kinas 2.45e+02
c	182	13	14.9	708	29	T61242	Human anti-RSV monocl 2.45e+02
c	183	13	14.9	723	15	Q94597	ShET1 enterotoxin loc 2.45e+02
c	184	13	14.9	767	15	Q90458	Chicken interferon ge 2.45e+02
c	185	13	14.9	795	19	T16027	PS201T6 toxin coding 2.45e+02
c	186	13	14.9	795	31	T47253	Delta-endotoxin codin 2.45e+02
c	187	13	14.9	795	13	Q85261	Delta-endotoxin 201T6 2.45e+02
c	188	13	14.9	837	13	Q74396	Isoform lgl1 of the g 2.45e+02
c	189	13	14.9	840	37	T84977	Mouse GDNFR alpha clo 2.45e+02
	190	13	14.9	841	26	T28529	S. pneumoniae detecti 2.45e+02
c	191	13	14.9	857	1	N93084	Lambda gt 11 pl(13T) 2.45e+02
c	192	13	14.9	867	13	Q74397	Isoform lgl2 of the g 2.45e+02
	193	13	14.9	885	30	T67698	H. pylori transporter 2.45e+02
c	194	13	14.9	918	31	T51788	Natural killer cell e 2.45e+02
	195	13	14.9	927	1	N92874	Expression plasmid pU 2.45e+02

	196	13	14.9	957	30	T68046	H. pylori transporter 2.45e+02
	197	13	14.9	961	12	Q67433	DNA encoding an antig 2.45e+02
	198	13	14.9	1020	1	N92873	Expression plasmid pU 2.45e+02
c	199	13	14.9	1030	32	T79905	Actin gene promoter. 2.45e+02
c	200	13	14.9	1045	1	Q04456	Bacillus thuringiensi 2.45e+02
c	201	13	14.9	1131	30	T68204	H. pylori secreted or 2.45e+02
	202	13	14.9	1138	24	T41465	gClq receptor cDNA. 2.45e+02
	203	13	14.9	1203	15	Q87971	Factor-XA cDNA. 2.45e+02
c	204	13	14.9	1215	2	N60836	5'region of bacterial 2.45e+02
	205	13	14.9	1303	18	T17863	Napin gene. 2.45e+02
	206	13	14.9	1313	27	T48101	Immunogenic type F bo 2.45e+02
	207	13	14.9	1314	15	Q87988	Porphyromonas gingiva 2.45e+02
	208	13	14.9	1365	36	T93683	Thermococcus AEDII12R 2.45e+02
	209	13	14.9	1390	2	Q13751	Peptidylglycine alpha 2.45e+02
	210	13	14.9	1429	17	T05625	Neurotrophin-6 cDNA. 2.45e+02
c	211	13	14.9	1437	14	Q81959	Prostaglandin FP rece 2.45e+02
	212	13	14.9	1458	14	Q79935	Murine Kin17 cDNA. 2.45e+02
c	213	13	14.9	1470	3	Q15322	Sucrose phosphorylase 2.45e+02
	214	13	14.9	1487	17	Q98953	gill insert encoding 2.45e+02
c	215	13	14.9	1493	37	V01889	Human OTK27 gene. 2.45e+02
c	216	13	14.9	1511	39	V14508	AmEPV entomopoxvirus 2.45e+02
c	217	13	14.9	1511	5	Q28302	AmEPV tk DNA. 2.45e+02
c	218	13	14.9	1511	11	Q66798	AmEPV thymidine-kinas 2.45e+02
	219	13	14.9	1512	9	Q55709	Human A1 adenosine re 2.45e+02
	220	13	14.9	1518	1	N92872	Expression plasmid pU 2.45e+02
	221	13	14.9	1563	36	T99964	Coding sequence for R 2.45e+02
c	222	13	14.9	1648	4	Q25532	Sequence of genomic c 2.45e+02
	223	13	14.9	1665	1	N92871	Expression plasmid pU 2.45e+02
c	224	13	14.9	1772	23	T38521	Chamaecyparis obtusa 2.45e+02
c	225	13	14.9	1825	2	N80259	Sequence encoding hum 2.45e+02
	226	13	14.9	1829	2	N60863	Sequence encoding p-n 2.45e+02
	227	13	14.9	1871	34	T73869	Cotton lipid transfer 2.45e+02
c	228	13	14.9	1875	29	T41614	Adzuki bean endo-xylo 2.45e+02
	229	13	14.9	2004	13	Q74680	Bacillus cereus 80 kD 2.45e+02
	230	13	14.9	2004	39	V16167	DNA encoding an 80 kD 2.45e+02
	231	13	14.9	2004	38	T73995	B. cereus 80 kD VIP1A 2.45e+02
	232	13	14.9	2004	20	T13941	B. cereus VIP1 protei 2.45e+02
	233	13	14.9	2079	1	N94527	Expression plasmid pU 2.45e+02
	234	13	14.9	2138	38	T95297	Rat glial cell line-d 2.45e+02
	235	13	14.9	2244	36	T99965	Gene for RCH1-related 2.45e+02
	236	13	14.9	2273	17	T05626	Neurotrophin-6 gene. 2.45e+02
c	237	13	14.9	2326	13	Q79399	Human NMDAR2 receptor 2.45e+02
	238	13	14.9	2378	37	T84975	Rat glial cell derive 2.45e+02
c	239	13	14.9	2448	1	Q03453	cDNA molecule which h 2.45e+02
c	240	13	14.9	2448	9	Q52648	Sequence encoding pro 2.45e+02
c	241	13	14.9	2485	4	Q25594	Rabbit GnT I cDNA clo 2.45e+02
	242	13	14.9	2497	1	Q03444	HindIII S fragment of 2.45e+02
	243	13	14.9	2520	20	T09941	Anther earl gene prom 2.45e+02
c	244	13	14.9	2582	2	N60610	Vector sequence deriv 2.45e+02
c	245	13	14.9	2585	2	N70629	DNA -a sequence of Be 2.45e+02
	246	13	14.9	2610	1	N93060	cDNA of plasmid pXA45 2.45e+02

	247	13	14.9	2625	15	Q87970	Peptidyl C-terminal a 2.45e+02
	248	13	14.9	2655	38	T73994	B. cereus VIP1A(a) co 2.45e+02
	249	13	14.9	2655	39	V16166	DNA encoding a 100 kD 2.45e+02
	250	13	14.9	2655	20	T13940	B. cereus VIP1A(a) ge 2.45e+02
	251	13	14.9	2655	13	Q74679	Bacillus cereus 100 k 2.45e+02
	252	13	14.9	2809	31	T73280	S. cerevisiae origin 2.45e+02
	253	13	14.9	2809	15	Q91287	ORC2 gene encoding a 2.45e+02
c	254	13	14.9	2901	3	Q14810	Btm PG14 72kDa Cry in 2.45e+02
c	255	13	14.9	3000	12	Q62136	Rice starch branching 2.45e+02
c	256	13	14.9	3000	10	Q54675	5' UTR and exons 1 an 2.45e+02
c	257	13	14.9	3025	13	Q79402	Human NMDAR2 receptor 2.45e+02
c	258	13	14.9	3075	37	V06252	Murine puromycin-sens 2.45e+02
c	259	13	14.9	3076	1	N80063	Equine gamma interfer 2.45e+02
c	260	13	14.9	3153	37	V06251	Human puromycin-sensi 2.45e+02
c	261	13	14.9	3215	11	Q70739	X-prolyl-dipeptidyl-a 2.45e+02
c	262	13	14.9	3243	13	Q79401	Human NMDAR2 receptor 2.45e+02
	263	13	14.9	3275	3	N97129	Partial sequence of m 2.45e+02
c	264	13	14.9	3281	8	Q47743	Truncated IBV spike p 2.45e+02
	265	13	14.9	3315	1	N90791	cDNA of plasmid pXA79 2.45e+02
c	266	13	14.9	3350	31	T72718	2-Oxoglutarate:accept 2.45e+02
	267	13	14.9	3383	3	Q20269	Gene encoding AE-III 2.45e+02
c	268	13	14.9	3489	25	T18211	Major neutralising an 2.45e+02
c	269	13	14.9	3545	2	N60680	Sequence encoding spi 2.45e+02
c	270	13	14.9	3600	2	N70196	Sequence encoding a r 2.45e+02
	271	13	14.9	3616	39	V00245	Rat Ret ligand retLl 2.45e+02
c	272	13	14.9	3645	2	N60679	Sequence encoding S1 2.45e+02
c	273	13	14.9	3698	13	Q79400	Human NMDAR2 receptor 2.45e+02
c	274	13	14.9	3717	10	Q56917	Glutamic acid recepto 2.45e+02
c	275	13	14.9	3731	5	Q32631	Porcine PAM-8. 2.45e+02
c	276	13	14.9	3747	37	V06250	Human puromycin-sensi 2.45e+02
	277	13	14.9	3786	13	Q73778	STE6-CFTR chimera H6. 2.45e+02
	278	13	14.9	3801	1	N81446	Sequence of Brassica 2.45e+02
	279	13	14.9	3840	12	Q73776	STE6-CFTR chimera H4. 2.45e+02
	280	13	14.9	3840	13	Q73777	STE6-CFTR chimera H5. 2.45e+02
	281	13	14.9	3962	32	T47566	H6 promoted cytomegal 2.45e+02
c	282	13	14.9	3969	11	Q65666	N-methyl-D-aspartic a 2.45e+02
	283	13	14.9	4041	38	T73999	B. cereus VIP1A(a)/VI 2.45e+02
	284	13	14.9	4041	39	V16172	DNA sequence encoding 2.45e+02
	285	13	14.9	4041	20	T13944	VIP2A(a) and VIP1A(a) 2.45e+02
	286	13	14.9	4138	2	N60683	Sequence of premelibi 2.45e+02
c	287	13	14.9	4274	13	Q79407	Human NMDAR2 receptor 2.45e+02
c	288	13	14.9	4289	13	Q79405	Human NMDAR2 receptor 2.45e+02
	289	13	14.9	4320	11	Q65428	Lambda CGN1-2 napin g 2.45e+02
	290	13	14.9	4325	15	Q94586	Lambda CGN1-2 napin r 2.45e+02
	291	13	14.9	4325	29	T65475	Gamma-ACGN1-2 napin g 2.45e+02
	292	13	14.9	4325	40	V15266	Brassica campestris n 2.45e+02
	293	13	14.9	4325	3	Q13969	Napin 1-2 sequence. 2.45e+02
c	294	13	14.9	4325	13	Q79404	Human NMDAR2 receptor 2.45e+02
c	295	13	14.9	4340	13	Q79372	Human N-methyl-D-aspa 2.45e+02
c	296	13	14.9	4349	13	Q79406	Human NMDAR2 receptor 2.45e+02
c	297	13	14.9	4364	10	Q63730	Protein tyrosine kina 2.45e+02

c	298	13	14.9	4364	13	Q79403	Human NMDAR2 receptor 2.45e+02
c	299	13	14.9	4368	8	Q49902	Glutamic acid recepto 2.45e+02
c	300	13	14.9	4375	38	T76892	Arabidopsis floral me 2.45e+02
	301	13	14.9	4529	8	Q46121	Porcine sodium ion/gl 2.45e+02
c	302	13	14.9	4529	8	Q46121	Porcine sodium ion/gl 2.45e+02
c	303	13	14.9	4640	15	Q91999	Mouse tie-2 receptor 2.45e+02
c	304	13	14.9	4663	1	N80085	BamHI fragment of equ 2.45e+02
c	305	13	14.9	4695	13	Q79378	Human N-methyl-D-aspa 2.45e+02
c	306	13	14.9	4738	13	Q50997	Rat NMDA receptor sub 2.45e+02
	307	13	14.9	4752	10	Q58032	Tyrosinase promoter. 2.45e+02
c	308	13	14.9	4818	33	T62177	Candida boidinni ATCC 2.45e+02
c	309	13	14.9	4818	32	T79909	Actin gene. 2.45e+02
	310	13	14.9	5100	6	Q37760	Glucosyltransferease 2.45e+02
	311	13	14.9	5137	38	V12500	Mouse metastasis-asso 2.45e+02
	312	13	14.9	5302	33	T47567	Cytomegalovirus gL ge 2.45e+02
	313	13	14.9	5798	32	T47565	Canarypox virus DNA c 2.45e+02
c	314	13	14.9	5820	32	T71205	HaSNPV polyhedrin gen 2.45e+02
	315	13	14.9	6049	20	T13939	B. cereus VIP2A(a) ge 2.45e+02
	316	13	14.9	6049	38	T73993	B. cereus VIP2A(a) co 2.45e+02
	317	13	14.9	6049	39	V16165	DNA encoding vegetati 2.45e+02
	318	13	14.9	6106	13	Q74678	Bacillus cereus vegat 2.45e+02
	319	13	14.9	8540	24	T41900	Vector pAPEX-3p. 2.45e+02
	320	13	14.9	8540	17	T08476	Eukaryotic expression 2.45e+02
	321	13	14.9	8575	7	Q40324	Sequence of pHEBO30. 2.45e+02
c	322	13	14.9	8760	12	Q79473	Porcine pro-interleuk 2.45e+02
c	323	13	14.9	8806	17	T09349	Rat IP3 receptor gene 2.45e+02
c	324	13	14.9	9326	4	Q25051	DNA complementary to 2.45e+02
c	325	13	14.9	9370	13	Q75165	AF-4 transcript (cDNA 2.45e+02
c	326	13	14.9	9391	13	Q75164	AF-4 transcript (cDNA 2.45e+02
c	327	13	14.9	10136	22	T34578	Kinetochore protein C 2.45e+02
c	328	13	14.9	10190	28	T33183	Fragment of human OCI 2.45e+02
	329	13	14.9	10580	6	Q38104	Vector pHEBO-DR-CAT. 2.45e+02
	330	13	14.9	10596	23	T40348	Plasmid pCisEBON for 2.45e+02
	331	13	14.9	10596	9	Q51731	Plasmid pCisEBON for 2.45e+02
c	332	13	14.9	10684	24	T33758	Control region isolat 2.45e+02
	333	13	14.9	10798	4	Q28505	Tomato vacuolar inver 2.45e+02
	334	13	14.9	10965	4	Q28507	Tomato vacuolar inver 2.45e+02
	335	13	14.9	11616	12	Q67390	Construct pHEBO-DR-RE 2.45e+02
	336	13	14.9	11616	6	Q38105	Vector pHEBO-DR-LUC. 2.45e+02
c	337	13	14.9	12151	10	Q54676	Rice starch branching 2.45e+02
c	338	13	14.9	12151	11	Q62137	Rice starch branching 2.45e+02
c	339	13	14.9	14176	34	T84564	Swinepox virus HindII 2.45e+02
	340	13	14.9	14237	39	V05384	Human cathepsin K gen 2.45e+02
c	341	13	14.9	19307	20	T27558	Shuttle vector pAdDel 2.45e+02
	342	13	14.9	20303	31	T71699	Human deoxycytidylate 2.45e+02
c	343	13	14.9	21126	2	N60970	TL-DNA region of A.rh 2.45e+02
	344	13	14.9	26764	31	T71696	Human deoxycytidylate 2.45e+02
c	345	13	14.9	134525	1	Q04525	Total base sequence o 2.45e+02
c	346	13	14.9	580073	27	T58840	Mycoplasma genitalium 2.45e+02
	347	12	13.8	15	29	T56329	Mouse TNF-a hammerhea 7.87e+02
	348	12	13.8	18	35	T68589	Plasmid pSM700 upstre 7.87e+02

	349	12	13.8	21	40	V17510	Probe MT80 for human 7.87e+02
	350	12	13.8	21	24	T44655	Human papillomavirus 7.87e+02
	351	12	13.8	21	17	Q98624	Human papilloma virus 7.87e+02
	352	12	13.8	21	8	Q48109	Toxigenic E.coli STp 7.87e+02
c	353	12	13.8	28	24	T30521	Primer 5'H23 for HNK- 7.87e+02
c	354	12	13.8	30	3	Q22426	PCR primer oligo-85. 7.87e+02
c	355	12	13.8	30	36	T70529	Human immunoglobulin 7.87e+02
c	356	12	13.8	31	16	Q93718	Human stromelysin ham 7.87e+02
c	357	12	13.8	31	16	Q93719	Human stromelysin ham 7.87e+02
	358	12	13.8	34	35	T68587	Plasmid pSM700 constr 7.87e+02
c	359	12	13.8	40	32	T87122	IFN-gamma 2'F RNA lig 7.87e+02
c	360	12	13.8	41	18	Q91105	Oligomer containing 1 7.87e+02
c	361	12	13.8	45	30	T43960	HHP-29 polymorphic se 7.87e+02
	362	12	13.8	57	20	T23186	Human gene signature 7.87e+02
	363	12	13.8	62	19	T20954	Human gene signature 7.87e+02
c	364	12	13.8	62	22	T25926	Human gene signature 7.87e+02
c	365	12	13.8	68	10	Q76448	Human genome fragment 7.87e+02
	366	12	13.8	69	36	T97764	Nucleic acid inhibito 7.87e+02
c	367	12	13.8	69	6	Q41032	pFP-ZsP11 constructio 7.87e+02
	368	12	13.8	73	6	Q41038	pFP-Zsp11 constructio 7.87e+02
	369	12	13.8	87	6	Q41022	P11 promoter mutant P 7.87e+02
	370	12	13.8	90	6	Q41023	P11 promoter mutant s 7.87e+02
c	371	12	13.8	91	19	T22078	Human gene signature 7.87e+02
c	372	12	13.8	94	3	Q14469	SCA gene linker Lab14 7.87e+02
	373	12	13.8	95	6	Q39037	Oligonucleotide ident 7.87e+02
c	374	12	13.8	101	21	T24667	Human gene signature 7.87e+02
	375	12	13.8	104	11	Q67968	Sequence of vaccinia 7.87e+02
	376	12	13.8	104	2	N60057	Transcriptional regul 7.87e+02
	377	12	13.8	105	2	N60056	Transcriptional regul 7.87e+02
c	378	12	13.8	107	19	T19780	Human gene signature 7.87e+02
	379	12	13.8	108	2	N60059	Transcriptional regul 7.87e+02
c	380	12	13.8	111	21	T24489	Human gene signature 7.87e+02
c	381	12	13.8	117	23	T17667	Functional biotin apt 7.87e+02
	382	12	13.8	120	31	T48114	Ubiquitin encoding fr 7.87e+02
c	383	12	13.8	121	20	T23451	Human gene signature 7.87e+02
	384	12	13.8	126	21	T22556	Human gene signature 7.87e+02
	385	12	13.8	129	36	T91724	Mouse SmLIM/CRP2 gene 7.87e+02
	386	12	13.8	135	18	T19727	Human gene signature 7.87e+02
c	387	12	13.8	140	18	T19339	Human gene signature 7.87e+02
c	388	12	13.8	145	22	T18572	Human chondrocyte RNA 7.87e+02
	389	12	13.8	147	19	T20892	Human gene signature 7.87e+02
c	390	12	13.8	156	18	T19322	Human gene signature 7.87e+02
	391	12	13.8	167	5	Q30994	Notch clone hN5k sequ 7.87e+02
	392	12	13.8	170	33	T73985	Rat THY-1 synthetic g 7.87e+02
	393	12	13.8	170	23	T13570	Oligo 3 for construct 7.87e+02
	394	12	13.8	173	22	T25083	Human gene signature 7.87e+02
	395	12	13.8	182	10	Q76563	Human genome fragment 7.87e+02
	396	12	13.8	184	23	T36032	rchd036 gene differen 7.87e+02
c	397	12	13.8	191	10	Q77009	Human genome fragment 7.87e+02
c	398	12	13.8	198	18	T19651	Human gene signature 7.87e+02
	399	12	13.8	219	20	T24106	Human gene signature 7.87e+02

c	400	12	13.8	222	20	T24127	Human gene signature 7.87e+02
c	401	12	13.8	228	18	T05711	Human ubiquitin gene. 7.87e+02
	402	12	13.8	229	20	T23952	Human gene signature 7.87e+02
	403	12	13.8	229	6	Q38520	3' end of Ri-alpha tr 7.87e+02
c	404	12	13.8	231	19	T20125	Human gene signature 7.87e+02
c	405	12	13.8	239	3	N40125	Sequence encoding por 7.87e+02
	406	12	13.8	243	31	T47122	Alphal region of clas 7.87e+02
c	407	12	13.8	243	4	Q23865	Heat-stable toxin (ST 7.87e+02
c	408	12	13.8	249	34	T89608	Hepatitis C virus H 3 7.87e+02
	409	12	13.8	250	21	T24635	Human gene signature 7.87e+02
	410	12	13.8	258	30	T67573	H. pylori secreted or 7.87e+02
c	411	12	13.8	258	4	Q24952	HIV-1 vpu gene amplif 7.87e+02
	412	12	13.8	261	30	T68194	H. pylori secreted or 7.87e+02
	413	12	13.8	270	11	Q77337	Human genome fragment 7.87e+02
c	414	12	13.8	280	40	T99590	Secreted protein AS15 7.87e+02
c	415	12	13.8	285	10	Q44342	Sequence of fragment 7.87e+02
c	416	12	13.8	285	34	T89595	Hepatitis C virus H c 7.87e+02
	417	12	13.8	286	19	T20372	Human gene signature 7.87e+02
	418	12	13.8	287	8	Q60961	Human brain Expressed 7.87e+02
c	419	12	13.8	294	22	T25305	Human gene signature 7.87e+02
	420	12	13.8	297	29	T44909	Partial sequence of C 7.87e+02
	421	12	13.8	304	4	Q27776	Human type IV collage 7.87e+02
c	422	12	13.8	306	10	Q44343	Sequence of fragment 7.87e+02
c	423	12	13.8	311	28	T59831	Beta-casein gene frag 7.87e+02
c	424	12	13.8	312	10	Q44344	Sequence of fragment 7.87e+02
c	425	12	13.8	314	21	T26541	Human gene signature 7.87e+02
c	426	12	13.8	319	20	T22908	Human gene signature 7.87e+02
	427	12	13.8	320	18	T13043	Cotton fibre-specific 7.87e+02
	428	12	13.8	320	31	T70050	Cotton fibre specific 7.87e+02
	429	12	13.8	320	24	T30260	Cotton fibre cell-spe 7.87e+02
c	430	12	13.8	321	34	T89598	Hepatitis C virus H c 7.87e+02
c	431	12	13.8	324	1	N81555	EPO 125 encoding eryt 7.87e+02
	432	12	13.8	328	20	T22636	Human gene signature 7.87e+02
c	433	12	13.8	330	11	Q68744	CHA255 light chain va 7.87e+02
c	434	12	13.8	330	1	Q04657	Sequence encoding met 7.87e+02
c	435	12	13.8	330	1	Q04594	Sequence encoding reg 7.87e+02
c	436	12	13.8	333	19	T20634	Human gene signature 7.87e+02
c	437	12	13.8	340	36	T84817	Staphylococcus aureus 7.87e+02
	438	12	13.8	343	22	T24789	Human gene signature 7.87e+02
c	439	12	13.8	347	20	T22626	Human gene signature 7.87e+02
c	440	12	13.8	360	40	V09795	DNA encoding the ligh 7.87e+02
c	441	12	13.8	371	40	V04475	Secreted protein AS15 7.87e+02
c	442	12	13.8	405	8	Q59596	Human brain Expressed 7.87e+02
	443	12	13.8	408	6	Q37084	T cell receptor alpha 7.87e+02
c	444	12	13.8	420	25	T43436	MAb Br-3 light chain 7.87e+02
c	445	12	13.8	420	5	Q08604	Br-3 Light Chain V Re 7.87e+02
c	446	12	13.8	423	30	T67644	H. pylori surface mem 7.87e+02
c	447	12	13.8	435	3	Q15163	V186 region of lambda 7.87e+02
c	448	12	13.8	436	20	T23265	Human gene signature 7.87e+02
c	449	12	13.8	441	29	T59189	Nucleic acid coding f 7.87e+02
	450	12	13.8	442	26	T28513	S. aureus detection p 7.87e+02

	451	12	13.8	449	1	N90391	Recombinant DNA probe 7.87e+02
c	452	12	13.8	450	13	Q80820	Gene X cistron. 7.87e+02
	453	12	13.8	452	8	Q50346	Mutant elastase gene 7.87e+02
	454	12	13.8	461	32	T78154	DNA sequence of L1 am 7.87e+02
	455	12	13.8	461	24	T44718	Human papillomavirus 7.87e+02
	456	12	13.8	469	33	T80579	Human plasma platelet 7.87e+02
	457	12	13.8	469	15	Q87964	Human acetyl hydrolas 7.87e+02
	458	12	13.8	469	37	T96146	Human plasma platelet 7.87e+02
	459	12	13.8	486	33	T73953	Rat THY-1 synthetic g 7.87e+02
c	460	12	13.8	494	2	N70776	Sequence encoding rho 7.87e+02
c	461	12	13.8	513	19	T20793	Human gene signature 7.87e+02
	462	12	13.8	527	18	T19026	Human gene signature 7.87e+02
c	463	12	13.8	537	24	T30559	3G3 variable lambda c 7.87e+02
	464	12	13.8	552	30	T67888	H. pylori transmembra 7.87e+02
c	465	12	13.8	556	34	T84560	MDR Mycobacterium HTP 7.87e+02
	466	12	13.8	558	15	Q92645	Recombinant mite alle 7.87e+02
c	467	12	13.8	564	16	Q89755	Plasmid pAR6 partial 7.87e+02
c	468	12	13.8	564	15	Q87557	Human T cell lymphotr 7.87e+02
c	469	12	13.8	567	8	Q49869	Haemopoietic stem cel 7.87e+02
c	470	12	13.8	569	18	T19479	Human gene signature 7.87e+02
c	471	12	13.8	601	5	Q29874	Pheromone receptor cl 7.87e+02
	472	12	13.8	617	19	T19853	Human gene signature 7.87e+02
	473	12	13.8	621	31	T47123	cDNA encoding soluble 7.87e+02
	474	12	13.8	626	35	T75627	Canine herpesvirus ge 7.87e+02
c	475	12	13.8	637	4	N10028	Sequence which encode 7.87e+02
c	476	12	13.8	638	24	T39910	Maize acetyl CoA carb 7.87e+02
c	477	12	13.8	642	33	T85093	Mouse monoclonal anti 7.87e+02
c	478	12	13.8	654	29	T61280	Humanised anti-L-sele 7.87e+02
c	479	12	13.8	661	14	Q79933	Arabidopsis thaliana 7.87e+02
c	480	12	13.8	662	4	N10026	Sequence of part of t 7.87e+02
c	481	12	13.8	675	35	T96973	Plasmid puD-BL5 human 7.87e+02
c	482	12	13.8	675	29	T65490	G protein conjugative 7.87e+02
	483	12	13.8	685	4	Q25063	Human HLA SB-beta cDN 7.87e+02
	484	12	13.8	687	1	Q04000	Sequence complementar 7.87e+02
c	485	12	13.8	687	1	Q04000	Sequence complementar 7.87e+02
	486	12	13.8	690	15	Q92015	DNA encoding an MHC p 7.87e+02
c	487	12	13.8	700	13	Q78989	Human immunoglobulin 7.87e+02
c	488	12	13.8	710	2	Q10689	Insecticidal toxin hy 7.87e+02
c	489	12	13.8	714	12	Q67358	Anti-traseolide 02/01 7.87e+02
c	490	12	13.8	717	36	T89950	Aequorea victoria gfp 7.87e+02
c	491	12	13.8	717	36	T89949	Aequorea victoria gfp 7.87e+02
c	492	12	13.8	717	36	T75682	Aequorea victoria gfp 7.87e+02
c	493	12	13.8	717	36	T75699	Aequorea victoria hum 7.87e+02
c	494	12	13.8	720	39	V14336	Wild type green fluor 7.87e+02
c	495	12	13.8	731	3	Q21036	Region specific to My 7.87e+02
	496	12	13.8	745	12	Q71372	DNA probe (Y1) used f 7.87e+02
	497	12	13.8	747	38	T97175	DR alpha-DAP chimeric 7.87e+02
c	498	12	13.8	752	4	N30196	Sequence encoding por 7.87e+02
c	499	12	13.8	752	4	N30186	Sequence encoding por 7.87e+02
	500	12	13.8	756	3	Q15507	FL-specific scTCR, is 7.87e+02
	501	12	13.8	756	3	Q15129	FL-specific scTCR, is 7.87e+02

	502	12	13.8	768	30	T68108	H. pylori cytoplasmic 7.87e+02
c	503	12	13.8	770	3	Q23859	ScFvB18 construct mut 7.87e+02
c	504	12	13.8	770	3	Q23861	ScFvB18 construct mut 7.87e+02
c	505	12	13.8	770	3	Q23863	ScFvB18 construct mut 7.87e+02
c	506	12	13.8	770	3	Q23860	ScFvB18 construct mut 7.87e+02
	507	12	13.8	788	1	Q01787	Partial sequence of g 7.87e+02
c	508	12	13.8	809	4	Q26809	DNA encoding caGM-CSF 7.87e+02
c	509	12	13.8	858	11	Q64817	pRAS111 between HindI. 7.87e+02
	510	12	13.8	897	15	Q92649	Recombinant mite alle 7.87e+02
c	511	12	13.8	900	5	Q29373	Mouse beta-RAR promot 7.87e+02
c	512	12	13.8	900	29	T07264	Mutant LAG1 coding se 7.87e+02
c	513	12	13.8	915	14	Q83764	Plasmid 16965. 7.87e+02
c	514	12	13.8	915	31	T72213	DNA encoding leaderle 7.87e+02
c	515	12	13.8	915	31	T72211	DNA encoding mature h 7.87e+02
c	516	12	13.8	915	14	Q83756	Plasmid pDM25486. 7.87e+02
c	517	12	13.8	918	35	T94894	Human G protein conju 7.87e+02
c	518	12	13.8	944	15	Q93253	Modified jellyfish gr 7.87e+02
c	519	12	13.8	947	2	Q14001	Porcine IL-1alpha gen 7.87e+02
c	520	12	13.8	951	38	V06397	Human melanocortin-1 7.87e+02
c	521	12	13.8	951	31	T68792	Melanocortin-1 recept 7.87e+02
	522	12	13.8	972	30	T68932	Bovine PIT-1 exon 6 s 7.87e+02
	523	12	13.8	981	37	T76682	Methanococcus jannasc 7.87e+02
	524	12	13.8	1004	3	Q21414	Encodes osmotin-like 7.87e+02
	525	12	13.8	1032	30	T77452	H. pylori cytoplasmic 7.87e+02
c	526	12	13.8	1032	31	T72200	DNA encoding leaderle 7.87e+02
c	527	12	13.8	1032	14	Q83747	Plasmid 41620. 7.87e+02
c	528	12	13.8	1032	31	T72194	DNA encoding leaderle 7.87e+02
	529	12	13.8	1039	2	N60687	Sequence encoding tri 7.87e+02
c	530	12	13.8	1039	12	Q66958	Herpesvirus saimiri H 7.87e+02
c	531	12	13.8	1039	17	T03337	Ly-6 terminal CIP HSV 7.87e+02
c	532	12	13.8	1039	15	Q99112	HVS-15 full-length cD 7.87e+02
	533	12	13.8	1041	2	N60407	Schizosaccharyomyces 7.87e+02
	534	12	13.8	1055	18	T15275	E.coli K12 dnaJ codin 7.87e+02
c	535	12	13.8	1063	39	V12466	Human oestrogen sulph 7.87e+02
c	536	12	13.8	1065	17	V16020	Ubiquitin-TFP1 gene f 7.87e+02
c	537	12	13.8	1095	13	Q72574	P. pastoris alcohol o 7.87e+02
c	538	12	13.8	1095	33	T80523	Partial alcohol oxida 7.87e+02
	539	12	13.8	1099	4	N30113	Sequence complementar 7.87e+02
	540	12	13.8	1099	4	N30068	Sequence of cDNA prob 7.87e+02
	541	12	13.8	1099	2	N60977	Sequence encoding hum 7.87e+02
c	542	12	13.8	1104	14	Q83754	Plasmid 25498. 7.87e+02
c	543	12	13.8	1104	31	T72201	DNA encoding leaderle 7.87e+02
c	544	12	13.8	1113	30	T68234	H. pylori cytoplasmic 7.87e+02
c	545	12	13.8	1130	23	T09250	Human somatostatin re 7.87e+02
c	546	12	13.8	1157	17	T01343	Rat T:alpha:1 alpha-t 7.87e+02
	547	12	13.8	1191	30	T66686	Soybean peroxidase SE 7.87e+02
	548	12	13.8	1200	4	Q25096	Vaccinia virus B15R g 7.87e+02
	549	12	13.8	1200	8	Q46243	B15R gene. 7.87e+02
c	550	12	13.8	1209	40	V16670	Polynucleotide sequen 7.87e+02
c	551	12	13.8	1209	27	T45882	Human colon specific 7.87e+02
	552	12	13.8	1215	30	T68041	H. pylori cytoplasmic 7.87e+02

	553	12	13.8	1227	23	T37306	Murine gonadotropin-r	7.87e+02
c	554	12	13.8	1240	4	Q24177	Tox2a gene.	7.87e+02
c	555	12	13.8	1254	1	N92937	cDNA from Plasmodium	7.87e+02
c	556	12	13.8	1256	6	Q38083	Encodes kangaroo eye	7.87e+02
c	557	12	13.8	1256	7	Q41217	Kangaroo mu-crystalli	7.87e+02
c	558	12	13.8	1257	12	Q70661	ScPV pRAS110 and pRAS	7.87e+02
c	559	12	13.8	1260	30	T72045	Sequence around marke	7.87e+02
c	560	12	13.8	1260	30	T43925	Sequence around marke	7.87e+02
c	561	12	13.8	1270	10	Q44340	Sequence encoding a p	7.87e+02
c	562	12	13.8	1278	10	Q55273	Jun v I gene.	7.87e+02
c	563	12	13.8	1284	36	T77139	Single chain antigen	7.87e+02
	564	12	13.8	1293	8	Q46244	B18R gene.	7.87e+02
c	565	12	13.8	1294	40	V15231	Class II acyl-ACP thi	7.87e+02
c	566	12	13.8	1294	15	Q92302	Leek class II thioest	7.87e+02
c	567	12	13.8	1347	12	Q67347	VCAM 2D-IgG.	7.87e+02
c	568	12	13.8	1351	7	Q45655	Human somatostatin re	7.87e+02
c	569	12	13.8	1355	15	Q87284	Aspartic proteinase g	7.87e+02
c	570	12	13.8	1356	12	Q70659	ScFV pRAS108 and pRAS	7.87e+02
c	571	12	13.8	1357	30	T66555	Rabies virus nucleopr	7.87e+02
	572	12	13.8	1359	1	Q02419	Promoter of pCP-B1 ex	7.87e+02
c	573	12	13.8	1431	3	Q22534	DNA sequence of plasm	7.87e+02
	574	12	13.8	1443	37	T88777	Banana ACC oxidase ge	7.87e+02
	575	12	13.8	1455	21	T30577	Leuconostoc dextranic	7.87e+02
	576	12	13.8	1461	21	T30576	Leuconostoc lactis gl	7.87e+02
	577	12	13.8	1470	35	T76589	Para-nitrobenzyl este	7.87e+02
	578	12	13.8	1484	37	T88776	Banana ACC oxidase ge	7.87e+02
	579	12	13.8	1494	33	T80580	Mouse plasma platelet	7.87e+02
	580	12	13.8	1494	29	T66137	Murine PAF-AH coding	7.87e+02
	581	12	13.8	1494	37	T96128	Mouse plasma platelet	7.87e+02
	582	12	13.8	1494	34	T87064	Mouse platelet-activa	7.87e+02
c	583	12	13.8	1500	34	T78902	Carboxy-terminal regi	7.87e+02
c	584	12	13.8	1524	3	N50163	Sequence encoding the	7.87e+02
c	585	12	13.8	1530	15	Q86694	Candida tropicalis pK	7.87e+02
c	586	12	13.8	1534	34	T90335	YC1 gene.	7.87e+02
c	587	12	13.8	1549	29	T59978	3' untranslated regio	7.87e+02
c	588	12	13.8	1585	33	T62365	Erythropoietin genomi	7.87e+02
	589	12	13.8	1586	4	Q27987	Protease from S. Aure	7.87e+02
c	590	12	13.8	1586	33	T62366	Erythropoietin genomi	7.87e+02
c	591	12	13.8	1587	33	T62368	Erythropoietin genomi	7.87e+02
c	592	12	13.8	1633	9	Q51227	Human MSH-R gene.	7.87e+02
	593	12	13.8	1644	2	N60445	Sequence encoding tum	7.87e+02
	594	12	13.8	1644	17	T15423	Mouse tumour necrosis	7.87e+02
	595	12	13.8	1649	7	Q41064	Clone derived when ch	7.87e+02
c	596	12	13.8	1675	7	Q43208	Sequence encoding soy	7.87e+02
	597	12	13.8	1680	33	T86265	Human brain Neuroglyc	7.87e+02
	598	12	13.8	1696	3	Q20959	L.dextranicus glucose	7.87e+02
c	599	12	13.8	1743	36	T85939	Shigella flexneri Ipa	7.87e+02
c	600	12	13.8	1748	15	Q88327	Arabidopsis thaliana	7.87e+02
c	601	12	13.8	1758	4	Q31765	cDNA clone pf7 encode	7.87e+02
c	602	12	13.8	1776	40	V15232	Class II acyl-ACP thi	7.87e+02
c	603	12	13.8	1776	15	Q92300	Leek class II thioest	7.87e+02

c	604	12	13.8	1776	14	Q86521	Heliothis virescens r	7.87e+02
	605	12	13.8	1780	3	Q15138	Clone ptACC5 encoding	7.87e+02
	606	12	13.8	1781	3	Q15137	Clone ptACC4 encoding	7.87e+02
	607	12	13.8	1783	3	Q15136	Clone ptACC3 encoding	7.87e+02
c	608	12	13.8	1788	40	V13120	Bacillus thuringiensi	7.87e+02
	609	12	13.8	1789	37	V01865	Staphylococcus aureus	7.87e+02
c	610	12	13.8	1789	23	T34657	NIa protease of the P	7.87e+02
	611	12	13.8	1793	3	Q15139	Clone ptACC6 encoding	7.87e+02
c	612	12	13.8	1797	23	T34661	NIa protease of the U	7.87e+02
	613	12	13.8	1800	39	V09713	ACC synthase DNA.	7.87e+02
	614	12	13.8	1809	37	T79353	Staphylococcus aureus	7.87e+02
	615	12	13.8	1818	3	Q15135	Clone ptACC2 encoding	7.87e+02
c	616	12	13.8	1835	2	Q10957	Luciferase gene.	7.87e+02
	617	12	13.8	1838	37	T79352	Staphylococcus aureus	7.87e+02
c	618	12	13.8	1841	21	T30651	P. gingivalis haemagg	7.87e+02
c	619	12	13.8	1864	2	N71405	Sequence of ANS-1 whi	7.87e+02
	620	12	13.8	1872	6	Q39220	Carrot reverse transc	7.87e+02
c	621	12	13.8	1876	34	T87067	Chicken platelet-acti	7.87e+02
	622	12	13.8	1885	22	T34524	Receptor for hyaluron	7.87e+02
	623	12	13.8	1902	31	T69651	Tumour suppressor gen	7.87e+02
	624	12	13.8	1906	1	Q04531	Human prolidase gene.	7.87e+02
c	625	12	13.8	1918	24	T45650	Mouse ST-S1 cDNA, use	7.87e+02
c	626	12	13.8	1920	2	N60108	Sequence encoding the	7.87e+02
	627	12	13.8	1921	27	T45123	S. typhimurium mviA+	7.87e+02
c	628	12	13.8	1923	13	Q77672	Sequence for controll	7.87e+02
c	629	12	13.8	1929	7	Q43590	VCAM-6D/ICAM-1 DNA.	7.87e+02
c	630	12	13.8	1932	7	Q46661	VCAM-6D/ICAM-2 DNA.	7.87e+02
	631	12	13.8	1934	25	T42302	Maize ribosomal inhib	7.87e+02
c	632	12	13.8	1941	7	Q46662	VCAM-6D/VCAM4-1 DNA.	7.87e+02
c	633	12	13.8	1941	7	Q43586	VCAM-6D DNA.	7.87e+02
c	634	12	13.8	1953	29	T68316	cDNA encoding human h	7.87e+02
c	635	12	13.8	1953	22	T28769	Human homologue of ba	7.87e+02
	636	12	13.8	1961	22	T34523	Receptor for hyaluron	7.87e+02
c	637	12	13.8	1974	13	Q77673	Sequence for controll	7.87e+02
	638	12	13.8	1975	6	Q36031	Chicken nov coding se	7.87e+02
	639	12	13.8	1977	2	Q11986	Transcription activat	7.87e+02
c	640	12	13.8	1982	1	N90225	Malaria-specific Plas	7.87e+02
c	641	12	13.8	2002	18	T12320	Anti-procathepsin B m	7.87e+02
	642	12	13.8	2061	31	T69652	Tumour suppressor gen	7.87e+02
	643	12	13.8	2064	3	Q14178	Gene encoding ascorba	7.87e+02
c	644	12	13.8	2091	3	Q23321	Soluble VCAM variant	7.87e+02
	645	12	13.8	2104	7	Q43722	Sequence of DNA clone	7.87e+02
	646	12	13.8	2105	13	Q75120	ELM1 locus encoding p	7.87e+02
c	647	12	13.8	2106	30	T67494	H. pylori outer membr	7.87e+02
	648	12	13.8	2116	6	Q35987	Tomato hsp80 cDNA seq	7.87e+02
	649	12	13.8	2123	37	V02801	Mouse RHAMM cDNA.	7.87e+02
c	650	12	13.8	2132	38	V06392	B. thuringiensis 14 a	7.87e+02
c	651	12	13.8	2134	29	T07263	LAG1 coding sequence.	7.87e+02
c	652	12	13.8	2150	35	T91037	Yeast checkpoint cont	7.87e+02
	653	12	13.8	2160	33	T85328	Truncated murine TRAF	7.87e+02
	654	12	13.8	2166	36	T93682	Thermotoga maritima M	7.87e+02

c	655	12	13.8	2168	36	T91114	Absidia griseola var.	7.87e+02
c	656	12	13.8	2181	7	Q43207	Sequence encoding mic	7.87e+02
	657	12	13.8	2186	1	Q02292	Major portion of Equi	7.87e+02
	658	12	13.8	2192	13	Q74634	POU domain transcript	7.87e+02
	659	12	13.8	2196	13	Q80100	ADPGPP small subunit	7.87e+02
	660	12	13.8	2205	13	Q74635	POU domain transcript	7.87e+02
c	661	12	13.8	2205	7	Q43587	VCAM/ICAM-1 DNA.	7.87e+02
c	662	12	13.8	2208	7	Q43588	VCAM/ICAM-2 DNA.	7.87e+02
	663	12	13.8	2212	13	Q79676	2.3 kb cbh1 promoter	7.87e+02
c	664	12	13.8	2220	39	V16209	cDNA encoding human v	7.87e+02
c	665	12	13.8	2220	5	Q29043	1E7-2G7 antigen cDNA.	7.87e+02
c	666	12	13.8	2220	10	Q44394	Sequence of bases 24-	7.87e+02
c	667	12	13.8	2220	3	Q21005	VCAM variant with 7 I	7.87e+02
c	668	12	13.8	2226	13	Q72485	Tumour rejection anti	7.87e+02
	669	12	13.8	2230	40	V15704	Tomato ACC synthase c	7.87e+02
c	670	12	13.8	2230	38	V06389	B. thuringiensis 14 a	7.87e+02
	671	12	13.8	2232	2	Q11125	Prolactin promoter an	7.87e+02
	672	12	13.8	2240	8	Q47668	Sequence of the rat b	7.87e+02
	673	12	13.8	2243	17	T00395	Stearoyl-ACP-desatura	7.87e+02
	674	12	13.8	2243	3	Q20187	Soybean seed stearoyl	7.87e+02
	675	12	13.8	2247	37	T95248	M. catarrhalis 4223 t	7.87e+02
	676	12	13.8	2247	37	T80398	Staphylococcus aureus	7.87e+02
c	677	12	13.8	2247	32	T61187	Human phospholipase A	7.87e+02
c	678	12	13.8	2247	21	T29822	Human phospholipase A	7.87e+02
c	679	12	13.8	2247	11	Q64162	PLA2 phospholipase ge	7.87e+02
c	680	12	13.8	2247	12	Q74416	Human phospholipase A	7.87e+02
c	681	12	13.8	2247	5	Q29420	DNA encoding cytosoli	7.87e+02
	682	12	13.8	2261	10	Q58016	Sequence of altered c	7.87e+02
	683	12	13.8	2287	37	T95251	M. catarrhalis R1 tra	7.87e+02
	684	12	13.8	2296	39	V12457	Human HP4 prostagland	7.87e+02
c	685	12	13.8	2305	5	Q32361	MAGE-51 genomic DNA.	7.87e+02
c	686	12	13.8	2305	13	Q72486	Tumour rejection anti	7.87e+02
	687	12	13.8	2321	11	Q64562	M. matalensis CCK B r	7.87e+02
c	688	12	13.8	2323	36	T86836	DNA encoding S. marce	7.87e+02
c	689	12	13.8	2323	25	T49479	Serratia marcescens c	7.87e+02
c	690	12	13.8	2323	10	Q57966	Bacterial chitinase-A	7.87e+02
c	691	12	13.8	2327	17	T03736	Phytase gene.	7.87e+02
c	692	12	13.8	2337	1	N91483	NP cDNA to genomic ju	7.87e+02
c	693	12	13.8	2391	30	T68246	H. pylori outer membr	7.87e+02
c	694	12	13.8	2400	1	Q06622	Asparagine synthetase	7.87e+02
	695	12	13.8	2418	33	T58629	Human brain-specific	7.87e+02
	696	12	13.8	2419	16	T05086	MZ2-MEL antigen E pre	7.87e+02
	697	12	13.8	2419	13	Q72476	Tumour rejection anti	7.87e+02
	698	12	13.8	2420	14	Q85435	Human melanoma antige	7.87e+02
	699	12	13.8	2420	13	Q72472	Tumour rejection anti	7.87e+02
c	700	12	13.8	2426	1	N80751	Sense strand of ApaI	7.87e+02
c	701	12	13.8	2435	16	Q94155	Human Fas ligand gene	7.87e+02
	702	12	13.8	2438	1	N90373	DNA encoding amphireg	7.87e+02
	703	12	13.8	2448	30	T63320	Yeast mitochondria CO	7.87e+02
	704	12	13.8	2450	27	T43378	Human cytokine respon	7.87e+02
	705	12	13.8	2465	2	Q13136	Hyp operon.	7.87e+02

	706	12	13.8	2466	11	Q66397	Protein kinase cDNA,	7.87e+02
	707	12	13.8	2470	28	T62855	Saccharomyces cerevis	7.87e+02
c	708	12	13.8	2481	37	T87940	Rat neuropeptide Y re	7.87e+02
c	709	12	13.8	2483	24	T41947	ACC oxidase gene prom	7.87e+02
c	710	12	13.8	2509	15	Q92776	Human thymopoietin ge	7.87e+02
	711	12	13.8	2533	1	Q06000	Sequence encoding pol	7.87e+02
c	712	12	13.8	2538	15	Q92232	Brevibacterium flavum	7.87e+02
	713	12	13.8	2547	17	T15543	Borrelia garinii 66 k	7.87e+02
	714	12	13.8	2574	14	Q76019	DNA encoding gp120 fr	7.87e+02
c	715	12	13.8	2604	37	T73595	Rat neuropeptide Y re	7.87e+02
	716	12	13.8	2622	35	T90461	Thermoanaerobacter th	7.87e+02
	717	12	13.8	2634	15	Q87841	Human neuronal calciu	7.87e+02
	718	12	13.8	2652	23	T34655	Testis-associated hel	7.87e+02
c	719	12	13.8	2658	14	Q81477	RAR-alpha-403 dominan	7.87e+02
c	720	12	13.8	2687	14	Q84716	DNA encoding C. jejun	7.87e+02
c	721	12	13.8	2694	39	V09364	GMCSF/HIV-1 chimeric	7.87e+02
	722	12	13.8	2712	15	Q87836	Human neuronal calciu	7.87e+02
c	723	12	13.8	2723	2	N70893	MYMV strand (b).	7.87e+02
c	724	12	13.8	2727	1	N91188	BamHI/SalI insert of	7.87e+02
c	725	12	13.8	2733	28	T14604	Human serotonin varia	7.87e+02
c	726	12	13.8	2809	1	N90244	Alpha-factor profibro	7.87e+02
c	727	12	13.8	2809	1	Q02842	pKS311 Plasmid carryi	7.87e+02
c	728	12	13.8	2871	13	Q72501	Mitosis-associated nu	7.87e+02
c	729	12	13.8	2928	5	Q29338	RAR-alpha gene.	7.87e+02
c	730	12	13.8	2940	14	Q81476	RAR-alpha clone lambd	7.87e+02
c	731	12	13.8	2940	1	N90124	DNA of clone phRAR1	7.87e+02
	732	12	13.8	2970	15	Q87835	Human neuronal calciu	7.87e+02
c	733	12	13.8	2973	9	Q53501	Sequence of bean mali	7.87e+02
c	734	12	13.8	2982	1	N91185	BamHI insert from pDM	7.87e+02
c	735	12	13.8	3036	40	V20474	Human PML/RARalpha on	7.87e+02
c	736	12	13.8	3036	5	Q29334	myl/RAR-alpha fusion.	7.87e+02
c	737	12	13.8	3058	36	T77137	Single chain antigen	7.87e+02
	738	12	13.8	3063	17	Q97526	Human DNA repair prot	7.87e+02
c	739	12	13.8	3066	1	Q04025	Gene pJH2 encoding ba	7.87e+02
c	740	12	13.8	3080	1	Q06688	Vascular cell adhesio	7.87e+02
c	741	12	13.8	3084	39	V09955	HIV-1/CD4 chimeric nu	7.87e+02
c	742	12	13.8	3095	18	T02918	Human glucose-6-phosp	7.87e+02
c	743	12	13.8	3095	18	T02922	Human glucose-6-phosp	7.87e+02
c	744	12	13.8	3097	24	T28951	Helicobacter-specific	7.87e+02
c	745	12	13.8	3156	2	N60128	Sequence of the envel	7.87e+02
c	746	12	13.8	3164	33	T85401	Human cadherin-5 codi	7.87e+02
c	747	12	13.8	3164	29	T61921	Full length human cad	7.87e+02
	748	12	13.8	3167	32	T73392	DNA encoding ORF27 an	7.87e+02
c	749	12	13.8	3188	12	Q74631	Bacillus globigii Bgl	7.87e+02
c	750	12	13.8	3188	29	T60035	Blight resistance ind	7.87e+02
	751	12	13.8	3198	38	T92702	Candida CaRhol gene.	7.87e+02
c	752	12	13.8	3201	3	Q15144	pVE36 Bt ICP coding s	7.87e+02
c	753	12	13.8	3201	3	Q14529	pPS029 Bt ICP coding	7.87e+02
	754	12	13.8	3202	29	T47520	Human hepatoma-derive	7.87e+02
c	755	12	13.8	3211	3	N50347	Positve clone (phage	7.87e+02
	756	12	13.8	3212	40	V03313	Archaeoglobus lithotr	7.87e+02

	757	12	13.8	3220	35	T90424	K-ras genomic fragmen 7.87e+02
	758	12	13.8	3267	15	Q88049	Glucan Lyase 1 gene. 7.87e+02
	759	12	13.8	3267	15	Q87603	Fungus-infected G.lem 7.87e+02
c	760	12	13.8	3278	30	T62358	Kluyveromyces lactis 7.87e+02
	761	12	13.8	3294	12	Q70179	Sequence encoding let 7.87e+02
	762	12	13.8	3309	9	Q50646	Human Ews gene (genom 7.87e+02
c	763	12	13.8	3328	24	T43277	DOA4 coding sequence. 7.87e+02
c	764	12	13.8	3365	7	Q41287	Ubiquitin fusion prot 7.87e+02
c	765	12	13.8	3411	24	T28950	Helicobacter-specific 7.87e+02
c	766	12	13.8	3465	1	N90708	Bacillus thuringiensi 7.87e+02
c	767	12	13.8	3465	18	T05271	CryIA(c)/cryI(b) chim 7.87e+02
c	768	12	13.8	3465	20	T18720	CryIA(b)/CryIA(c) chi 7.87e+02
c	769	12	13.8	3468	2	Q03456	Sequence encoding ins 7.87e+02
c	770	12	13.8	3487	3	Q21602	Sequence comprising P 7.87e+02
	771	12	13.8	3500	1	N90229	Mouse tumour necrosis 7.87e+02
c	772	12	13.8	3508	15	Q86693	Candida tropicalis pK 7.87e+02
c	773	12	13.8	3511	22	T33259	PML-RAR-alpha DNA seq 7.87e+02
c	774	12	13.8	3511	22	T33246	PML-RAR-alpha RNA seq 7.87e+02
c	775	12	13.8	3531	13	Q72786	Bacillus thuringiensi 7.87e+02
c	776	12	13.8	3586	10	Q56643	Human tyrosinase gene 7.87e+02
c	777	12	13.8	3596	9	Q52699	Corticosterone-bindin 7.87e+02
c	778	12	13.8	3623	1	N81647	Sequence encoding mod 7.87e+02
c	779	12	13.8	3623	2	N90905	Sequence encoding ins 7.87e+02
c	780	12	13.8	3629	1	N91243	Bacillus thuringiensi 7.87e+02
c	781	12	13.8	3629	16	Q91510	Dextranase gene seque 7.87e+02
c	782	12	13.8	3636	2	N70233	Sequence encoding an 7.87e+02
c	783	12	13.8	3636	1	N81646	Sequence encoding mod 7.87e+02
c	784	12	13.8	3636	2	N90904	Sequence encoding ins 7.87e+02
	785	12	13.8	3637	1	N91475	Human topoisomerase I 7.87e+02
c	786	12	13.8	3639	20	T06481	Cystathionine gamma s 7.87e+02
	787	12	13.8	3645	16	Q91432	Variant human DNA top 7.87e+02
c	788	12	13.8	3670	4	Q22501	Human MIP gene. 7.87e+02
c	789	12	13.8	3675	21	T17045	B.thuringiensis insec 7.87e+02
	790	12	13.8	3699	39	V14340	Plasmid pBSGFP expres 7.87e+02
	791	12	13.8	3817	33	T72814	Human SPT5 cDNA. 7.87e+02
c	792	12	13.8	3826	25	T49478	Bacillus thuringiensi 7.87e+02
c	793	12	13.8	3826	11	Q66292	Crystal delta-endotox 7.87e+02
c	794	12	13.8	3861	1	Q05979	Recombinant AcNPV-HIV 7.87e+02
c	795	12	13.8	3870	2	N70683	Sequence of cry-1-2 g 7.87e+02
c	796	12	13.8	3889	38	T99948	Human N-acetylgalacto 7.87e+02
c	797	12	13.8	3890	1	N91835	B. thuringiensis kurs 7.87e+02
c	798	12	13.8	3954	29	T59975	Full length tipE prot 7.87e+02
	799	12	13.8	3960	8	Q48732	CT gene. 7.87e+02
	800	12	13.8	3974	26	T43360	Cotton FbLate2-82A ge 7.87e+02
	801	12	13.8	3993	37	T88023	Human TRAF5 cDNA. 7.87e+02
c	802	12	13.8	4013	2	N60042	Sequence encoding the 7.87e+02
c	803	12	13.8	4014	24	T42566	Bt2 toxin coding sequ 7.87e+02
c	804	12	13.8	4014	11	Q65577	Bacillus thuringiensi 7.87e+02
c	805	12	13.8	4020	2	N71016	Sequence of LAV/HTLV 7.87e+02
	806	12	13.8	4042	17	T04132	Helicobacter pylori v 7.87e+02
c	807	12	13.8	4061	25	T44518	Hib hxuC + hxuB gene 7.87e+02

c	808	12	13.8	4091	1	Q04035	Stem cell leukaemia (7.87e+02
c	809	12	13.8	4104	12	Q68997	Human protocadherin-4 7.87e+02
c	810	12	13.8	4104	20	T03621	Protocadherin clone 4 7.87e+02
c	811	12	13.8	4113	1	Q06644	Recombinant AcNPV-HIV 7.87e+02
c	812	12	13.8	4118	25	T44520	NTHI hxuC + hxuB gene 7.87e+02
c	813	12	13.8	4131	7	Q43662	Acetobacter cdg3 oper 7.87e+02
	814	12	13.8	4146	37	T93281	Restriction enzyme Ac 7.87e+02
c	815	12	13.8	4156	22	T30166	Human endothelin-bomb 7.87e+02
	816	12	13.8	4165	39	V13998	Human eps15 protein e 7.87e+02
	817	12	13.8	4165	19	T16483	cDNA encoding human E 7.87e+02
c	818	12	13.8	4193	14	Q82834	Human stromalin-2 DNA 7.87e+02
c	819	12	13.8	4199	1	Q05330	Stem cell leukaemia (7.87e+02
	820	12	13.8	4201	1	N90571	Bovine repeat element 7.87e+02
c	821	12	13.8	4201	9	Q50337	Bt2 toxin gene sequen 7.87e+02
	822	12	13.8	4307	4	Q26078	Lambda clone 101. 7.87e+02
	823	12	13.8	4320	32	T73388	DNA encoding HSP72 an 7.87e+02
c	824	12	13.8	4343	10	Q56805	Bacillus thuringiensi 7.87e+02
c	825	12	13.8	4360	32	T36284	DNA encoding crystal 7.87e+02
c	826	12	13.8	4360	7	Q39017	Sequence of CryIA(b) 7.87e+02
c	827	12	13.8	4360	1	N81066	Bacterial insecticida 7.87e+02
c	828	12	13.8	4360	12	Q74441	Bacillus thuringiensi 7.87e+02
c	829	12	13.8	4360	2	N70534	Sequence encoding pro 7.87e+02
c	830	12	13.8	4429	5	Q29685	CCVInsavc-1 spike gen 7.87e+02
c	831	12	13.8	4438	2	Q13073	Modified CryIA(b) gen 7.87e+02
c	832	12	13.8	4472	19	T05406	Human cell cycle gene 7.87e+02
c	833	12	13.8	4488	24	T35520	Human thrombopoietin 7.87e+02
c	834	12	13.8	4527	24	T33899	HIV-1 SF33 genome. 7.87e+02
c	835	12	13.8	4583	1	N90183	DNA sequence of plant 7.87e+02
c	836	12	13.8	4583	29	T59737	Vector pAMVBTS encodi 7.87e+02
c	837	12	13.8	4583	3	Q15174	Plasmid pAMVBT4. 7.87e+02
c	838	12	13.8	4650	12	Q68999	Product of alternativ 7.87e+02
c	839	12	13.8	4650	20	T03623	Alternatively spliced 7.87e+02
c	840	12	13.8	4651	25	T44518	NTHI hxuC + hxuB gene 7.87e+02
	841	12	13.8	4659	1	Q01491	DNA sequence encoding 7.87e+02
c	842	12	13.8	4673	4	Q27189	P. yoelii SSP2 antige 7.87e+02
	843	12	13.8	4682	6	Q34557	GST-II-27 cDNA. 7.87e+02
c	844	12	13.8	4684	29	T65017	Aspergillus oryzae al 7.87e+02
	845	12	13.8	4696	19	T13330	Rat type I insulin-li 7.87e+02
	846	12	13.8	4716	9	Q58610	Mammalian son of seve 7.87e+02
c	847	12	13.8	4739	32	T78203	Physiologically activ 7.87e+02
	848	12	13.8	4747	20	T09019	Arabidopsis thaliana 7.87e+02
	849	12	13.8	4766	10	Q70102	Malarial PfEMP3 epito 7.87e+02
c	850	12	13.8	4800	4	N81533	Sequence of the peplo 7.87e+02
	851	12	13.8	4806	1	N91768	Active region of oste 7.87e+02
c	852	12	13.8	4839	2	Q06827	Alpha amylase pullula 7.87e+02
c	853	12	13.8	5073	31	T71314	cDNA encoding C. eleg 7.87e+02
	854	12	13.8	5300	35	T90447	Thermoanaerobacter th 7.87e+02
	855	12	13.8	5309	2	Q10280	Complete sequence of 7.87e+02
c	856	12	13.8	5376	1	N90732	Sequence of plasmid p 7.87e+02
	857	12	13.8	5495	35	T75590	Canine herpesvirus ge 7.87e+02
c	858	12	13.8	5500	2	N70513	Entire sequence of th 7.87e+02

	859	12	13.8	5501	11	Q64896	HER4. 7.87e+02
	860	12	13.8	5501	19	T18532	Receptor tyrosine kin 7.87e+02
	861	12	13.8	5518	34	T73870	Cotton fibre promoter 7.87e+02
c	862	12	13.8	5518	34	T73870	Cotton fibre promoter 7.87e+02
	863	12	13.8	5534	6	Q35088	Tomato hsp80 genomic 7.87e+02
	864	12	13.8	5547	34	T73865	Cotton fibre promoter 7.87e+02
c	865	12	13.8	5547	34	T73865	Cotton fibre promoter 7.87e+02
	866	12	13.8	5555	11	Q64898	HER4 with alternate 3 7.87e+02
	867	12	13.8	5555	19	T18533	Receptor tyrosine kin 7.87e+02
	868	12	13.8	5579	23	T30056	S. pneumoniae detecti 7.87e+02
c	869	12	13.8	5655	40	V04632	Arabidopsis thaliana 7.87e+02
	870	12	13.8	5674	13	Q72477	Tumour rejection anti 7.87e+02
	871	12	13.8	5674	5	Q32352	MAGE-1 nucleic acid. 7.87e+02
	872	12	13.8	5724	17	Q98902	Tumour rejection anti 7.87e+02
c	873	12	13.8	5775	2	N80920	Sequence encoding the 7.87e+02
	874	12	13.8	5836	33	T84444	Human prostate-specif 7.87e+02
	875	12	13.8	5836	15	Q96296	Human prostrate-speci 7.87e+02
	876	12	13.8	6042	20	T09018	Arabidopsis thaliana 7.87e+02
	877	12	13.8	6060	28	T48454	Human beta-globin gen 7.87e+02
c	878	12	13.8	6126	2	Q13053	CFTR delta I507. 7.87e+02
c	879	12	13.8	6128	2	Q13068	CFTR 556 del A. 7.87e+02
c	880	12	13.8	6128	2	Q13072	CFTR 3659 del C. 7.87e+02
c	881	12	13.8	6129	2	Q13055	CFTR I148T. 7.87e+02
c	882	12	13.8	6129	14	Q84724	DNA encoding an antib 7.87e+02
c	883	12	13.8	6129	2	Q13056	CFTR G178R. 7.87e+02
c	884	12	13.8	6129	15	Q94067	DNA encoding the CFTC 7.87e+02
c	885	12	13.8	6129	2	Q13071	CFTR 1717 -1G -> A. 7.87e+02
c	886	12	13.8	6129	2	Q13066	CFTR Y1092X. 7.87e+02
c	887	12	13.8	6129	2	Q13054	CFTR G85E. 7.87e+02
c	888	12	13.8	6129	2	Q13067	CFTR 129G -> C. 7.87e+02
c	889	12	13.8	6129	2	Q13059	CFTR G542X. 7.87e+02
c	890	12	13.8	6129	2	Q13060	CFTR S549R. 7.87e+02
c	891	12	13.8	6129	14	Q87795	Cystic fibrosis trans 7.87e+02
c	892	12	13.8	6129	2	Q13065	CFTR L1077P. 7.87e+02
c	893	12	13.8	6129	2	Q13064	CFTR P574H. 7.87e+02
c	894	12	13.8	6128	2	Q13061	CFTR G551D. 7.87e+02
c	895	12	13.8	6130	4	Q24280	Cystic fibrosis gene. 7.87e+02
c	896	12	13.8	6146	4	Q23660	Cystic Fibrosis trans 7.87e+02
c	897	12	13.8	6238	39	V14339	Plasmid pFRED7 expres 7.87e+02
	898	12	13.8	6314	7	Q40341	Sequence of a cDNA cl 7.87e+02
	899	12	13.8	6361	39	V14341	Plasmid pFRED13 expre 7.87e+02
	900	12	13.8	6371	22	T30194	Peripheral nervous sy 7.87e+02
	901	12	13.8	6404	22	T30195	Peripheral nervous sy 7.87e+02
	902	12	13.8	6416	4	Q24046	DNA aequence of plasm 7.87e+02
c	903	12	13.8	6767	1	Q05883	Plasmid clone pUC-SK1 7.87e+02
c	904	12	13.8	6785	3	Q20065	Plasmid pAZ112. 7.87e+02
c	905	12	13.8	6795	4	Q27433	HVTA antigen (gA) gen 7.87e+02
c	906	12	13.8	6901	13	Q73703	Rice sucrose phosphat 7.87e+02
	907	12	13.8	7059	10	Q44750	Soybean lipoxygenase 7.87e+02
c	908	12	13.8	7070	15	Q86897	S. cerevisiae fks2 ge 7.87e+02
	909	12	13.8	7076	17	Q86906	Vibrio cholerae disru 7.87e+02

	910	12	13.8	7076	3	Q20186	HlyA gene with ctxB a 7.87e+02
	911	12	13.8	7101	19	T29247	C. difficile toxin B 7.87e+02
	912	12	13.8	7244	40	V15705	Tomato ACC synthase L 7.87e+02
c	913	12	13.8	7257	31	T68656	Green fluorescent pro 7.87e+02
c	914	12	13.8	7294	11	Q66626	Right-half vector M13 7.87e+02
c	915	12	13.8	7294	4	Q24174	M13IX421. 7.87e+02
c	916	12	13.8	7294	11	Q66630	Right-half vector M13 7.87e+02
c	917	12	13.8	7317	37	V03651	M13-based vector M13I 7.87e+02
c	918	12	13.8	7320	4	Q24171	M13IX22. 7.87e+02
c	919	12	13.8	7320	11	Q66627	Left-half vector M13I 7.87e+02
c	920	12	13.8	7394	4	Q24175	M13ED04. 7.87e+02
c	921	12	13.8	7394	11	Q66631	Left-half vector M13E 7.87e+02
c	922	12	13.8	7389	6	Q33479	HDPACK1. 7.87e+02
	923	12	13.8	7400	17	T09310	Tobacco mosaic virus 7.87e+02
c	924	12	13.8	7400	38	T93024	Simian hepatitis A vi 7.87e+02
c	925	12	13.8	7409	4	Q24173	M13IXD03. 7.87e+02
c	926	12	13.8	7409	11	Q66629	Left-half vector M13E 7.87e+02
c	927	12	13.8	7445	11	Q66628	Vector M13IX30 for ra 7.87e+02
c	928	12	13.8	7445	37	V03650	M13-based vector M13I 7.87e+02
c	929	12	13.8	7474	31	T71320	Plasmid pTB73 encodin 7.87e+02
c	930	12	13.8	7557	37	V03653	M13-based vector M13I 7.87e+02
c	931	12	13.8	7566	7	Q42160	Plasmid pPS0212 contg 7.87e+02
c	932	12	13.8	7639	7	Q42159	Plasmid pJD884 contg. 7.87e+02
c	933	12	13.8	7652	7	Q45682	Sequencing vector JAN 7.87e+02
c	934	12	13.8	7705	3	Q15214	Human fibronectin gen 7.87e+02
c	935	12	13.8	7705	1	N70596	Sequence encoding hum 7.87e+02
c	936	12	13.8	7705	1	Q05502	Sequence encoding hum 7.87e+02
c	937	12	13.8	7729	37	V03652	M13-based vector M13I 7.87e+02
c	938	12	13.8	7803	12	Q70009	Fibrinogen-alpha. 7.87e+02
c	939	12	13.8	8316	12	Q74450	myoD retroviral expre 7.87e+02
c	940	12	13.8	8342	13	Q75209	ALL-1 (acute lymphocy 7.87e+02
	941	12	13.8	8353	38	T89396	Methods for diagnosin 7.87e+02
c	942	12	13.8	8391	19	T16333	MLL gene 8.3 kb fragm 7.87e+02
c	943	12	13.8	8392	9	Q53478	MLL gene 8.3 kb BamHI 7.87e+02
	944	12	13.8	8585	1	Q05749	Microspore-specific c 7.87e+02
c	945	12	13.8	8585	28	T48867	Brassica napus micros 7.87e+02
c	946	12	13.8	8710	17	Q89650	pSC11 FIPV E1. 7.87e+02
c	947	12	13.8	8797	27	T50856	DNA cassette for amph 7.87e+02
c	948	12	13.8	8920	11	Q62924	Carbamoyl-phosphate-s 7.87e+02
c	949	12	13.8	9020	17	T07652	pSC11 FIPV N. 7.87e+02
c	950	12	13.8	9088	2	N60476	Sequence of lymphaden 7.87e+02
c	951	12	13.8	9115	12	Q74449	myoD retroviral expre 7.87e+02
c	952	12	13.8	9192	2	Q11943	Nucleotide sequence o 7.87e+02
c	953	12	13.8	9193	2	N60365	Sequence of LAV virus 7.87e+02
c	954	12	13.8	9213	2	N60288	Sequence of the HTLV- 7.87e+02
c	955	12	13.8	9272	13	Q79353	Human genomic clone h 7.87e+02
c	956	12	13.8	9425	7	Q45921	HTLV-III DNA (Clone B 7.87e+02
c	957	12	13.8	9427	7	Q45920	HTLV-III DNA (Clone B 7.87e+02
c	958	12	13.8	9642	31	T71324	Plasmid pTB113 used f 7.87e+02
c	959	12	13.8	9709	16	Q96140	HIV-1 NL4-3 genomic D 7.87e+02
c	960	12	13.8	9709	3	Q22488	HIV-1 proviral clone 7.87e+02

c	961	12	13.8	9718	19	T14180	Attenuated HIV-1 stra 7.87e+02
c	962	12	13.8	9737	5	Q31936	ARV-2 proviral DNA fr 7.87e+02
c	963	12	13.8	9737	2	N60140	Sequence of ARV-2 (9B 7.87e+02
c	964	12	13.8	9745	2	N60240	HTLV-III virus (HIV v 7.87e+02
c	965	12	13.8	9749	7	Q45922	HTLV-III DNA (Clone H 7.87e+02
	966	12	13.8	9919	40	V04631	Arabidopsis thaliana 7.87e+02
c	967	12	13.8	9955	5	Q31625	mGP1-2 containing T7 7.87e+02
c	968	12	13.8	9955	1	Q05877	mGP1-2 plasmid sequen 7.87e+02
c	969	12	13.8	10443	31	T71319	Plasmid pTB72 encodin 7.87e+02
	970	12	13.8	10564	22	T03854	Human fibrinogen gamm 7.87e+02
c	971	12	13.8	10592	35	T75616	Canine herpesvirus ge 7.87e+02
c	972	12	13.8	10974	2	N60055	Sequence of VB131 clo 7.87e+02
c	973	12	13.8	11093	27	T50855	DNA cassette for amph 7.87e+02
c	974	12	13.8	11298	34	T86756	Human high affinity I 7.87e+02
c	975	12	13.8	11357	9	Q51024	Human FcERI beta chai 7.87e+02
c	976	12	13.8	11866	3	N40141	Sequence of human fac 7.87e+02
c	977	12	13.8	11873	3	N40176	Sequence of human fac 7.87e+02
c	978	12	13.8	13058	8	Q48231	Vector containing ACV 7.87e+02
	979	12	13.8	13104	8	Q46852	Clone of recombinant 7.87e+02
c	980	12	13.8	13414	31	T71321	Plasmid pCB50 encodin 7.87e+02
c	981	12	13.8	13585	17	T11549	Tumour rejection anti 7.87e+02
	982	12	13.8	13807	32	T62328	Exopolysaccharide ope 7.87e+02
c	983	12	13.8	15101	18	T16239	Lelystad virus genomi 7.87e+02
c	984	12	13.8	15108	5	Q32002	Lelystad Agent genome 7.87e+02
c	985	12	13.8	15581	39	V14355	Plasmid pNLnSG11 expr 7.87e+02
	986	12	13.8	17410	40	V15205	Human osteogenic prot 7.87e+02
	987	12	13.8	17410	18	T18381	hOP-1 genomic DNA. 7.87e+02
	988	12	13.8	17410	9	Q53142	Sequence encoding ost 7.87e+02
	989	12	13.8	17415	17	T11639	Human osteogenic prot 7.87e+02
	990	12	13.8	19012	6	Q36952	HSA gene. 7.87e+02
c	991	12	13.8	19124	32	T72882	Plasmodium var-7 gene 7.87e+02
c	992	12	13.8	24417	38	T97221	Pseudomonas aeruginos 7.87e+02
	993	12	13.8	28994	39	V15826	Genomic DNA for inter 7.87e+02
c	994	12	13.8	28994	39	V15826	Genomic DNA for inter 7.87e+02
c	995	12	13.8	29392	40	V15422	Mouse poly Ig recepto 7.87e+02
c	996	12	13.8	40352	40	V02032	MAGE-B cluster DNA se 7.87e+02
	997	12	13.8	40875	33	T80043	Insert from cosmid 10 7.87e+02
c	998	12	13.8	43804	38	T86375	Chicken embryo lethal 7.87e+02
	999	12	13.8	133894	17	T13635	AcNPV genomic DNA clo 7.87e+02
c	1000	12	13.8	133894	17	T13635	AcNPV genomic DNA clo 7.87e+02

ALIGNMENTS

```
RESULT    1
ID   T22084 standard; cDNA to mRNA; 88 BP.
AC   T22084;
DT   02-AUG-1996  (first entry)
DE   Human gene signature HUMGS03628.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
```



```
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K, Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 1038; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer.. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   88 BP;    34 A;    7 C;    15 G;    31 T;

  Query Match          100.0%;  Score 87;  DB 19;  Length 88;
  Best Local Similarity 100.0%;  Pred. No. 3.08e-52;
  Matches   88;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 gatcaatctttgatggtgagggttttagaaaggaaaaatatagtaaaatgtgtaatttgt 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      1 GATCAATCTTTGATGGTGAGGGTTTTAGAAAGGAAAAATATAGTAAAATGTGTAATTTGT 60

Db     61 cttaataaaatctatctntacatctaaa 88
          ||||||||||||||||||||||||||||
Qy     61 CTTAATAAAATCTATCTNTACATCTAAA 88


RESULT    2
ID   Q48589 standard; DNA; 1667 BP.
AC   Q48589;
DT   16-FEB-1994  (first entry)
DE   Alcohol oxidase gene promoter.
KW   Expression cassette; expression vector; inducible; methanol;
```



```
KW   glycerol; yeast; ss.
OS   Saccharomyces cerevisiae.
FH   Key             Location/Qualifiers
FT   tata_signal     1329..1335
FT                   /*tag=  a
PN   EP-558024-A.
PD   01-SEP-1993.
PF   26-FEB-1993; 103040.
PR   28-FEB-1992; JP-043361.
PA   (SUNR ) SUNTORY LTD.
PI   Hatanaka H,  Kondo H,  Sakai Y,  Shibano Y,  Tani Y;
DR   WPI; 93-274483/35.
PT   New expression cassettes inducible by methanol and/or glycerol -
PT   having promoter and terminator derived from alcohol oxidase gene
PS   Claim 1; Page 14-15; 37pp; English.
CC   An expression cassette which contains the alcohol oxidase promoter
CC   is induced when in the presence of methanol/glycerol.  Under such
CC   conditions any heterologous gene located downstream of the promoter
CC   will be expressed.  The expression cassette can be used to construct
CC   an expression vector which in turn can be used to transform cells.
CC   When transformed cells are cultured in the presence of methanol/
CC   glycerol expression of the heterologous gene is induced such that
CC   large quantities of the desired product is produced.  The
CC   transformed cells are used particularly for the production of
CC   enzymes such as adenylate kinases, cytochrome C and peroxidases.
SQ   Sequence   1667 BP;    546 A;    276 C;    204 G;    641 T;

  Query Match          19.5%;  Score 17;  DB 8;  Length 1667;
  Best Local Similarity 100.0%;  Pred. No. 1.46e+00;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    802 cacatttactatatttt 818
          |||||||||||||||||
Cp     52 CACATTTTACTATATTT 36


RESULT    3
ID   Q41037 standard; DNA; 69 BP.
AC   Q41037;
DT   29-JUL-1993  (first entry)
DE   pFP-Zspl1 construction oligomer #3.
KW   Fowl pox virus; FPV; P2; promoter; vaccina virus; VV; mutation; apoAI;
KW   multiple cloning site; plasmid; thymidine kinase; tk; Marck's disease;
KW   intergenic region; post-translational modification; factor II; apoAII;
KW   factor V; factor VII; factor VIII; factor IX; tick-borne encephalitis;
KW   hepatitis E; factor X; factor XI; factor XII; factor XIII; protein C;
KW   protein S; von Willebrand factor; hepatitis B; hepatitis C; malaria;
KW   plasminogen; apolipo protein; viral antigen; infectious bronchitis;
KW   TBE; tetanus; HIV; HSV; pertussis; Newcastle disease; vaccine; PCR; ss.
OS   Synthetic.
PN   AU9219580-A.
```



```
PD   04-MAR-1993.
PF   10-JUL-1992; 019580.
PR   26-AUG-1991; EP-114300.
PA   (IMMO ) IMMUNO AG.
PI   Dorner F,  Falkner FG,  Scheiflinger F;
DR   WPI; 93-126452/16.
PT   Prepn. of recombinant fowl pox virus for e.g. expression of
PT   proteins - by inserting foreign DNA into enlarged inter-genic
PT   region so FPV TK-gene remains intact and codes for entire
PT   thymidine kinase
PS   Disclosure; Page 31; 92pp; English.
CC   The sequences given in Q41027-45 are oligomers which were used in the
CC   construction of the insertion plasmids of the invention.  These
CC   plasmids comprised the fowl pox virus (FPV) P2 promoter sequence (see
CC   Q41013 and Q41018-20), multiple cloning sites (see Q41015-16), the FPV
CC   thymidine kinase (tk) gene and the tk gene intergenic region (see also
CC   Q41011-12).  The P2 promoter belongs to one of the strongest natural
CC   promoters in vaccina virus (VV) infected cells.  These plasmids may be
CC   used to express foreign proteins which require post-translational
CC   modification eg. Factors II, V, VII, VIII, IX, X, XI, XII, XIII,
CC   proteins C and S, von Willebrand factor, plasminogen and derivatives,
CC   apolipo proteins, eg. apoAI and apoAII, and viral antigens, eg.
CC   hepatitis B- antigens and the antigens of hepatitis C and E viruses,
CC   of tickborne encephalitis (TBE) virus, antigens of HIV, HSV and whole
CC   or partial sequences of antigens which cause pertussis, infectious
CC   bronchitis, tetanus, malaria, Marck's disease and Newcastle disease,
CC   these antigens being useful as vaccines.
SQ   Sequence   69 BP;    30 A;    10 C;    4 G;    25 T;

  Query Match          18.4%;  Score 16;  DB 6;  Length 69;
  Best Local Similarity 100.0%;  Pred. No. 5.57e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     34 attttactatattttt 49
          ||||||||||||||||
Cp     49 ATTTTACTATATTTTT 34


RESULT    4
ID   T30632 standard; cDNA; 266 BP.
AC   T30632;
DT   21-FEB-1997  (first entry)
DE   Probe nucleic acid #15.
KW   Target binding assembly; nucleic acid recognition unit; NF-kappa-B; SP1;
KW   TATA; human papillomavirus; HPV; HPV E2; human immunodeficiency virus;
KW   HIV LTR; Tat binding unit; probe; assembly sequence; asymmetry sequence;
KW   nuclear localisation signal sequence; human cell; HPV LTR; HIV; bacteria;
KW   virus; ss.
OS   Synthetic.
PN   WO9617956-A2.
PD   13-JUN-1996.
```

```
PF   07-DEC-1995; U15944.
PR   09-DEC-1994; US-353476.
PA   (GENE-) GENE POOL INC.
PI   Weininger AM, Weininger S;
DR   WPI; 96-287199/29.
PT   Probe nucleic acids, target binding assemblies, etc - for detection
PT   and localisation of specific nucleic acid sequences, esp. HIV and
PT   HPV
PS   Disclosure; Page 84; 172pp; English.
CC   T30615-T30634 represent probe nucleic acids of the invention.  The probe
CC   of the invention contains a target binding region (TBR), a booster
CC   binding region (BBR), and an optional support or attachment (OSA).  The
CC   target binding assembly (TBA) recognised by the TBR (see T30581-T30614)
CC   of the probe, contains at least one nucleic acid recognition unit (NAR),
CC   and optionally a linker sequence, an assembly sequence (see
CC   R95994-R95998), an asymmetry sequence (see R95999-R96006), a nuclear
CC   localisation signal sequence (see R96007), and an OSA.  The assembly
CC   sequence and asymmetry sequences are responsible for the folding and
CC   association of the NARs.  The NARs (see R95965-R95993) are selected from
CC   NF-kappa-B, SP1, TATA, human papillomavirus (HPV) E2, HPV LTR, human
CC   immunodeficiency virus (HIV) LTR and Tat binding units.  The linker
CC   sequence is an oligopeptide, which does not interfere with NAR function,
CC   but provides stability and control over the spacing of the NAR from the
CC   rest of the TBA.  The OSA is an attached support or indicator, or other
CC   means of localisation of the probe.  The probe can be used in a method
CC   for detecting or localising a specific target nucleic acid sequence
CC   (TNA).  The method is highly sensitive, and has a high degree of
CC   specificity.  The method can be used for detecting specific nucleic acid
CC   sequences, including those found in human cells, in HIV, HPV, and other
CC   nucleic acid containing systems, including bacteria and viruses.
SQ   Sequence   266 BP;    93 A;    34 C;    56 G;    83 T;

  Query Match          18.4%;  Score 16;  DB 25;  Length 266;
  Best Local Similarity 100.0%;  Pred. No. 5.57e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    155 gaaaaatatagtaaaa 170
          ||||||||||||||||
Qy     33 GAAAAATATAGTAAAA 48


RESULT    5
ID   V18817 standard; DNA; 1095 BP.
AC   V18817;
DT   11-JUN-1998  (first entry)
DE   Human papillomavirus-16 E2 DNA.
KW   Human papillomavirus-16; HPV-16; E2 protein; infection; treatment;
KW   wart; epidermodysplasia verruciformis; laryngeal papilloma;
KW   cervical carcinoma; ds.
OS   Human papillomavirus.
FH   Key             Location/Qualifiers
```



```
FT   CDS             1..1095
FT                   /*tag=  a
FT                   /note= "stop codon not given"
PN   WO9801148-A1.
PD   15-JAN-1998.
PF   08-JUL-1997; U11815.
PR   09-JUL-1996; US-677206.
PA   (HARD ) HARVARD COLLEGE.
PI   Dowhanick JJ, Howley PM;
DR   WPI; 98-100815/09.
DR   P-PSDB; W50005.
PT   Treating papilloma virus infections, by repressing E6 and E7
PT   oncogenic proteins - using E2 polypeptide or nucleic acid encoding
PT   it, useful for treating warts and cervical cancer
PS   Disclosure; Pages 48-49; 71pp; English.
CC   The present sequence encodes the human papillomavirus-16 (HPV-16)
CC   E2 protein.
CC   Papillomavirus infections can be treated by administering an E2
CC   protein, comprising a DNA-binding and transcription activating
CC   domain, or the gene encoding the E2 protein to infected cells. The
CC   method is applicable to livestock, zoo animals, pets and humans,
CC   specifically for the treatment, prevention or reversal of warts,
CC   e.g. plantar, common, Butcher's common, flat or genital warts, or
CC   epidermodysplasia verruciformis, or growth of papillomavirus
CC   transformed cells, especially laryngeal papilloma or cervical
CC   carcinoma.
SQ   Sequence   1095 BP;    397 A;    203 C;    210 G;    285 T;

  Query Match          18.4%;  Score 16;  DB 40;  Length 1095;
  Best Local Similarity 100.0%;  Pred. No. 5.57e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    526 gaaaaatatagtaaaa 541
          ||||||||||||||||
Qy     33 GAAAAATATAGTAAAA 48


RESULT    6
ID   Q32251 standard; cDNA; 1376 BP.
AC   Q32251;
DT   27-APR-1993  (first entry)
DE   BDV structural glycoprotein gE1.
KW   Bovine Viral Diarrhoea Virus; Osloss; gp62; Pestivirus; Togaviridae;
KW   vaccine; ss.
OS   Bovine Viral Diarrhoea Virus.
FH   Key             Location/Qualifiers
FT   cds             10..1371
FT                   /*tag=  a
FT                   /product= gE1
PN   EP-518757-A.
PD   16-DEC-1992.
```



```
PF   10-JUN-1992; 401601.
PR   11-JUN-1991; FR-007077.
PA   (INMR ) RHONE MERIEUX SA.
PA   (EUBI-) SOC EURO BIOTECHNOLOGIE DROIT BELGE SA.
PI   Chappuis G, Pin J, Lecomte CMTG, Renard AJJ;
DR   WPI; 92-417606/51.
DR   P-PSDB; R29511.
PT   New nucleic acid sequences encoding bovine diarrhoea virus
PT   glycoprotein(s) - and derived, recombinant viruses, expression
PT   vectors, proteins etc., useful in protective vaccines
PS   Claim 3; Page 21-22; 37pp; French.
CC   This sequence encodes the gE1 structural protein from the Osloss
CC   strain of BDV (gE1 is called gp62 by Collett et al., Virology 165:
CC   200-208, 1988). The epitopes of neutralising antibodies which
CC   recognise BDV have been localised to the viral structural proteins.
CC   The invention includes sequences with homology to the nucleotide
CC   sequence encoding gE1, at least the first 273 amino acids of gE2
CC   (also called gp53) and combinations of gE1 with at least part of
CC   gE2 and opt. the other structural (non-glycosylated) protein C
CC   (also called p20).  See also Q32249-Q32255.
SQ   Sequence   1376 BP;    402 A;    289 C;    365 G;    320 T;

  Query Match          18.4%;  Score 16;  DB 5;  Length 1376;
  Best Local Similarity 100.0%;  Pred. No. 5.57e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    268 aagacaaattacacat 283
          ||||||||||||||||
Cp     63 AAGACAAATTACACAT 48


RESULT    7
ID   Q32252 standard; cDNA; 2210 BP.
AC   Q32252;
DT   27-APR-1993  (first entry)
DE   Encodes BDV structural glycoproteins gE1 and part of gE2.
KW   Bovine Viral Diarrhoea Virus; Osloss; gp53; Pestivirus; Togaviridae;
KW   vaccine; gp62; ss.
OS   Bovine Viral Diarrhoea Virus.
FH   Key             Location/Qualifiers
FT   cds             10..2205
FT                   /*tag=  a
FT                   /note= "encodes gE1 and first part of gE2"
PN   EP-518757-A.
PD   16-DEC-1992.
PF   10-JUN-1992; 401601.
PR   11-JUN-1991; FR-007077.
PA   (INMR ) RHONE MERIEUX SA.
PA   (EUBI-) SOC EURO BIOTECHNOLOGIE DROIT BELGE SA.
PI   Chappuis G, Pin J, Lecomte CMTG, Renard AJJ;
DR   WPI; 92-417606/51.
```



```
DR   P-PSDB; R29512.
PT   New nucleic acid sequences encoding bovine diarrhoea virus
PT   glycoprotein(s) - and derived, recombinant viruses, expression
PT   vectors, proteins etc., useful in protective vaccines
PS   Claim 6; Page 23-25; 37pp; French.
CC   This sequence contains a translation initiation codon, the gE1
CC   coding sequence, the first 273 codons of gE2 and a translation stop
CC   codon. The structural glycoprotein coding sequences are from the
CC   Osloss strain of BDV (gE1 is called gp62 and gE2 is called gp53
CC   by Collett et al., Virology 165: 200-208, 1988). The epitopes of
CC   neutralising antibodies which recognise BDV have been localised
CC   to the viral structural proteins. The invention includes sequences
CC   with homology to the nucleotide sequence encoding at least the
CC   first 273 amino acids of gE2, of gE1 alone and of the combination
CC   of gE1 and gE2 with the other structural protein C (also called
CC   p20).  See also Q32249-Q32255.
SQ   Sequence   2210 BP;    637 A;    457 C;    603 G;    513 T;

  Query Match          18.4%;  Score 16;  DB 5;  Length 2210;
  Best Local Similarity 100.0%;  Pred. No. 5.57e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    268 aagacaaattacacat 283
          ||||||||||||||||
Cp     63 AAGACAAATTACACAT 48


RESULT    8
ID   T59675 standard; DNA; 2406 BP.
AC   T59675;
DT   09-MAY-1997  (first entry)
DE   DNA encoding hRIP, a protein critical for HIV replication.
KW   hRIP; HIV; human immunodeficiency virus; replication; Rev;
KW   retrovirus; antiviral; virucide; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             67..1755
FT                   /*tag=  a
FT   repeat_unit     726..727
FT                   /*tag=  b
FT                   /rpt_type= DIRECT
FT                   /note= "R1"
FT   repeat_unit     783..794
FT                   /*tag=  c
FT                   /rpt_type= DIRECT
FT                   /note= "R1"
FT   repeat_unit     802..806
FT                   /*tag=  d
FT                   /rpt_type= DIRECT
FT                   /note= "PG"
FT   repeat_unit     828..833
```

```
FT                   /*tag=  e
FT                   /rpt_type= DIRECT
FT                   /note= "FG"
FT   repeat_unit     909..920
FT                   /*tag=  f
FT                   /rpt_type= DIRECT
FT                   /note= "R1"
FT   repeat_unit     941..946
FT                   /*tag=  g
FT                   /rpt_type= DIRECT
FT                   /note= "FG"
FT   repeat_unit     1038..1067
FT                   /*tag=  h
FT                   /rpt_type= DIRECT
FT                   /note= "R2"
FT   repeat_unit     1121..1125
FT                   /*tag=  i
FT                   /rpt_type= DIRECT
FT                   /note= "FG"
FT   repeat_unit     1180..1208
FT                   /*tag=  j
FT                   /rpt_type= DIRECT
FT                   /note= "R2"
FT   repeat_unit     1278..1283
FT                   /*tag=  k
FT                   /rpt_type= DIRECT
FT                   /note= "FG"
FT   repeat_unit     1345..1349
FT                   /*tag=  l
FT                   /rpt_type= DIRECT
FT                   /note= "FG"
FT   repeat_unit     1360..1373
FT                   /*tag=  m
FT                   /rpt_type= DIRECT
FT                   /note= "R3"
FT   repeat_unit     1405..1418
FT                   /*tag=  n
FT                   /rpt_type= DIRECT
FT                   /note= "R3"
FT   repeat_unit     1453..1457
FT                   /*tag=  o
FT                   /rpt_type= DIRECT
FT                   /note= "FG"
FT   repeat_unit     1486..1490
FT                   /*tag=  p
FT                   /rpt_type= DIRECT
FT                   /note= "FG"
FT   repeat_unit     1621..1625
FT                   /*tag=  q
FT                   /rpt_type= DIRECT
FT                   /note= "FG"
```



```
FT   repeat_unit     1650..1635
FT                   /*tag=  r
FT                   /rpt_type= DIRECT
FT                   /note= "FG"
FT   repeat_unit     1678..1691
FT                   /*tag=  s
FT                   /rpt_type= DIRECT
FT                   /note= "R3"
FT   repeat_unit     1735..1748
FT                   /*tag=  t
FT                   /rpt_type= DIRECT
FT                   /note= "R3"
PN   WO9701648-A1.
PD   16-JAN-1997.
PF   27-JUN-1996; U11078.
PR   27-JUN-1995; US-000551.
PA   (UYMA-) UNIV MASSACHUSETTS MEDICAL CENT.
PI   Frits CC,  Green MR;
DR   WPI; 97-100218/09.
DR   P-PSDB; W12713.
PT   New isolated protein necessary for HIV replication - which mediates
PT   the Rev response, useful for developing prods. for the study and
PT   treatment of viral infection
PS   Example 1; Page 37-40; 73pp; English.
CC   A DNA sequence (T59675) codes for hRIP (W12713), a novel human
CC   cellular cofactor required to mediate the Rev response necessary
CC   for HIV replication.  hRIP binds to HIV Rev protein and is
CC   essential for Rev activity.  Full-length HIV-1, Rev mutants and
CC   HTLV-I Rex proteins were fused to LexA (1-202) by PCR.  A two-
CC   hybrid screen was performed using an activation domain-tagged HeLa
CC   cDNA library and the LexA construct as a bait in yeast strain
CC   EGY48.  Primary positive plasmids were retransformed into EGY48
CC   expressing LexA, LexA-Rev or LexA-revM10 and contg. the lacZ
CC   reporter plasmid pSH18-34.  A clone contg. a hRIP fragment was
CC   obtd. and used to screen a human placental cDNA library to obtain
CC   the full-length hRIP DNA.  The DNA can be used to produce
CC   recombinant hRIP in host cells for use in developing prods. for
CC   the study and treatment of retroviral infections.
SQ   Sequence   2406 BP;    686 A;    575 C;    494 G;    651 T;

  Query Match           18.4%;  Score 16;  DB 28;  Length 2406;
  Best Local Similarity 100.0%;  Pred. No. 5.57e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2245 atttttcctttctaaa 2260
          ||||||||||||||||
Cp     39 ATTTTTCCTTTCTAAA 24


RESULT   9
ID   Q32250 standard; cDNA; 2487 BP.
```



```
AC   Q32250;
DT   27-APR-1993  (first entry)
DE   Encodes BDV structural glycoproteins gE1 and gE2.
KW   Bovine Viral Diarrhoea Virus; Osloss; gp53; Pestivirus; Togaviridae;
KW   vaccine; gp62; ss.
OS   Bovine Viral Diarrhoea Virus.
FH   Key             Location/Qualifiers
FT   cds             10..2481
FT                   /*tag=  a
FT                   /product= gE1_and_gE2
PN   EP-518757-A.
PD   16-DEC-1992.
PF   10-JUN-1992; 401601.
PR   11-JUN-1991; FR-007077.
PA   (INMR ) RHONE MERIEUX SA.
PA   (EUBI-) SOC EURO BIOTECHNOLOGIE DROIT BELGE SA.
PI   Chappuis G, Pin J, Lecomte CMTG, Renard AJJ;
DR   WPI; 92-417606/51.
DR   P-PSDB; R29510.
PT   New nucleic acid sequences encoding bovine diarrhoea virus
PT   glycoprotein(s) - and derived, recombinant viruses, expression
PT   vectors, proteins etc., useful in protective vaccines
PS   Claim 5; Page 18-20; 37pp; French.
CC   This sequence contains a translation initiation codon, the gE1
CC   coding sequence, the gE2 coding sequence and a translation stop
CC   codon. The structural glycoprotein coding sequences are from the
CC   Osloss strain of BDV (gE1 is called gp62 and gE2 is called gp53
CC   by Collett et al., Virology 165: 200-208, 1988). The epitopes of
CC   neutralising antibodies which recognise BDV have been localised
CC   to the viral structural proteins. The invention includes sequences
CC   with homology to the nucleotide sequence encoding at least the
CC   first 273 amino acids of gE2, of gE1 alone and of the combination
CC   of gE1 and gE2 with the other structural protein C (also called
CC   p20).  See also Q32249-Q32255.
SQ   Sequence   2487 BP;    721 A;    517 C;    665 G;    584 T;

  Query Match           18.4%;  Score 16;  DB 5;  Length 2487;
  Best Local Similarity 100.0%;  Pred. No. 5.57e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    268 aagacaaattacacat 283
          ||||||||||||||||
Cp     63 AAGACAAATTACACAT 48


RESULT   10
ID   T79881 standard; cDNA; 2583 BP.
AC   T79881;
DT   03-DEC-1997  (first entry)
DE   cDNA encoding HIV Rev-1 activation domain binding protein.
KW   HIV; human immunodeficiency virus; lentivirus; inhibitor; antisense;
```



```
KW   AIDS; acquired immune deficiency service; Rex; Rev; Rab; HTLV-1;
KW   human T cell leukaemia virus; ss.
OS   Human immunodeficiency virus.
FH   Key             Location/Qualifiers
FT   cds             244..1932
FT                   /*tag=  a
PN   WO9706257-A1.
PD   20-FEB-1997.
PF   09-AUG-1996; U12986.
PR   10-AUG-1995; US-513570.
PA   (UYDU-) UNIV DUKE.
PI   Bogerd HP,  Cullen BR;
DR   WPI; 97-154264/14.
DR   P-PSDB; W25156.
PT   New DNA encoding cellular co:factor for HIV Rev and related proteins
PT   - used for screening cpds. able to modulate rev-co:factor
PT   interaction for control of HIV and other viral infections
PS   Claim 3; Page 33-36; 55pp; English.
CC   T79881 encodes a cellular co-factor for HIV-1 Rev and other related
CC   proteins. The protein is called Rab (Rev/Rex activation domain
CC   binding) protein as it bind most strongly to HIV-1 Rev and HTLV-1
CC   Rex. HIV-1 Rev is required for the nucleocytoplasmic transport and
CC   hence translation of a class of incompletely spliced mRNAs that
CC   encode the viral structural proteins. In the absence of Rev, these late
CC   viral RNAs remain sequestered in the nucleus until they are either
CC   spliced or degraded. Rev function is therefore essential for the
CC   production of progeny virions by HIV-1 infected cells. Rex has an
CC   equivalent function for HTLV-1. Rab and DNA encoding it can be used
CC   for screening for compounds that inhibit formation of complexes
CC   between Rev/Rex and Rab, preventing activation of these proteins.
CC   Such compounds are useful for treating lentivirus or HTLV infection.
SQ   Sequence   2583 BP;    702 A;    640 C;    572 G;    669 T;

  Query Match           18.4%;  Score 16;  DB 34;  Length 2583;
  Best Local Similarity 100.0%;  Pred. No. 5.57e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2422 atttttcctttctaaa 2437
          ||||||||||||||||
Cp     39 ATTTTTCCTTTCTAAA 24


RESULT   11
ID   T61925 standard; cDNA; 2625 BP.
AC   T61925;
DT   14-MAY-1997  (first entry)
DE   Full length human cadherin-11 cDNA.
KW   Ca2+ dependent; cell adhesion protein; foetal; cadherin; rat;
KW   brain; human; antibody; purification; determination;
KW   tissue expression; binding antagonist; calcium ion; ss.
OS   Homo sapiens.
```

```
FH   Key             Location/Qualifiers
FT   cds             156..2546
FT                   /*tag=  a
PN   US5597725-A.
PD   28-JAN-1997.
PF   17-APR-1992; 872643.
PR   17-APR-1992; US-872643.
PR   19-APR-1993; US-049460.
PR   26-JAN-1994; US-188228.
PA   (DOHE-) DOHENY EYE INST.
PI   Suzuki S;
DR   WPI; 97-108328/10.
DR   P-PSDB; W13134.
PT   Antibodies to cadherin proteins - useful as cadherin antagonists,
PT   etc.
PS   Example 2; Columns 91-96; 59pp; English.
CC   The present sequence encodes full length human cadherin-11, which
CC   is a Ca2+ dependent cell adhesion protein. The human cadherin cDNA
CC   was isolated from a foetal brain cDNA library, using probes based
CC   on homologous rat cadherin cDNA.
CC   Antibodies or fragments that specifically bind the human cadherin
CC   can be used to purify the cadherin, determine its tissue expression
CC   and antagonise its ligand/antiligand binding activities.
SQ   Sequence   2625 BP;    730 A;    661 C;    675 G;    559 T;

  Query Match           18.4%;  Score 16;  DB 28;  Length 2625;
  Best Local Similarity 100.0%;  Pred. No. 5.57e+00;
  Matches    16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1914 accctcaccatcaaag 1929
         ||||||||||||||||
Cp     23 ACCCTCACCATCAAAG 8


RESULT   12
ID   T85403 standard; cDNA; 2625 BP.
AC   T85403;
DT   03-NOV-1997  (first entry)
DE   Human cadherin-11 coding sequence.
KW   Human; cadherin; rat; calcium-dependent cell adhesion protein;
KW   superfamily; cytoskeleton; eatenin; cancer; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             156..2546
FT                   /*tag=  a
FT                   /product= Cadherin-11
PN   US5646250-A.
PD   08-JUL-1997.
PF   17-APR-1992; 872643.
PR   19-APR-1993; US-049460.
PR   17-APR-1992; US-872643.
```



```
PR   01-NOV-1994; US-332638.
PA   (DOHE-) DOHENY EYE INST.
PI   Suzuki S;
DR   WPI; 97-362997/33.
PT   Human and rat cadherin polypeptide(s) - mediate cell-cell adhesion
PS   Example 2; Column 87-90; 56pp; English.
CC   This sequence encodes human cadherin-11.  The invention specifically
CC   provides details of human cadherin-5, -8, -11, -12 and -13, and rat
CC   cadherin-8, -5, -11 and -13.  Cadherins are calcium-dependent cell
CC   adhesion proteins.  They are glycosylated integral membrane proteins
CC   that have an N-terminal extracellular domain that determines binding
CC   specificity, a hydrophobic membrane spanning region and a C-terminal
CC   cytoplasmic domain, which is highly conserved among members of the
CC   superfamily.  The C-terminal domain interacts with the cytoskeleton
CC   through eatenins and other cytoskeleton-associated proteins.  The
CC   novel cadherin proteins may be used in the analysis of the role of
CC   cadherins in various cancers.  Sequence analysis of the cadherin
CC   proteins also allows investigation of the structure and function of
CC   cadherin.  The cadherin proteins may be isolated by using anti-cadherin
CC   antibodies.  These antibodies may also be used to modulate the activity
CC   of cadherin and to determine the tissue specific distribution of cadherin
CC   proteins.  Each subclass of cadherins has a unique tissue distribution
CC   pattern.
SQ   Sequence   2625 BP;    730 A;    661 C;    675 G;    559 T;

  Query Match           18.4%;  Score 16;  DB 33;  Length 2625;
  Best Local Similarity 100.0%;  Pred. No. 5.57e+00;
  Matches    16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1914 accctcaccatcaaag 1929
         ||||||||||||||||
Cp     23 ACCCTCACCATCAAAG 8


RESULT   13
ID   Q32253 standard; cDNA; 3224 BP.
AC   Q32253;
DT   27-APR-1993  (first entry)
DE   Encodes fusion of BDV structural proteins C, gE1 and gE2.
KW   Bovine Viral Diarrhoea Virus; Osloss; p20; Pestivirus; Togaviridae;
KW   vaccine; gp53; gp62; ss.
OS   Bovine Viral Diarrhoea Virus.
FH   Key             Location/Qualifiers
FT   cds             9..3218
FT                   /*tag=  a
PN   EP-518757-A.
PD   16-DEC-1992.
PF   10-JUN-1992; 401601.
PR   11-JUN-1991; FR-007077.
PA   (INMR ) RHONE MERIEUX SA.
PA   (EUBI-) SOC EURO BIOTECHNOLOGIE DROIT BELGE SA.
```



```
PI   Chappuis G, Pin J, Lecomte CMTG, Renard AJJ;
DR   WPI; 92-417606/51.
DR   P-PSDB; R29513.
PT   New nucleic acid sequences encoding bovine diarrhoea virus
PT   glycoprotein(s) - and derived, recombinant viruses, expression
PT   vectors, proteins etc., useful in protective vaccines
PS   Disclosure; Page 26-29; 37pp; French.
CC   This fusion sequence encodes a translation initiation
CC   codon-sequence coding for C-sequence coding for gE1-sequence coding
CC   for gE2-stop codon. The sequences are from the Osloss strain of BDV
CC   (protein C is called p20, gE1 is called gp62 and gE2 is called gp53
CC   by Collett et al., Virology 165: 200-208, 1988). The epitopes of
CC   neutralising antibodies which recognise BDV have been localised to
CC   the viral structural proteins. See also Q32249-Q32255.
SQ   Sequence   3224 BP;    989 A;    656 C;    851 G;    728 T;

  Query Match           18.4%;  Score 16;  DB 5;  Length 3224;
  Best Local Similarity 100.0%;  Pred. No. 5.57e+00;
  Matches    16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1005 aagacaaattacacat 1020
         ||||||||||||||||
Cp     63 AAGACAAATTACACAT 48


RESULT   14
ID   Q44392 standard; cDNA to mRNA; 3712 BP.
AC   Q44392;
DT   14-SEP-1994  (first entry)
DE   Sequence of human OSF-4-1 cDNA.
KW   OSF-4-1; cadherin; growth factor; osteogenesis; osteoblast; therapy;
KW   diagnosis; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             461..2851
FT                   /*tag=  a
PN   EP-585801-A.
PD   09-MAR-1994.
PF   25-AUG-1993; 113602.
PR   28-AUG-1992; JP-230028.
PA   (FARH ) HOECHST JAPAN LTD.
PI   Amann E,  Kawai S,  Okazaki M,  Takeshita S,  Tsujimura A;
DR   WPI; 94-076152/10.
DR   P-PSDB; R49731.
PT   New bone related, cadherin-like OSF-4 proteins - for treatment
PT   and diagnosis of bone metabolic disease, and nucleic acid
PT   encoding them
PS   Claim 3; Page 18-22; 34pp; English.
CC   cDNA libraries were constructed from the mouse osteoblastic cell
CC   line MC3T3-E1 and from mouse liver tissue, amplified by PCR, and
CC   then as much common DNA as possible removed by hybridisation between
```



```
CC   the 2 libraries. Residual E1-specific DNA was amplified, inserted
CC   into lambda gt10 and screened by plaque hybridisation. A minibank of
CC   273 E-specific clones was recovered, their inserts amplified and
CC   used to screen total RNA from both cell types. One clone specific
CC   for E1 was identified and sequenced. The insert from this clone was
CC   used to screen cDNA prepd. from E1 RNA and the longest posn. insert
CC   cloned in pGEM 112f (+) to give pKOT164. This insert was sequenced;
CC   it encoded the 796 AA mouse precursor protein (Q44391/R49730). The
CC   insert was also used to screen a cDNA bank prepd. from human
CC   osteosarcoma to identify 2 clones encoding the 2 human precursor
CC   proteins - OSF-4-1 and OSF-4-2 (Q44392/R49731 and Q44393/R49732
CC   respectively).
SQ   Sequence   3712 BP;    1074 A;    856 C;    897 G;    885 T;

  Query Match           18.4%;  Score 16;  DB 10;  Length 3712;
  Best Local Similarity 100.0%;  Pred. No. 5.57e+00;
  Matches    16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2219 accctcaccatcaaag 2234
         ||||||||||||||||
Cp     23 ACCCTCACCATCAAAG 8


RESULT   15
ID   Q44393 standard; cDNA to mRNA; 3914 BP.
AC   Q44393;
DT   14-SEP-1994  (first entry)
DE   Sequence of human OSF-4-2 cDNA.
KW   OSF-4-2; cadherin; growth factor; osteogenesis; osteoblast; therapy;
KW   diagnosis; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             491..2569
FT                   /*tag=  a
PN   EP-585801-A.
PD   09-MAR-1994.
PF   25-AUG-1993; 113602.
PR   28-AUG-1992; JP-230028.
PA   (FARH ) HOECHST JAPAN LTD.
PI   Amann E,  Kawai S,  Okazaki M,  Takeshita S,  Tsujimura A;
DR   WPI; 94-076152/10.
DR   P-PSDB; R49732.
PT   New bone related, cadherin-like OSF-4 proteins - for treatment
PT   and diagnosis of bone metabolic disease, and nucleic acid
PT   encoding them
PS   Claim 3; Page 23-27; 34pp; English.
CC   cDNA libraries were constructed from the mouse osteoblastic cell
CC   line MC3T3-E1 and from mouse liver tissue, amplified by PCR, and
CC   then as much common DNA as possible removed by hybridisation between
CC   the 2 libraries. Residual E1-specific DNA was amplified, inserted
CC   into lambda gt10 and screened by plaque hybridisation. A minibank of
```

```
CC   273 E-specific clones was recovered, their inserts amplified and
CC   used to screen total RNA from both cell types. One clone specific
CC   for E1 was identified and sequenced. The insert from this clone was
CC   used to screen cDNA prepd. from E1 RNA and the longest posn. insert
CC   cloned in pGEM 112f (+) to give pKOT164. This insert was sequenced;
CC   it encoded the 796 AA mouse precursor protein (Q44391/R49730). The
CC   insert was also used to screen a cDNA bank prepd. from human
CC   osteosarcoma to identify 2 clones encoding the 2 human precursor
CC   proteins - OSF-4-1 and OSF-4-2 (Q44392/R49731 and Q44393/R49732
CC   respectively).
SQ   Sequence   3914 BP;    1105 A;    920 C;    947 G;    942 T;

  Query Match           18.4%;  Score 16;  DB 10;  Length 3914;
  Best Local Similarity 100.0%;  Pred. No. 5.57e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2249 accctcaccatcaaag 2264
          ||||||||||||||||
Cp     23 ACCCTCACCATCAAAG 8


RESULT   16
ID   T36899 standard; cDNA; 4057 BP.
AC   T36899;
DT   22-NOV-1996  (first entry)
DE   Mouse neuron restrictive silencer factor cDNA.
KW   Neuron restrictive silencer factor; NRSF; differentiation;
KW   drug screening; ss.
OS   Mus musculus.
FH   Key             Location/Qualifiers
FT   misc_difference 306
FT                   /*tag= a
FT                   /note= "base 306 is given as r in Fig 12"
FT   cds             356..3286
FT                   /*tag= b
FT   misc_difference 429
FT                   /*tag= c
FT                   /note= "base 429 is given as s in Fig 12"
FT   misc_difference 3686
FT                   /*tag= d
FT                   /note= "base 3686 is given as y in Fig 12"
FT   misc_difference 4045
FT                   /*tag= e
FT                   /note= "base 4045 is given as s in Fig 12"
PN   WO9627665-A2.
PD   12-SEP-1996.
PF   01-MAR-1996; U02817.
PR   03-MAR-1995; US-398590.
PA   (CALY ) CALIFORNIA INST OF TECHNOLOGY.
PI   Anderson DJ,  Schoenherr CJ;
DR   WPI; 96-425426/42.
```



```
DR   P-PSDB; W02289.
PT   Recombinant neuron restrictive silencer factor protein - useful to
PT   control neuron cell precursor differentiation, and in drug screening
PS   Claim 6; Fig 12A-E; 59pp; English.
CC   A DNA clone (T36899) codes for mouse neuron-restrictive silencer
CC   factor (NRSF) (W02289), a protein that silences or suppresses the
CC   expression of neuron-specific genes.  The DNA may be incorporated
CC   into a vector and used for prodn. of recombinant NRSF in transformed
CC   host cells.  A partial sequence (T36898) for the human NRSF gene
CC   was also obtd.  NRSF genes contg. significant deletions may be used
CC   to disrupt expression of NRSF protein, thus allowing the expression
CC   of neuronal genes.
SQ   Sequence   4057 BP;    1123 A;    988 C;    1102 G;    840 T;

  Query Match           18.4%;  Score 16;  DB 23;  Length 4057;
  Best Local Similarity 100.0%;  Pred. No. 5.57e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   3169 ttttagaaaggaaaaa 3184
          ||||||||||||||||
Qy     23 TTTTAGAAAGGAAAAA 38


RESULT   17
ID   T09847 standard; DNA; 7902 BP.
AC   T09847;
DT   02-JAN-1997  (first entry)
DE   Human papilloma virus 16 genomic DNA.
KW   Human papilloma virus; HPV; ribozyme; cleavage; E6; E7;
KW   keratinocyte; infection; genital warts; neoplasia; cervical cancer;
KW   warts; ds.
OS   Human papilloma virus 16.
PN   WO9531552-A2.
PD   23-NOV-1995.
PF   15-MAY-1995; U06016.
PR   13-MAY-1994; US-242665.
PA   (UYNI-) UNIV NORTHERN ILLINOIS.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI   Dipaolo J,  Hampel A,  Siwkowski A;
DR   WPI; 96-010937/01.
PT   Synthetic ribozyme(s) for human papilloma virus (HPV) - useful in
PT   treatment of cervical cancer and in detection of HPV
PS   Disclosure; Page 25-29; 60pp; English.
CC   Synthetic ribozymes directed against human papilloma virus (HPV) 16
CC   may be used to cleave HPV or to detect HPV in human tissue.  The
CC   ribozymes cleave HPV to lower the production of the E6 and E7
CC   proteins which play a key role in keratinocyte transformation.  They
CC   can be used for the treatment of HPV infections which are associated
CC   with genital warts and genital neoplasms.  In particular they can be
CC   used for the treatment of cervical cancer optionally with other
CC   agents such as cisplatin.  Ribozymes are described in T40306-T40315.
```



```
SQ   Sequence   7902 BP;    2601 A;    1376 C;    1507 G;    2418 T;

  Query Match           18.4%;  Score 16;  DB 24;  Length 7902;
  Best Local Similarity 100.0%;  Pred. No. 5.57e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   3278 gaaaaatatagtaaaa 3293
          ||||||||||||||||
Qy     33 GAAAAATATAGTAAAA 48


RESULT   18
ID   T94723 standard; DNA; 7904 BP.
AC   T94723;
DT   16-FEB-1998  (first entry)
DE   Human papillomavirus type 16 E6 complete genome.
KW   Complete genome; circular; human papillomavirus type 16; HPV16 E6;
KW   cervical dysplasia; cervical cancer; cervical smear; ds.
OS   Human papillomavirus type 16.
FH   Key             Location/Qualifiers
FT   TATA_signal     17..23
FT                   /*tag= a
FT   TATA_signal     65..71
FT                   /*tag= b
FT   CDS             83..559
FT                   /*tag= c
FT                   /product= E6_transforming_protein
FT                   /note= "E6 ORF from 65 to 559; putative"
FT   mutation        350
FT                   /*tag= d
FT                   /note= "Mutated from the normal T (as in this case)
FT                        to G in the variant"
FT   CDS             562..858
FT                   /*tag= e
FT                   /product= E7_transforming_protein
FT                   /note= "E7 ORF 544 to 858; putative"
FT   CDS             865..2813
FT                   /*tag= f
FT                   /product= E1_replication_protein
FT                   /note= "E1 interrupted ORF from 859 to 2813; putative"
FT   CDS             2755..3852
FT                   /*tag= g
FT                   /product= E2_regulatory_protein
FT                   /note= "E2 ORF from 2725 to 3852; putative"
FT   CDS             3332..3619
FT                   /*tag= h
FT                   /product= E4
FT                   /note= "E4 ORF from 3332 to 3619; putative"
FT   CDS             3863..4099
FT                   /*tag= i
FT                   /product= E5
```



```
FT                   /note= "E5 ORF from 3863 to 4099; putative"
FT   polyA_signal    4213..4218
FT                   /*tag= j
FT                   /note= "Putative"
FT   CDS             4235..5656
FT                   /*tag= k
FT                   /product= L2_minor_capsid_protein
FT                   /note= "L2 ORF from 4133 to 5656; putative"
FT   TATA_signal     4289..4295
FT                   /*tag= l
FT   CDS             5559..7154
FT                   /*tag= m
FT                   /product= L1_major_capsid_protein
FT                   /note= "L1 ORF from 5526 to 7154; putative"
FT   polyA_signal    7260..7265
FT                   /*tag= n
PN   US5679509-A.
PD   21-OCT-1997.
PF   28-SEP-1993; 127906.
PR   30-SEP-1994; US-316239.
PR   28-SEP-1993; US-127906.
PA   (UYNE-) UNIV NEW MEXICO STATE.
PI   Parmenter CA,  Wheeler CM;
DR   WPI; 97-525714/48.
DR   P-PSDB; W35741.
PT   Evaluating risk of cervical dysplasia or cervical cancer - by
PT   deteting variant form of human papilloma virus 16
PS   Claim 1; Column 9-16; 33pp; English.
CC   Methods have been developed for distinguishing a subset of human
CC   papilloma virus (HPV) that is associated with an increased risk of
CC   developing cervical dysplasia or cervical cancer. The methods involve:
CC   (1) preparing a cervical sample to expose any HPV-16 E6 gene in the
CC   sample and determining if the base at position 350 of the E6 gene
CC   (see T94723 and T94724 for comparison) is T or G, where the presence of
CC   G at position 350 is associated with an increased risk of developing
CC   cervical dysplasia or cervical cancer; and (2) preparing a cervical
CC   sample to expose any HPV-16 E6 protein in the sample and determining
CC   if the amino acid at position 83 of the protein (see position 90 in
CC   W35741 and w35742 for comparison) is Val or Leu, where the presence of
CC   Val at position 83 that is associated with an increased risk of
CC   developing cervical dysplasia or cervical cancer. The present sequence
CC   represents the reference nucleotide sequence for HPV-16 E6. The 350G
CC   variant correlates well with Pap scores: 350T:350G ratios among 45
CC   HPV16 samples were 10:4 for negative Pap scores; 4:2 for CIN I, 1:6 for
CC   CIN II; 2:9 for CIN III; 0:3 for cancer.
SQ   Sequence   7904 BP;    2601 A;    1377 C;    1509 G;    2417 T;

  Query Match           18.4%;  Score 16;  DB 35;  Length 7904;
  Best Local Similarity 100.0%;  Pred. No. 5.57e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```

```
Db    3280 gaaaaatatagtaaaa 3295
           ||||||||||||||||
Qy      33 GAAAAATATAGTAAAA 48


RESULT   19
ID   T94724 standard; DNA; 7904 BP.
AC   T94724;
DT   16-FEB-1998  (first entry)
DE   Human papillomavirus type 16 E6 variant nucleotide sequence.
KW   Complete genome; circular; human papillomavirus type 16; HPV16 E6;
KW   cervical dysplasia; cervical cancer; cervical smear; ds.
OS   Human papillomavirus type 16.
FH   Key             Location/Qualifiers
FT   TATA_signal      17..23
FT                    /*tag=  a
FT   TATA_signal      65..71
FT                    /*tag=  b
FT   CDS              83..559
FT                    /*tag=  c
FT                    /product= E6_transforming_protein
FT                    /note= "E6 ORF from 65 to 559; putative"
FT   mutation         350
FT                    /*tag=  d
FT                    /note= "Mutated from the normal T in the reference
FT                     sequence (T94723)"
FT   CDS              562..858
FT                    /*tag=  e
FT                    /product= E7_transforming_protein
FT                    /note= "E7 ORF 544 to 858; putative"
FT   CDS              865..2813
FT                    /*tag=  f
FT                    /product= E1_replication_protein
FT                    /note= "E1 interrupted ORF from 859 to 2813; putative"
FT   CDS              2755..3852
FT                    /*tag=  g
FT                    /product= E2_regulatory_protein
FT                    /note= "E2 ORF from 2725 to 3852; putative"
FT   CDS              3332..3619
FT                    /*tag=  h
FT                    /product= E4
FT                    /note= "E4 ORF from 3332 to 3619; putative"
FT   CDS              3863..4099
FT                    /*tag=  i
FT                    /product= E5
FT                    /note= "E5 ORF from 3863 to 4099; putative"
FT   polyA_signal     4213..4218
FT                    /*tag=  j
FT                    /note= "Putative"
FT   CDS              4235..5656
FT                    /*tag=  k
```



```
FT                    /product= L2_minor_capsid_protein
FT                    /note= "L2 ORF from 4133 to 5656; putative"
FT   TATA_signal      4289..4295
FT                    /*tag=  l
FT   CDS              5559..7154
FT                    /*tag=  m
FT                    /product= L1_major_capsid_protein
FT                    /note= "L1 ORF from 5526 to 7154; putative"
FT   polyA_signal     7260..7265
FT                    /*tag=  n
PN   US5679509-A.
PD   21-OCT-1997.
PF   28-SEP-1993; 127906.
PR   30-SEP-1994; US-316239.
PR   28-SEP-1993; US-127906.
PA   (UYNE-) UNIV NEW MEXICO STATE.
PI   Parmenter CA,  Wheeler CM;
DR   WPI; 97-525714/48.
DR   P-PSDB; W35742.
PT   Evaluating risk of cervical dysplasia or cervical cancer - by
PT   detecting variant form of human papilloma virus 16
PS   Claim 1; Column 15-24; 33pp; English.
CC   Methods have been developed for distinguishing a subset of human
CC   papilloma virus (HPV) that is associated with an increased risk of
CC   developing cervical dysplasia or cervical cancer. The methods involve:
CC   (1) preparing a cervical sample to expose any HPV-16 E6 gene in the
CC   sample and determining if the base at position 350 of the E6 gene
CC   (see T94723 and T94724 for comparison) is T or G, where the presence of
CC   G at position 350 is associated with an increased risk of developing
CC   cervical dysplasia or cervical cancer; and (2) preparing a cervical
CC   sample to expose any HPV-16 E6 protein in the sample and determining
CC   if the amino acid at position 83 of the protein (see position 90 in
CC   W35741 and W35742 for comparison) is Val or Leu, where the presence of
CC   Val at position 83 that is associated with an increased risk of
CC   developing cervical dysplasia or cervical cancer. The present sequence
CC   represents the variant nucleotide sequence for HPV-16 E6. The 350G
CC   variant correlates well with Pap scores: 350T:350G ratios among 45
CC   HPV16 samples were 10:4 for negative Pap scores; 4:2 for CIN I, 1:6 for
CC   CIN II; 2:9 for CIN III; 0:3 for cancer.
SQ   Sequence   7904 BP;    2601 A;    1377 C;    1510 G;    2416 T;

  Query Match           18.4%;  Score 16;  DB 35;  Length 7904;
  Best Local Similarity 100.0%;  Pred. No. 5.57e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    3280 gaaaaatatagtaaaa 3295
           ||||||||||||||||
Qy      33 GAAAAATATAGTAAAA 48


RESULT   20
```



```
ID   N70543 standard; DNA; 12492 BP.
AC   N70543;
DT   11-MAR-1991  (first entry)
DE   Complete sequence of the bovine diarrhoea virus genome.
KW   BDV; vaccine; ds.
OS   Bovine diarrhoea virus.
FH   Key             Location/Qualifiers
FT   cds              384..12303
FT                    /*tag=  a
PN   EP-208672-A.
PD   14-JAN-1987.
PF   01-JUL-1986; 870095.
PR   08-JUL-1985; US-752981.
PA   (REGI-) REGION WALLONNE.
PA   (CHIR-) CHIRON CORP.
PA   (WALL-) WALLONNE REGIONAL.
PI   Renard A, Dina D, Martial J;
DR   WPI; 87-009295/02.
DR   P-PSDB; P70340.
PT   Complete nucleotide sequence of bovine diarrhoea virus genome -
PT   and its fragments, useful for making antigenic proteins useful in
PT   therapy and diagnosis.
PS   Claim 1; Fig 2; 54pp; English.
CC   The gene product may be produced from recombinant expression systems,
CC   and used to raise Abs for both active and passive immunotherapy and
CC   diagnosis. DNA and RNA fragments may also be used as diagnostic probes.
CC   The preferred expression system has the N-terminus of beta-
CC   galactosidase upstream of the BDV sequence, and the particle
CC   forming polypeptide is derived from HBsAg.
SQ   Sequence   12492 BP;    3966 A;    2544 C;    3213 G;    2769 T;

  Query Match           18.4%;  Score 16;  DB 1;  Length 12492;
  Best Local Similarity 100.0%;  Pred. No. 5.57e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    1380 aagacaaattacacat 1395
           ||||||||||||||||
Cp      63 AAGACAAATTACACAT 48


RESULT   21
ID   T11901 standard; DNA; 280 BP.
AC   T11901;
DT   03-SEP-1996  (first entry)
DE   Mycobacterium ITG 8723 16S-23S rRNA spacer region.
KW   Probe; detection; identification; microorganism; amplify;
KW   16S-23S rRNA spacer region; food; universal;
KW   species-specific; ss.
OS   Mycobacterium sp.
PN   WO9600298-A1.
PD   04-JAN-1996.
```



```
PF   23-JUN-1995; E02452.
PR   24-JUN-1994; EP-870106.
PR   07-APR-1995; EP-870032.
PA   (INNO-) INNOGENETICS NV.
PI   Jannes G,  Rossau R,  Van Heuverswyn H;
DR   WPI; 96-068882/07.
PT   Novel hybridisation assay for the detection of eubacteria - esp
PT   Mycobacterium, comprises amplifying the 16S-23S rRNA spacer region
PS   Claim 9; Fig 34; 248pp; English.
CC   The sequences given in T11889-905 represent the 16S-23S rRNA spacer
CC   regions derived from various microorganisms.  These sequences were
CC   used in the method of the invention for the detection and identification
CC   of at least one or more microorganisms.  The method comprises amplifying
CC   the 16S-23S rRNA spacer region and hybridising one or more probes to
CC   the amplified sequence.  These probes were specifically used to detect
CC   microorganisms in samples originating from food.  This spacer region is
CC   derived from Mycobacterium ITG 8723.
SQ   Sequence   280 BP;    54 A;    55 C;    88 G;    68 T;

  Query Match           17.2%;  Score 15;  DB 21;  Length 280;
  Best Local Similarity 100.0%;  Pred. No. 2.05e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     246 taaaatgtgtaatttt 260
           ||||||||||||||||
Qy      44 TAAAATGTGTAATTT 58


RESULT   22
ID   T22640 standard; cDNA to mRNA; 325 BP.
AC   T22640;
DT   15-AUG-1996  (first entry)
DE   Human gene signature HUMGS04294.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 1177; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
```

```
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   325 BP;    109 A;    55 C;    49 G;    112 T;

  Query Match           17.2%;  Score 15;  DB 20;  Length 325;
  Best Local Similarity 100.0%;  Pred. No. 2.05e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    288 tttgtcttaataaaa 302
          |||||||||||||||
Qy     56 TTTGTCTTAATAAAA 70


RESULT   23
ID   T20519 standard; cDNA to mRNA; 404 BP.
AC   T20519;
DT   22-JUL-1996  (first entry)
DE   Human gene signature HUMGS01730.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 658; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
```



```
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   404 BP;    95 A;    95 C;    92 G;    122 T;

  Query Match           17.2%;  Score 15;  DB 19;  Length 404;
  Best Local Similarity 100.0%;  Pred. No. 2.05e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    379 gtcttaataaaatct 393
          |||||||||||||||
Qy     59 GTCTTAATAAAATCT 73


RESULT   24
ID   Q70916 standard; DNA; 2726 BP.
AC   Q70916;
DT   27-MAR-1995  (first entry)
DE   Multidrug-resistance gene MDR-1.
KW   Multidrug-resistance; MDR-1 gene; ds.
OS   Synthetic.
PN   WO9417210-A.
PD   04-AUG-1994.
PF   25-JAN-1994; U00748.
PR   27-JAN-1993; US-010433.
PR   16-DEC-1993; US-168621.
PA   (ONCO-) ONCOR INC.
PI   Bhatnagar SK,  George AL;
DR   WPI; 94-264122/32.
PT   Enzymatic amplification of target nucleic acid sequences to form
PT   exact or modified copies - has increased fidelity and can identify
PT   point mutations or allele(s)
PS   Disclosure; Page 30; 69pp; English.
CC   This gene is used as an example of a target DNA in a method for the
CC   amplification of nucleic acid sequences in a mixture using various
CC   DNA probes/primers.  The method may be used to detect a mutation or
CC   allele in the target, e.g. that that causes a genetic disease.
SQ   Sequence   2726 BP;    818 A;    538 C;    642 G;    728 T;

  Query Match           17.2%;  Score 15;  DB 12;  Length 2726;
  Best Local Similarity 100.0%;  Pred. No. 2.05e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```



```
Db   2595 taaaatgtgtaattt 2609
          |||||||||||||||
Qy     44 TAAAATGTGTAATTT 58


RESULT   25
ID   T43322 standard; DNA; 2726 BP.
AC   T43322;
DT   01-SEP-1997  (first entry)
DE   Multidrug resistance gene-1.
KW   PCR; primer; amplify; polymerase chain reaction; ligase chain reaction;
KW   LCR; human; multidrug resistance gene; MDR-1; ds.
OS   Synthetic.
PN   WO9639537-A1.
PD   12-DEC-1996.
PF   04-JUN-1996; U08841.
PR   05-JUN-1995; US-461823.
PA   (ONCO-) ONCOR INC.
PI   Bhatnagar SK,  George AL,  Nazarenko I;
DR   WPI; 97-043158/04.
PT   Amplification method avoiding strand displacement by polymerase -
PT   used in the detection of mutation(s) and allele(s) associated with
PT   genetic disease and cancer
PS   Example 1; Page 50-51; 92pp; English.
CC   This sequence represents the multidrug resistance gene (MDR-1) amplified
CC   by the primers shown in T43320 and T43321. This sequence can be used as a
CC   target in the method of the invention, for enzymatically amplifying a
CC   target nucleic acid (TNA) sequence contained in a nucleic acid or mixture
CC   of nucleic acids while avoiding strand displacement by polymerase. The
CC   method comprises using three primers, one complementary to a first
CC   segment of the TNA, a second complementary to a second segment of the
CC   TNA, which is adjacent to the first primer, and a third which is similar
CC   to the first segment of the TNA, and is complementary to a portion of the
CC   first primer. The first two primers are hybridised to the TNA, and a
CC   fused amplification product is created from the TNA using these two
CC   primers. The fused product is dissociated, and hybridised to the third
CC   primer, which is then extended. The extended modified amplification
CC   product is hybridised to the first two primers, and these two primers are
CC   ligated. Each of the primers may be labelled using a different label so
CC   that the method can be used to detect the presence of a mutation or
CC   allele by detecting whether the labelled primer is contained within the
CC   fused amplification product or the extended amplification product. The
CC   method of the invention combines certain aspects of ligase chain reaction
CC   (LCR) and polymerase chain reaction (PCR), but is improved compared to
CC   LCR due to the reduced number of primers needed, and the fact that the
CC   entire TNA sequence does not need to be known.
SQ   Sequence   2726 BP;    818 A;    538 C;    642 G;    728 T;

  Query Match           17.2%;  Score 15;  DB 31;  Length 2726;
  Best Local Similarity 100.0%;  Pred. No. 2.05e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```



```
Db   2595 taaaatgtgtaattt 2609
          |||||||||||||||
Qy     44 TAAAATGTGTAATTT 58


RESULT   26
ID   Q70907 standard; DNA; 2726 BP.
AC   Q70907;
DT   27-MAR-1995  (first entry)
DE   Multidrug-resistance gene MDR-1.
KW   Multidrug-resistance; MDR-1 gene; ds.
OS   Synthetic.
PN   WO9417206-A.
PD   04-AUG-1994.
PF   12-NOV-1993; U10883.
PR   27-JAN-1993; US-010433.
PR   16-DEC-1993; US-168621.
PA   (ONCO-) ONCOR INC.
PI   Bhatnagar SK,  George AL;
DR   WPI; 94-264118/32.
PT   Enzymatic amplification of target nucleic acid sequences in a
PT   mixt. - to detect a mutation or allele in the target, e.g. that
PT   causes a genetic disease
PS   Disclosure; Page 30; 50pp; English.
CC   This gene is used as an example of a target DNA in a method for the
CC   amplification of nucleic acid sequences in a mixture using various
CC   DNA probes/primers.  The method may be used to detect a mutation or
CC   allele in the target, e.g. that that causes a genetic disease.
SQ   Sequence   2726 BP;    818 A;    538 C;    642 G;    728 T;

  Query Match           17.2%;  Score 15;  DB 12;  Length 2726;
  Best Local Similarity 100.0%;  Pred. No. 2.05e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2595 taaaatgtgtaattt 2609
          |||||||||||||||
Qy     44 TAAAATGTGTAATTT 58


RESULT   27
ID   Q04522 standard; DNA; 4378 BP.
AC   Q04522;
DT   01-OCT-1990  (first entry)
DE   Multidrug Resistance A1 gene
KW   Multidrug Resistance A1 gene; drug resistance of cancer cells; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             138..3981
FT                   /*tag= a
PN   J02100680-A.
```

```
PD   12-APR-1990.
PF   05-OCT-1988; 251475.
PR   05-OCT-1988; JO-251475.
PA   (SUNR) Suntory Ltd.
DR   WPI; 90-159707/21.
DR   P-PSDB; R04868.
PT   MDR related gene derived from human normal cells -
PT   codes specific amino acid sequence, used for diagnosing drug
PT   resistance of cancer cells
PS   Disclosure; p; Japanese.
CC   The gene is useful for diagnosis of drug resistance of cancer
CC   cells.
SQ   Sequence   4378 BP;   1313 A;   819 C;   1058 G;   1188 T;

  Query Match           17.2%;  Score 15;  DB 1;  Length 4378;
  Best Local Similarity 100.0%;  Pred. No. 2.05e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   4228 taaaatgtgtaattt 4242
         |||||||||||||||
Qy     44 TAAAATGTGTAATTT 58


RESULT   28
ID   Q72872 standard; cDNA; 4646 BP.
AC   Q72872;
DT   21-JUN-1995  (first entry)
DE   Human multidrug resistance gene (MRD-1) encoding a p-glycoprotein.
KW   Human; tyrosinase; compound; hair; follicle; liposome; lipophilic;
KW   lipophobic; stratum corneum; melanin; hair dye; glycoprotein; alopecia;
KW   chemotherapy; transforming growth factor; growth stimulant; aromatase;
KW   cyclosporin A; macromolecule; polymer; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             425..4267
FT                   /*tag= a
FT                   /product= p-glycoprotein
PN   WO9422468-A.
PD   13-OCT-1994.
PF   01-APR-1994; U03634.
PR   02-APR-1993; US-041553.
PR   13-JAN-1994; US-181471.
PA   (ANTI-) ANTICANCER INC.
PI   Li L,  Lishko VK;
DR   WPI; 94-332816/41.
DR   P-PSDB; Q72872.
PT   Lipsomes for delivering protein, nucleic acid etc. to hair
PT   follicles - e.g. to restore hair colour, prevent hair loss during
PT   chemotherapy, stimulate hair growth etc.
PS   Claim 8; Page 70-77; 100pp; English.
CC   The nucleotide sequence of the human multidrug resistance (MDR-1) gene
```



```
CC   which encodes a p-glycoprotein.  This is an example of a compound which
CC   can be delivered to hair follicles via a novel liposome composition.  The
CC   liposomes are comprised of a lipophilic or lipophobic compound which will
CC   selectively target the hair follicle (via the stratum corneum) without
CC   damaging or unwanted effects on cells outside the follicle.  Compounds
CC   e.g. p-glycoprotein, can be delivered to the hair follicle for treatment
CC   of chemotherapy-induced alopecia.  Other compounds targeted at hair
CC   follicles can include: tyrosinase (or the DNA encoding it - Q72871),
CC   melanin or hair dyes (to restore hair colour or condition); human
CC   transforming growth factor-alpha (Q72873)(for reversal of wavy hair);
CC   hair growth stimulants (e.g. cyclosporin A or aromatase) or antisense
CC   sequences.  The method allows compounds (e.g. macromolecules or
CC   polymers), which would not normally reach the hair follicles, to be
CC   delivered to these target areas.
SQ   Sequence   4646 BP;   1371 A;   893 C;   1128 G;   1254 T;

  Query Match           17.2%;  Score 15;  DB 12;  Length 4646;
  Best Local Similarity 100.0%;  Pred. No. 2.05e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   4515 taaaatgtgtaattt 4529
         |||||||||||||||
Qy     44 TAAAATGTGTAATTT 58


RESULT   29
ID   N70752 standard; cDNA; 4669 BP.
AC   N70752;
DT   21-MAY-1991  (first entry)
DE   Sequence of human multi-drug resistance-1 (mdr1) cDNA from clones
DE   lambda-HDR10,5 and 104.
KW   Chemo-therapy resistant tumour cell; P-glycoprotein; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             425..4267
FT                   /*tag= a
PN   WO8705943-A.
PD   08-OCT-1987.
PF   26-MAR-1987; U00758.
PR   28-MAR-1986; US-845610.
PR   01-AUG-1986; US-892575.
PA   (UNII ) UNIV OF ILLINOIS.
PI   Ronimson IB, Pastan IH, Gottesman MM;
DR   WPI; 87-291656/41.
DR   P-PSDB; P70452.
PT   DNA for multi-drug resistance in human cells - used to detect
PT   chemotherapy-resistant tumour cells and for producing
PT   polypeptide(s) for diagnosis and therapy
PS   Claim 2(a); Table 5, pp30-39; 61pp; English.
CC   The human multi-drug resistant KB carcinoma cell lines were used as
CC   the source of the mdr1 gene nucleic acid sequences (N70751). To
```



```
CC   obtain cDNA clones of the mdr1 gene (N70752), poly (A) and RNA was
CC   used. Analysis of the AA sequence presented in (N70752) indicates
CC   that the mdr1 gene product is likely to be a transmembrane protein.
CC   The presence of transmembrane domains and potential glycosylation
CC   sites is consistent with the mdr1 protein being related to the
CC   P-glycoprotein.
SQ   Sequence   4669 BP;   1393 A;   896 C;   1126 G;   1254 T;

  Query Match           17.2%;  Score 15;  DB 2;  Length 4669;
  Best Local Similarity 100.0%;  Pred. No. 2.05e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   4515 taaaatgtgtaattt 4529
         |||||||||||||||
Qy     44 TAAAATGTGTAATTT 58


RESULT   30
ID   Q52726 standard; cDNA; 4669 BP.
AC   Q52726;
DT   24-JUN-1994  (first entry)
DE   Sequence of human multi-drug resistant gene mdr1.
KW   p-glycoprotein; multi-drug resistant gene; mdr1; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             424..4267
FT                   /*tag= a
PN   WO9324613-A.
PD   09-DEC-1993.
PF   14-MAY-1993; U04707.
PR   22-MAY-1992; US-887712.
PA   (GENE-) GENETIC THERAPY INC.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI   Mcdonagh KT,  Nienhuis A,  Tolstoshev P;
DR   WPI; 93-405805/50.
DR   P-PSDB; R44297.
PT   DNA or RNA sequence for human multi-drug resistant gene MDR1 -
PT   encodes P-glycoprotein with altered splice site, used in cancer
PT   therapy
PS   Example; Fig 4; 64pp; English.
CC   pMDR2000 contains an mdr1 cDNA sequence (Q52726) described in PCT
CC   application no. WO87/0943m wherein the first 282 bp of the 5' UTR
CC   and the last 23 bp of the 3' UTR of the cDNA sequence have been
CC   removed. The codon YTN is unclear in the spec., and encodes the AA
CC   Leu.
SQ   Sequence   4669 BP;   1394 A;   893 C;   1126 G;   1254 T;

  Query Match           17.2%;  Score 15;  DB 9;  Length 4669;
  Best Local Similarity 100.0%;  Pred. No. 2.05e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```



```
Db   4515 taaaatgtgtaattt 4529
         |||||||||||||||
Qy     44 TAAAATGTGTAATTT 58


RESULT   31
ID   T43225 standard; DNA; 9048 BP.
AC   T43225;
DT   26-FEB-1997  (first entry)
DE   Brassica napus FCA gene.
KW   FCA; flowering; transgenic plant; oilseed rape; ss.
OS   Brassica napus.
FH   Key             Location/Qualifiers
FT   cds             2468..2470
FT                   /*tag= a
FT                   /codon_start= 2468..2470
FT                   /note= "translation start codon"
PN   WO9638560-A2.
PD   05-DEC-1996.
PF   03-JUN-1996; G01332.
PR   02-JUN-1995; GB-011196.
PA   (INNE-) INNES CENT INNOVATIONS LTD JOHN.
PI   Bancroft I,  Dean C,  Lister CK,  MacKnight RC;
DR   WPI; 97-034373/03.
DR   P-PSDB; W06449.
PT   Methods of influencing flowering characteristics of plants - by
PT   administration of FCA protein, DNA or antisense transcripts
PS   Claim 20; Fig 8a; 97pp; English.
CC   The FCA gene (T43225) of Brassica napus codes for a polypeptide
CC   (W06449) able to influence flowering characteristics, partic.
CC   flowering time.  It was isolated from a genomic library using
CC   a cDNA clone obtd. from the Arabidopsis FCA gene (T43224).  The
CC   gene fully complements a mutation in the Arabidopsis FCA gene
CC   and is thus a fully functional homologue.  Timing of flowering of
CC   transgenic plants can be delayed or hastened using FCA sense and
CC   antisense constructs, mutants and alleles.  FCA genes can also be
CC   used to isolate FCA homologues from other plant species.
SQ   Sequence   9048 BP;   2643 A;   1643 C;   1713 G;   3049 T;

  Query Match           17.2%;  Score 15;  DB 25;  Length 9048;
  Best Local Similarity 100.0%;  Pred. No. 2.05e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   5669 ttacacattttacta 5683
         |||||||||||||||
Cp     55 TTACACATTTTACTA 41


RESULT   32
ID   T43224 standard; DNA; 9763 BP.
AC   T43224;
```

```
DT   26-FEB-1997  (first entry)
DE   Arabidopsis FCA gene.
KW   FCA; flowering; transgenic plant; ss.
OS   Arabidopsis thaliana.
FH   Key             Location/Qualifiers
FT   misc_rna        1118
FT                   /*tag= a
FT                   /function= transcription start site
FT   exon            1140..1865
FT                   /*tag= b
FT                   /codon_start= 1532..1534
FT                   /note= "alternative start sites are at positions
FT                   1568..1570 and 1601..1603"
FT   intron          1866..2120
FT                   /*tag= c
FT   exon            2121..2191
FT                   /*tag= d
FT   intron          2192..2278
FT                   /*tag= e
FT   exon            2279..2357
FT                   /*tag= f
FT   intron          2358..4432
FT                   /*tag= g
FT   polya_site      2753
FT                   /*tag= h
FT                   /note= "polyA site of beta-transcript"
FT   misc_rna        3026..3028
FT                   /*tag= i
FT                   /note= "stop codon potentially used in
FT                   FCA variant alpha-beta"
FT   exon            4433..4515
FT                   /*tag= j
FT   intron          4516..4595
FT                   /*tag= k
FT   exon            4596..4647
FT                   /*tag= l
FT   intron          4648..4844
FT                   /*tag= m
FT   exon            4845..4912
FT                   /*tag= n
FT   intron          4913..5272
FT                   /*tag= o
FT   exon            5273..5339
FT                   /*tag= p
FT   intron          5340..5426
FT                   /*tag= q
FT   exon            5427..5507
FT                   /*tag= r
FT   intron          5508..5594
FT                   /*tag= s
FT   exon            5595..5659
```



```
FT                   /*tag= t
FT   intron          5660..5865
FT                   /*tag= u
FT   exon            5866..5925
FT                   /*tag= v
FT   intron          5926..6013
FT                   /*tag= w
FT   exon            6014..6150
FT                   /*tag= x
FT   intron          6151..6493
FT                   /*tag= y
FT   exon            6494..6577
FT                   /*tag= z
FT   intron          6578..6662
FT                   /*tag= aa
FT   exon            6663..7130
FT                   /*tag= ab
FT   misc_rna        7056..7377
FT                   /*tag= ac
FT                   /note= "alternative splicing around intron 11"
FT   intron          7131..7295
FT                   /*tag= ad
FT   exon            7296..7475
FT                   /*tag= ae
FT   intron          7476..7629
FT                   /*tag= af
FT   exon            7630..7809
FT                   /*tag= ag
FT   intron          7810..8119
FT                   /*tag= ah
FT   exon            8120..8152
FT                   /*tag= ai
FT   intron          8153..8321
FT                   /*tag= aj
FT   exon            8322..8354
FT                   /*tag= ak
FT   intron          8355..8537
FT                   /*tag= al
FT   exon            8538..8603
FT                   /*tag= am
FT   intron          8604..8735
FT                   /*tag= an
FT   exon            8736..8781
FT                   /*tag= ao
FT   misc_rna        8771..8773
FT                   /*tag= ap
FT                   /function= translation stop TAA
FT   intron          8782..8898
FT                   /*tag= aq
FT   exon            8899..9256
FT                   /*tag= ar
```



```
FT   polya_site      9256
FT                   /*tag= as
PN   WO9638560-A2.
PD   05-DEC-1996.
PF   03-JUN-1996; G01332.
PR   02-JUN-1995; GB-011196.
PA   (INNE-) INNES CENT INNOVATIONS LTD JOHN.
PI   Bancroft I,  Dean C,  Lister CK,  MacKnight RC;
DR   WPI; 97-034373/03.
DR   P-PSDB; W06448.
PT   Methods of influencing flowering characteristics of plants - by
PT   administration of FCA protein, DNA or antisense transcripts
PS   Claim 2; Fig 1; 97pp; English.
CC   The FCA gene (T43224) of Arabidopsis thaliana codes for a
CC   polypeptide (W06448) able to influence flowering characteristics,
CC   partic. flowering time.  It was identified by fca mutant
CC   complementation analysis using cosmid clones.  The FCA transcript
CC   is alternatively spliced; 5 forms of the transcript (see also
CC   T43226) are generated in cells.  The timing of flowering of
CC   transgenic plants can be delayed or hastened using FCA sense and
CC   antisense constructs, mutants and alleles.  The Arabidopsis FCA
CC   gene can also be used to isolate FCA homologues, e.g. from Brassica
CC   napus (see also T43225).
SQ   Sequence   9763 BP;    2623 A;    1834 C;    1898 G;    3408 T;

  Query Match          17.2%;  Score 15;  DB 25;  Length 9763;
  Best Local Similarity 100.0%;  Pred. No. 2.05e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   8100 ttttcctttctaaaa 8114
          |||||||||||||||
Cp     37 TTTTCCTTTCTAAAA 23


RESULT   33
ID   T28774 standard; cDNA; 14042 BP.
AC   T28774;
DT   31-JUL-1996  (first entry)
DE   Human placental calcium sensor protein cDNA.
KW   Calcium sensor protein; CSP; hyperthyroidism; osteoporosis;
KW   membranous glomerulonephritis; therapy; antisense; placenta; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             68..14035
FT                   /*tag= a
PN   WO9615801-A1.
PD   30-MAY-1996.
PF   22-NOV-1995; U15203.
PR   23-NOV-1994; US-344836.
PR   07-JUN-1995; US-487314.
PA   (RHON ) RHONE POULENC RORER PHARM INC.
```



```
PI   Amerstrom G,  Crumley GR,  Hjalm G,  Juhlin C,  Morse CC;
PI   Murray EM,  Rask L;
DR   WPI; 96-268334/27.
DR   P-PSDB; R97209.
PT   Isolated human calcium sensor protein - useful develop prods. for
PT   diagnosis and treatment of conditions, such as membranous
PT   glomerulonephritis
PS   Claim 2; Page 110-128; 236pp; English.
CC   Full-length human placental (T28774), kidney (T28775) and
CC   parathyroid (T28776) cDNA clones coding for 500 kDa calcium sensor
CC   proteins (R97209-11, respetively) were obtd. from corresponding
CC   human cDNA libraries by RT-PCR (see also T28780-95).  These sequences
CC   can be used for prodn. of recombinant CSP, a protein involved in
CC   calcium homeostasis.  Transformed host cells expressing CSP are
CC   used to identify CSP (ant)agonists and in diagnostic and therapeutic
CC   applns.  Antisense sequences capable of down-regulating or blocking
CC   CSP expression can also be produced.
SQ   Sequence   14042 BP;    3772 A;    3091 C;    3503 G;    3669 T;

  Query Match          17.2%;  Score 15;  DB 20;  Length 14042;
  Best Local Similarity 100.0%;  Pred. No. 2.05e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1501 taataaaatctatct 1515
          |||||||||||||||
Qy     63 TAATAAAATCTATCT 77


RESULT   34
ID   V05995 standard; cDNA; 14042 BP.
AC   V05995;
DT   08-JUN-1998  (first entry)
DE   Human placenta calcium sensor protein full-length cDNA clone.
KW   Calcium sensor; human; glycoprotein 330; gp330; placenta;
KW   low density lipoprotein; LDL; membraneous glomerulonephritis;
KW   Alzheimer's disease; gene therapy; hyperthyroidism; osteoporosis;
KW   ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             68..14035
FT                   /*tag= a
FT   sig_peptide     68..142
FT                   /*tag= b
FT   mat_peptide     143..14032
FT                   /*tag= c
PN   WO9744050-A1.
PD   27-NOV-1997.
PF   20-MAY-1997; U08682.
PR   23-MAY-1996; US-652877.
PA   (RHON ) RHONE-POULENC RORER PHARM INC.
PI   Akerstrom G, Crumley GR, Hjalm G, Juhlin C, Morse CC,
```

```
PI   Murray EM, Rask L;
DR   WPI; 98-018219/02.
DR   P-PSDB; W43312.
PT   New human calcium sensor protein and vector encoding inhibitor of gp
PT   330 mediated transport of amyloid beta protein - useful to treat
PT   membraneous glomerulonephritis and central nervous system disorders
PS   Claim 2; Page 112-130; 240pp; English.
CC   This is a full-length sequence coding for human placenta calcium
CC   sensor protein (hCSP) (see W43312).  It was obtained by PCR
CC   amplification using primers (see V06001-16) derived from another
CC   hCSP cDNA sequence (see V05994).  Full-length sequences for human
CC   kidney (see V05996) and parathyroid (see V05997) hCSP were
CC   similarly obtained.  hCSP is considered to be the human homologue
CC   of rat Heymann nephritis antigen gp330 on the basis of sequence
CC   similarity.  It is proposed that the calcium sensor is a universal
CC   sensor for recognition of variation in extracellular calcium, and
CC   plays a key role in calcium regulation via different organ systems.
CC   Vectors comprising hCSP cDNA and host cells expressing the cDNA are
CC   claimed.  Also claimed are methods: of using hCSP to treat human
CC   membraneous glomerulonephritis; of using hCSP to screen for
CC   agonists and antagonists of hCSP (for use e.g. in the treatment of
CC   osteoporosis or hyperthyroidism); of using hCSP antisense nucleic
CC   acids to down-regulate or block expression of hCSP; and of using
CC   human gp330 antisense RNA to treat a central nervous system
CC   disorder characterised by beta-amyloid protein accumulation,
CC   especially Alzheimer's disease.
SQ   Sequence   14042 BP;    3772 A;    3090 C;    3504 G;    3669 T;

  Query Match          17.2%;  Score 15;  DB 39;  Length 14042;
  Best Local Similarity 100.0%;  Pred. No. 2.05e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1501 taataaaatctatct 1515
           |||||||||||||||
Qy     63 TAATAAAATCTATCT 77


RESULT   35
ID   V05997 standard; cDNA; 14044 BP.
AC   V05997;
DT   08-JUN-1998  (first entry)
DE   Human parathyroid calcium sensor protein full-length cDNA clone.
KW   Calcium sensor; human; glycoprotein 330; gp330; parathyroid;
KW   low density lipoprotein; LDL; membraneous glomerulonephritis;
KW   Alzheimer's disease; gene therapy; hyperthyroidism; osteoporosis;
KW   ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             65..14032
FT                   /*tag=  a
FT   sig_peptide     65..139
```



```
FT                   /*tag=  b
FT   mat_peptide     140..14029
FT                   /*tag=  c
PN   WO9744050-A1.
PD   27-NOV-1997.
PF   20-MAY-1997; U08682.
PR   23-MAY-1996; US-652877.
PA   (RHON ) RHONE-POULENC RORER PHARM INC.
PI   Akerstrom G, Crumley GR, Hjalm G, Juhlin C, Morse CC,
PI   Murray EM, Rask L;
DR   WPI; 98-018219/02.
DR   P-PSDB; W43314.
PT   New human calcium sensor protein and vector encoding inhibitor of gp
PT   330 mediated transport of amyloid beta protein - useful to treat
PT   membraneous glomerulonephritis and central nervous system disorders
PS   Claim 2; Page 177-195; 240pp; English.
CC   This is a full-length sequence coding for human parathyroid calcium
CC   sensor protein (hCSP) (see W43314).  It was obtained by PCR
CC   amplification using primers (see V06001-16) derived from another
CC   hCSP cDNA sequence (see V05994).  Full-length sequences for human
CC   placenta (see V05995) and kidney (see V05996) hCSP were similarly
CC   obtained.  hCSP is considered to be the human homologue of rat
CC   Heymann nephritis antigen gp330 on the basis of sequence
CC   similarity.  It is proposed that the calcium sensor is a universal
CC   sensor for recognition of variation in extracellular calcium, and
CC   plays a key role in calcium regulation via different organ systems.
CC   Vectors comprising hCSP cDNA and host cells expressing the cDNA are
CC   claimed.  Also claimed are methods: of using hCSP to treat human
CC   membraneous glomerulonephritis; of using hCSP to screen for
CC   agonists and antagonists of hCSP (for use e.g. in the treatment of
CC   osteoporosis or hyperthyroidism); of using hCSP antisense nucleic
CC   acids to down-regulate or block expression of hCSP; and of using
CC   human gp330 antisense RNA to treat a central nervous system
CC   disorder characterised by beta-amyloid protein accumulation,
CC   especially Alzheimer's disease.
SQ   Sequence   14044 BP;    3774 A;    3093 C;    3502 G;    3673 T;

  Query Match          17.2%;  Score 15;  DB 39;  Length 14044;
  Best Local Similarity 100.0%;  Pred. No. 2.05e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1498 taataaaatctatct 1512
           |||||||||||||||
Qy     63 TAATAAAATCTATCT 77


RESULT   36
ID   T28776 standard; cDNA; 14044 BP.
AC   T28776;
DT   31-JUL-1996  (first entry)
DE   Human parathyroid calcium sensor protein cDNA.
```



```
KW   Calcium sensor protein; CSP; hyperthyroidism; osteoporosis;
KW   membranous glomerulonephritis; therapy; antisense; parathyroid; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             65..14032
FT                   /*tag=  a
PN   WO9615801-A1.
PD   30-MAY-1996.
PF   22-NOV-1995; U15203.
PR   23-NOV-1994; US-344836.
PR   07-JUN-1995; US-487314.
PA   (RHON ) RHONE POULENC RORER PHARM INC.
PI   Amerstrom G,  Crumley GR,  Hjalm G,  Juhlin C,  Morse CC;
PI   Murray EM,  Rask L;
DR   WPI; 96-268334/27.
DR   P-PSDB; R97211.
PT   Isolated human calcium sensor protein - useful develop prods. for
PT   diagnosis and treatment of conditions, such as membranous
PT   glomerulonephritis
PS   Claim 2; Page 175-193; 236pp; English.
CC   Full-length human placental (T28774), kidney (T28775) and
CC   parathyroid (T28776) cDNA clones coding for 500 kDa calcium sensor
CC   proteins (R97209-11, respetively) were obtd. from corresponding
CC   human cDNA libraries by RT-PCR (see also T28780-95).  These sequences
CC   can be used for prodn. of recombinant CSP, a protein involved in
CC   calcium homeostasis. Transformed host cells expressing CSP are
CC   used to identify CSP (ant)agonists and in diagnostic and therapeutic
CC   applns.  Antisense sequences capable of down-regulating or blocking
CC   CSP expression can also be produced.
SQ   Sequence   14044 BP;    3774 A;    3092 C;    3503 G;    3673 T;

  Query Match          17.2%;  Score 15;  DB 20;  Length 14044;
  Best Local Similarity 100.0%;  Pred. No. 2.05e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1498 taataaaatctatct 1512
           |||||||||||||||
Qy     63 TAATAAAATCTATCT 77


RESULT   37
ID   V05996 standard; cDNA; 14080 BP.
AC   V05996;
DT   08-JUN-1998  (first entry)
DE   Human kidney calcium sensor protein full-length cDNA clone.
KW   Calcium sensor; human; glycoprotein 330; gp330; kidney;
KW   low density lipoprotein; LDL; membraneous glomerulonephritis;
KW   Alzheimer's disease; gene therapy; hyperthyroidism; osteoporosis;
KW   ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
```



```
FT   CDS             105..14072
FT                   /*tag=  a
FT   sig_peptide     105..179
FT                   /*tag=  b
FT   mat_peptide     180..14069
FT                   /*tag=  c
PN   WO9744050-A1.
PD   27-NOV-1997.
PF   20-MAY-1997; U08682.
PR   23-MAY-1996; US-652877.
PA   (RHON ) RHONE-POULENC RORER PHARM INC.
PI   Akerstrom G, Crumley GR, Hjalm G, Juhlin C, Morse CC,
PI   Murray EM, Rask L;
DR   WPI; 98-018219/02.
DR   P-PSDB; W43313.
PT   New human calcium sensor protein and vector encoding inhibitor of gp
PT   330 mediated transport of amyloid beta protein - useful to treat
PT   membraneous glomerulonephritis and central nervous system disorders
PS   Claim 2; Page 145-163; 240pp; English.
CC   This is a full-length sequence coding for human kidney calcium
CC   sensor protein (hCSP) (see W43313).  It was obtained by PCR
CC   amplification using primers (see V06001-16) derived from another
CC   hCSP cDNA sequence (see V05994).  Full-length sequences for human
CC   placenta (see V05995) and parathyroid (see V05997) hCSP were
CC   similarly obtained.  hCSP is considered to be the human homologue
CC   of rat Heymann nephritis antigen gp330 on the basis of sequence
CC   similarity.  It is proposed that the calcium sensor is a universal
CC   sensor for recognition of variation in extracellular calcium, and
CC   plays a key role in calcium regulation via different organ systems.
CC   Vectors comprising hCSP cDNA and host cells expressing the cDNA are
CC   claimed.  Also claimed are methods: of using hCSP to treat human
CC   membraneous glomerulonephritis; of using hCSP to screen for
CC   agonists and antagonists of hCSP (for use e.g. in the treatment of
CC   osteoporosis or hyperthyroidism); of using hCSP antisense nucleic
CC   acids to down-regulate or block expression of hCSP; and of using
CC   human gp330 antisenae RNA to treat a central nervous system
CC   disorder characterised by beta-amyloid protein accumulation,
CC   especially Alzheimer's disease.
SQ   Sequence   14080 BP;    3784 A;    3100 C;    3518 G;    3677 T;

  Query Match          17.2%;  Score 15;  DB 39;  Length 14080;
  Best Local Similarity 100.0%;  Pred. No. 2.05e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1538 taataaaatctatct 1552
           |||||||||||||||
Qy     63 TAATAAAATCTATCT 77


RESULT   38
ID   T28775 standard; cDNA; 14080 BP.
```

```
AC   T28775;
DT   31-JUL-1996  (first entry)
DE   Human kidney calcium sensor protein cDNA.
KW   Calcium sensor protein; CSP; hyperthyroidism; osteoporosis;
KW   membranous glomerulonephritis; therapy; antisense; kidney; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             105..14072
FT                   /*tag= a
PN   WO9615801-A1.
PD   30-MAY-1996.
PF   22-NOV-1995; U15203.
PR   23-NOV-1994; US-344836.
PR   07-JUN-1995; US-487314.
PA   (RHON ) RHONE POULENC RORER PHARM INC.
PI   Amerstrom G,  Crumley GR,  Hjalm G,  Juhlin C,  Morse CC;
PI   Murray EM,  Rask L;
DR   WPI; 96-268334/27.
DR   P-PSDB; R97210.
PT   Isolated human calcium sensor protein - useful develop prods. for
PT   diagnosis and treatment of conditions, such as membranous
PT   glomerulonephritis
PS   Claim 2; Page 143-161; 236pp; English.
CC   Full-length human placental (T28774), kidney (T28775) and
CC   parathyroid (T28776) cDNA clones coding for 500 kDa calcium sensor
CC   proteins (R97209-11, respetively) were obtd. from corresponding
CC   human cDNA libraries by RT-PCR (see also T28780-95).  These sequences
CC   can be used for prodn. of recombinant CSP, a protein involved in
CC   calcium homeostasis.  Transformed host cells expressing CSP are
CC   used to identify CSP (ant)agonists and in diagnostic and therapeutic
CC   applns.  Antisense sequences capable of down-regulating or blocking
CC   CSP expression can also be produced.
SQ   Sequence   14080 BP;    3783 A;    3101 C;    3518 G;    3677 T;

  Query Match          17.2%;  Score 15;  DB 20;  Length 14080;
  Best Local Similarity 100.0%;  Pred. No. 2.05e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1538 taataaaatctatct 1552
          ||||||||||||||
Qy     63 TAATAAAATCTATCT 77


RESULT   39
ID   V05994 standard; cDNA; 14086 BP.
AC   V05994;
DT   08-JUN-1998  (first entry)
DE   Human calcium protein sensor cDNA clone.
KW   Calcium sensor; human; glycoprotein 330; gp330; placenta;
KW   low density lipoprotein; LDL; membraneous glomerulonephritis;
KW   Alzheimer's disease; gene therapy; hyperthyroidism; osteoporosis;
```



```
KW   ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             107..14074
FT                   /*tag= a
FT   sig_peptide     107..181
FT                   /*tag= b
FT   mat_peptide     182..14071
FT                   /*tag= c
PN   WO9744050-A1.
PD   27-NOV-1997.
PF   20-MAY-1997; U08682.
PR   23-MAY-1996; US-652877.
PA   (RHON ) RHONE-POULENC RORER PHARM INC.
PI   Akerstrom G, Crumley GR, Hjalm G, Juhlin C, Morse CC,
PI   Murray EM, Rask L;
DR   WPI; 98-018219/02.
DR   P-PSDB; W43311.
PT   New human calcium sensor protein and vector encoding inhibitor of gp
PT   330 mediated transport of amyloid beta protein - useful to treat
PT   membraneous glomerulonephritis and central nervous system disorders
PS   Claim 2; Page 80-98; 240pp; English.
CC   This is a full-length sequence coding for human calcium sensor
CC   protein (hCSP) (see W43311).  It is derived from cDNA clone CAS-2
CC   (see V05993), which was isolated from a human placenta cDNA
CC   library, and from a cDNA clone obtained by probe screening of
CC   a human kidney cDNA library.  PCR primers (see V06001-16) based
CC   on this sequence were used in the PCR amplification of full-length
CC   clones (see V05995-97) coding for placental, kidney and parathyroid
CC   hCSPs (see W43312-14).  hCSP is considered to be the human
CC   homologue of rat Heymann nephritis antigen gp330 on the basis of
CC   sequence similarity.  It is proposed that the calcium sensor is a
CC   universal sensor for recognition of variation in extracellular
CC   calcium, and plays a key role in calcium regulation via different
CC   organ systems.  Vectors comprising hCSP cDNA and host cells
CC   expressing the cDNA are claimed.  Also claimed are methods: of
CC   using hCSP to treat human membraneous glomerulonephritis; of using
CC   hCSP to screen for agonists and antagonists of hCSP (for use e.g.
CC   in the treatment of osteoporosis or hyperthryoidism); of using hCSP
CC   antisense nucleic acids to down-regulate or block expression of
CC   hCSP; and of using human gp330 antisense RNA to treat a central
CC   nervous system disorder characterised by beta-amyloid protein
CC   accumulation, especially Alzheimer's disease.
SQ   Sequence   14086 BP;    3788 A;    3099 C;    3520 G;    3679 T;

  Query Match          17.2%;  Score 15;  DB 39;  Length 14086;
  Best Local Similarity 100.0%;  Pred. No. 2.05e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1540 taataaaatctatct 1554
          ||||||||||||||
```



```
Qy     63 TAATAAAATCTATCT 77


RESULT   40
ID   T28773 standard; cDNA; 14086 BP.
AC   T28773;
DT   31-JUL-1996  (first entry)
DE   Human calcium sensor protein cDNA.
KW   Calcium sensor protein; CSP; hyperthyroidism; osteoporosis;
KW   membranous glomerulonephritis; therapy; antisense; kidney; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             107..14074
FT                   /*tag= a
FT   primer_bind     complement (3..21)
FT                   /*tag= b
FT                   /note= "primer P1s"
FT   primer_bind     complement (1056..1074)
FT                   /*tag= c
FT                   /note= "primer G20s"
FT   primer_bind     1294..1313
FT                   /*tag= d
FT                   /note= "primer G7as"
FT   primer_bind     complement (2111..2128)
FT                   /*tag= e
FT                   /note= "primer G26s"
FT   primer_bind     2522..2542
FT                   /*tag= f
FT                   /note= "primer G29as"
FT   primer_bind     complement (4340..4358)
FT                   /*tag= g
FT                   /note= "primer G16s"
FT   primer_bind     4520..4542
FT                   /*tag= h
FT                   /note= "primer G16as"
FT   primer_bind     complement (6159..6180)
FT                   /*tag= i
FT                   /note= "primer E4s"
FT   primer_bind     6398..6415
FT                   /*tag= j
FT                   /note= "primer E2as"
FT   primer_bind     complement (8572..8593)
FT                   /*tag= k
FT                   /note= "primer B5s"
FT   primer_bind     8689..8706
FT                   /*tag= l
FT                   /note= "primer B9as"
FT   primer_bind     complement (10720..10740)
FT                   /*tag= m
FT                   /note= "primer G19s"
FT   primer_bind     10898..10912
```



```
FT                   /*tag= n
FT                   /note= "primer 23.5"
FT   primer_bind     complement (12898..12919)
FT                   /*tag= o
FT                   /note= "primer G35s"
FT   primer_bind     13011..13031
FT                   /*tag= p
FT                   /note= "primer G36as"
PN   WO9615801-A1.
PD   30-MAY-1996.
PF   22-NOV-1995; U15203.
PR   23-NOV-1994; US-344836.
PR   07-JUN-1995; US-487314.
PA   (RHON ) RHONE POULENC RORER PHARM INC.
PI   Amerstrom G,  Crumley GR,  Hjalm G,  Juhlin C,  Morse CC;
PI   Murray EM,  Rask L;
DR   WPI; 96-268334/27.
DR   P-PSDB; R97208.
PT   Isolated human calcium sensor protein - useful develop prods. for
PT   diagnosis and treatment of conditions, such as membranous
PT   glomerulonephritis
PS   Claim 2; Page 78-96; 236pp; English.
CC   A cDNA clone (T28773) codes for a 500 kDa human calcium sensor
CC   protein (CSP = R97208).  It was obtd. by RT-PCR amplification and
CC   32P-labeled probe screening of human kidney cDNA libraries and by
CC   sequencing of placental cDNA clone pCAS-2 (see also T28772).  Other
CC   full-length CSP cDNA clones were also obtd. from the placenta, kidney
CC   and parathyroid (see also T28774-76).  These sequences can be used for
CC   prodn. of recombinant CSP.  Transformed host cells expressing CSP are
CC   used to identify CSP (ant)agonists and in diagnostic and therapeutic
CC   applns.  Antisense sequences capable of down-regulating or blocking
CC   CSP expression can also be produced.
SQ   Sequence   14086 BP;    3789 A;    3101 C;    3518 G;    3678 T;

  Query Match          17.2%;  Score 15;  DB 20;  Length 14086;
  Best Local Similarity 100.0%;  Pred. No. 2.05e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1540 taataaaatctatct 1554
          ||||||||||||||
Qy     63 TAATAAAATCTATCT 77


RESULT   41
ID   T04088 standard; DNA; 30 BP.
AC   T04088;
DT   16-MAY-1996  (first entry)
DE   Trypsin inhibitory protein cDNA antisense primer-5.
KW   Trypsin; inhibitor; human T98G cells; human placenta; pancreatitis;
KW   shock; multiple organ failure; disseminated intravascular coagulation;
KW   DIC; ss.
```

```
OS   Synthetic.
PN   J07242700-A.
PD   19-SEP-1995.
PF   04-MAR-1994; 059906.
PR   04-MAR-1994; JP-059906.
PA   (MOCH ) MOCHIDA PHARM CO LTD.
DR   WPI; 95-355285/46.
PT   Polypeptide having trypsin inhibitory activity - for the treatment
PT   of pancreatitis, shock, multi-organ failure, etc.
PS   Example 9; Page 16; 25pp; Japanese.
CC   A polypeptide having trypsin inhibitory activity, mol.wt. 22-28 kD
CC   or 17-23 kD (as determined by SDS-PAGE under reducing or non-
CC   reducing conditions, respectively) and an N-terminal sequence as in
CC   R79913 is claimed. The polypeptide was isolated from human T98G
CC   cells and is useful for treating pancreatitis, shock, multiple
CC   organ failure and disseminated intravascular coagulation. The
CC   present sequence is that of an antisense RACE-PCR primer used for
CC   cloning coding sequences for the full-length polypeptide into
CC   plasmid pUC118.
SQ   Sequence   30 BP;    8 A;    4 C;    7 G;    11 T;

  Query Match           16.1%;  Score 14;  DB 17;  Length 30;
  Best Local Similarity  93.3%;  Pred. No. 7.26e+01;
  Matches   14;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db      15 gtacagatagatttt 29
           ||| ||||||||||
Cp      81 GTANAGATAGATTTT 67


RESULT   42
ID   T35614 standard; RNA; 39 BP.
AC   T35614;
DT   03-SEP-1996  (first entry)
DE   M.leprae pre-rRNA probe #6.
KW   Probe; Mycobacterium; M.tuberculosis; M.bovis; M.avium; M.lufu; M.habana;
KW   M.intracellulare; M.leprae; M.simiae; M.paratuberculosis; M.marinum;
KW   pre-rRNA; lysis; antimicrobial agent; rifampicin; antibiotic-resistance;
KW   ss.
OS   Synthetic.
PN   WO9534574-A1.
PD   21-DEC-1995.
PF   16-JUN-1995; U07755.
PR   16-JUN-1994; US-261068.
PA   (MICR-) MICROPROBE CORP.
PI   Britischgi TB,  Cangelosi GA;
DR   WPI; 96-097380/10.
PT   Determn. of the sensitivity of mycobacterial cells to antimicrobial
PT   agents  - using oligo:nucleotide probes specific for rRNA precursors
PS   Claim 17; Page 60; 79pp; English.
CC   T35571-T35625 represent probes for mycobacterial pre-rRNA.  T35609-T35616
```



```
CC   are probes specific for M.leprae pre-rRNA.  These sequences are used in a
CC   method for detecting pre-rRNA in mycobacterial cells.  In this method,
CC   the cells are treated enzymatically or mechanically to expose the cell
CC   membrane.  The cell is then treated with lysis reagents to release (but
CC   not degrade) the pre-rRNA.  The pre-rRNA is then detected using at least
CC   one of these probe sequences.  The method can be used to determine the
CC   sensitivity of mycobacterial cells to antimicrobial agents (such as
CC   rifampicin), and thereby detect antibiotic-resistant cells.  The method
CC   is also useful to determine the antimicrobial efficacy of a compound.
SQ   Sequence   39 BP;    8 A;    3 C;    7 G;    21 U;

  Query Match           16.1%;  Score 14;  DB 21;  Length 39;
  Best Local Similarity  57.1%;  Pred. No. 7.26e+01;
  Matches    8;  Conservative    6;  Mismatches   0;  Indels   0;  Gaps   0;

Db      9 aaaauguguaauuu 22
          ||||:|:|:||:::
Qy     45 AAAATGTGTAATTT 58


RESULT   43
ID   V12977 standard; rRNA; 39 BP.
AC   V12977;
DT   18-MAY-1998  (first entry)
DE   Mycobacterium leprae pre-rRNA target sequence SEQ ID NO:37.
KW   Mycobacterium; pre-rRNA; precursor ribosomal RNA; target sequence;
KW   probe; hybridisation; antibiotic resistance; cell lysis; ss.
OS   Mycobacterium leprae.
FH   Key             Location/Qualifiers
FT   precursor_RNA   1..39
FT                   /*tag= a
FT                   /note= "pre-rRNA open region"
PN   US5712095-A.
PD   27-JAN-1998.
PF   07-JUN-1995; 485602.
PR   07-JUN-1995; US-485602.
PR   16-JUN-1994; US-261068.
PA   (BECT ) BECTON DICKINSON CO.
PI   Britschgi TB, Cangelosi GA;
DR   WPI; 98-119975/11.
PT   Assay for pre-ribosomal RNA in mycobacterium cells - by
PT   hybridisation with specific probes after cell lysis
PS   Claim 11; Column 87; 54pp; English.
CC   The present sequence represents a mycobacterial pre-ribosomal RNA
CC   (pre-rRNA) target sequence. The present invention describes a method for
CC   detecting pre-rRNA in cells of a mycobacterial sample. The method
CC   comprises: (a) treating the cells to release pre-rRNA by: (i)
CC   pretreating the cells by enzymatic degradation using both lysozyme and
CC   protease until their cell walls are rendered porous to expose their cell
CC   membranes, making the cells susceptible to lysis; (ii) contacting the
CC   pretreated cells with a combination of a magnesium chelator, a nonionic
```



```
CC   detergent and an anionic detergent; and (iii) heating the cells to 75-99
CC   degrees Celsius until the mycobacterial cells are lysed; and (b)
CC   detecting the pre-rRNA using at least one oligonucleotide probe which is
CC   capable of selectively hybridising to a region of the pre-rRNA that is
CC   not present in a mature mycobacterial rRNA. The probes can be used to
CC   identify many Mycobacterium spp. including M. tuberculosis, M. leprae,
CC   M.  habana, M. avium, M. bovis, M. lufu, M. paratuberculosis, M.
CC   marinum, M. simiae and/or M. intracellulare. The probes may also be used
CC   to measure mycobacterial response to inhibitors of RNA and protein
CC   synthesis and may therefore be used to screen new antimycobacterial
CC   drugs. Mycobacteria have a slow growth rate. By using the probes, drug
CC   developers can now identify compounds that are more effective, but less
CC   stable than those previously identified.
SQ   Sequence   39 BP;    8 A;    3 C;    7 G;    21 U;

  Query Match           16.1%;  Score 14;  DB 39;  Length 39;
  Best Local Similarity  57.1%;  Pred. No. 7.26e+01;
  Matches    8;  Conservative    6;  Mismatches   0;  Indels   0;  Gaps   0;

Db      9 aaaauguguaauuu 22
          ||||:|:|:||:::
Qy     45 AAAATGTGTAATTT 58


RESULT   44
ID   Q58024 standard; DNA; 43 BP.
AC   Q58024;
DT   14-SEP-1994  (first entry)
DE   Sequence of PCR primer used for the production of altered
DE   cellobiohydrolase 1 (cbh1) promoter.
KW   Promoter; cellobiohydrolase 1; cbh1; PCR primer; ss.
OS   Synthetic.
PN   WO9404673-A.
PD   03-MAR-1994.
PF   19-AUG-1993; FI0330.
PR   19-AUG-1992; US-932485.
PA   (ALKO-) ALKO OY AB.
PI   Ilmen MH,  Nakari TH,  Nevalainen KMH,  Onnela M,  Pentitilae ME;
DR   WPI; 94-083192/10.
PT   Cloning promoters active in a desired environmental condition -
PT   used partic. for expression of genes in Trichoderma fungal hosts
PT   in glucose-contg. medium
PS   Example; page 38; 120pp; English.
CC   pMLO16del0(2) was used as a PCR template to yield a 280 bp fragment
CC   E (primers = Q58023 and Q58024), beginning at the promoter
CC   internal polylinker and ending at -720, and a 720 bp fragment F
CC   (primers = Q58026 and Q58028), beginning at -720 and ending at
CC   KspI at -16. Fragments D and E were purified from agarose gel
CC   and digested with BstEII-XbaI and XbaI-KspI respectively,& ligated
CC   to the 7.8 kb BstEII-KspI fragment of pMI-25 to produce produce
CC   pMI-27. The cbh1 promoter of pMI-27 has sequence alterations at
```



```
CC   posns. -1505-1500 (genomic sequence GTGGGG, altered sequence
CC   TCTAAA)and at posn. -720 to -715 (genomic sequence GTGGGG, altered
CC   sequence TCTAGA) upstream of the translation initiation codon of intact
CC   cbh1  promoter. The sequence of the altred cbh1 promoter in pMI-27
CC   is provided in Q58017.
SQ   Sequence   43 BP;    9 A;    10 C;    10 G;    14 T;

  Query Match           16.1%;  Score 14;  DB 10;  Length 43;
  Best Local Similarity 100.0%;  Pred. No. 7.26e+01;
  Matches   14;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     19 aaatgtgtaatttg 32
          ||||||||||||||
Qy     46 AAATGTGTAATTTG 59


RESULT   45
ID   T23965 standard; cDNA to mRNA; 63 BP.
AC   T23965;
DT   27-AUG-1996  (first entry)
DE   Human gene signature HUMGS05919.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 1497; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
```

```
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   63 BP;    22 A;    6 C;    9 G;    26 T;

  Query Match          16.1%;  Score 14;  DB 20;  Length 63;
  Best Local Similarity 100.0%;  Pred. No. 7.26e+01;
  Matches   14;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     46 tgtcttaataaaat 59
          ||||||||||||||
Qy     58 TGTCTTAATAAAAT 71


RESULT   46
ID   Q58025 standard; DNA; 73 BP.
AC   Q58025;
DT   14-SEP-1994  (first entry)
DE   Sequence of PCR primer used to construct altered cellobiohydrolase 1
DE   (cbh1) promoter
KW   Promoter; cellobiohydrolase 1; cbh1; PCR; primer; ss.
OS   Synthetic.
PN   WO9404673-A.
PD   03-MAR-1994.
PF   19-AUG-1993; FI0330.
PR   19-AUG-1992; US-932485.
PA   (ALKO-) ALKO OY AB.
PI   Ilmen MH,  Nakari TH,  Nevalainen KMH,  Onnela M,  Pentitilae ME;
DR   WPI; 94-083192/10.
PT   Cloning promoters active in a desired environmental condition -
PT   used partic. for expression of genes in Trichoderma fungal hosts
PT   in glucose-contg. medium
PS   Example; Page 38; 120pp; English.
CC   pMLO16del0(2) was used as a PCR template to yield a 280 bp fragment
CC   G (primers = Q58023 and Q58025), beginning at the promoter
CC   internal polylinker and ending at -720, and a 720 bp fragment H
CC   (primers = Q58026 and Q58028), beginning at -720 and ending at
CC   KspI at -16. Fragments G and H were purified from agarose gel
CC   and digested with BstEII-XbaI and XbaI-KspI respectively, & ligated
CC   to the 7.8 kb BstEII-KspI fragment of pMI-25 to produce produce
CC   pMI-28. The cbh1 promoter of pMI-28 has sequence alterations at
CC   posns. -1505-1500 (genomic sequence GTGGGG, altered sequence
CC   TCTAAA), -1001-996 (genomic sequence CTGGGG, altered sequence
CC   TCTAAA), and at posn. -720 to -715 (genomic sequence GTGGGG, altered
CC   sequence TCTAGA) upstream of the translation initiation codon of intact
CC   cbh1  promoter. The sequence of the altred cbh1 promoter in pMI-28
CC   is provided in Q58018.
SQ   Sequence   73 BP;    18 A;    16 C;    18 G;    21 T;

  Query Match          16.1%;  Score 14;  DB 10;  Length 73;
  Best Local Similarity 100.0%;  Pred. No. 7.26e+01;
  Matches   14;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```



```
Db     19 aaatgtgtaatttg 32
          ||||||||||||||
Qy     46 AAATGTGTAATTTG 59


RESULT   47
ID   T22257 standard; cDNA to mRNA; 190 BP.
AC   T22257;
DT   13-SEP-1996  (first entry)
DE   Human gene signature HUMGS03828.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 1084; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   190 BP;    58 A;    31 C;    18 G;    81 T;

  Query Match          16.1%;  Score 14;  DB 20;  Length 190;
  Best Local Similarity 100.0%;  Pred. No. 7.26e+01;
  Matches   14;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     16 tatttttccttttct 29
          ||||||||||||||
```



```
Cp     40 TATTTTTCCTTTCT 27


RESULT   48
ID   T26282 standard; cDNA to mRNA; 195 BP.
AC   T26282;
DT   10-OCT-1996  (first entry)
DE   Human gene signature HUMGS08521.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 2046; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   195 BP;    66 A;    22 C;    31 G;    75 T;

  Query Match          16.1%;  Score 14;  DB 21;  Length 195;
  Best Local Similarity 100.0%;  Pred. No. 7.26e+01;
  Matches   14;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     70 attttactatattt 83
          ||||||||||||||
Cp     49 ATTTTACTATATTT 36
```



```
RESULT   49
ID   T21383 standard; cDNA to mRNA; 227 BP.
AC   T21383;
DT   01-AUG-1996  (first entry)
DE   Human gene signature HUMGS02747.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 873; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   227 BP;    84 A;    27 C;    31 G;    77 T;

  Query Match          16.1%;  Score 14;  DB 19;  Length 227;
  Best Local Similarity 100.0%;  Pred. No. 7.26e+01;
  Matches   14;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     56 tttactatattttt 69
          ||||||||||||||
Cp     47 TTTACTATATTTTT 34


RESULT   50
ID   T26396 standard; cDNA to mRNA; 275 BP.
AC   T26396;
```

DT 06-DEC-1996 (first entry)
DE Human gene signature HUMGS08637.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN WO9514772-A1.
PD 01-JUN-1995.
PF 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATS/) MATSUBARA K.
PA (OKUB/) OKUBO K.
PI Matsubara K, OKubo K;
DR WPI; 95-206931/27.
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
PS Claim 1; Page 2075; 2245pp; Japanese.
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in T19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
SQ Sequence 275 BP; 84 A; 40 C; 58 G; 79 T;

Query Match 16.1%; Score 14; DB 23; Length 275;
Best Local Similarity 100.0%; Pred. No. 7.26e+01;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 197 aagacaaattacac 210
||||||||||||||
Cp 63 AAGACAAATTACAC 50

Search completed: Tue Dec 22 22:11:58 1998
Job time : 138 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Tue Dec 22 21:57:24 1998; MasPar time 166.25 Seconds
1166.533 Million cell updates/sec

Tabular output not generated.

Title: >US-08-530-112A-3084
Description: (1-88) from US08530112A.seq
Perfect Score: 87
N.A. Sequence: 1 GATCAATCTTTGATGGTGAG.....AATCTATCTNTACATCTAAA 88
Comp: CTAGTTAGAAACTACCACTC.....TTAGATAGANATGTAGATTT

Scoring table: TABLE jmetric
Gap 60

Nmatch STD : Dbase 0; Query 0

Searched: 567134 seqs, 1101898692 bases x 2

Post-processing: Minimum Match 0%
Listing first 1000 summaries

Database: embl55
1:em_ba 2:em_fun 3:em_htg 4:em_hum1 5:em_hum2 6:em_in
7:em_om 8:em_or 9:em_ov 10:em_pat 11:em_ph 12:em_pl
13:em_ro 14:em_vi
Database: genbank107
15:gb_ba1 16:gb_ba2 17:gb_htg 18:gb_in 19:gb_om 20:gb_ov
21:gb_pat 22:gb_ph 23:gb_pl1 24:gb_pl2 25:gb_pr1
26:gb_pr2 27:gb_pr3 28:gb_ro 29:gb_st 30:gb_sts 31:gb_sy
32:gb_un 33:gb_vi

Statistics: Mean 7.354; Variance 3.090; scale 2.380

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
c	1	76	87.4	307	30	G05576	human STS WI-8386.	7.13e-50
	2	21	24.1	120067	24	T25C13	Arabidopsis thaliana B	4.12e-02
	3	19	21.8	48690	26	AP000027	Homo sapiens genomic D	9.68e-01
c	4	19	21.8	121567	17	AC005727	*** SEQUENCING IN PROG	9.68e-01
c	5	19	21.8	124515	17	HS12803	Human DNA sequence ***	9.68e-01
c	6	19	21.8	180551	26	HUAC004158	Homo sapiens Chromosom	9.68e-01
	7	18	20.7	1903	28	MMU17092	Mus musculus ORRla tra	4.41e+00
c	8	18	20.7	2438	30	G06425	human STS WI-7161.	4.41e+00
c	9	18	20.7	10884	25	HUMIBP3	Human insulin-like gro	4.41e+00
	10	18	20.7	110079	28	AC002315	Mouse BAC-146N21 Chrom	4.41e+00
	11	18	20.7	194985	28	AC002406	Mouse chromosome X BAC	4.41e+00
c	12	18	20.7	199496	17	HS357K22	Human DNA sequence ***	4.41e+00
	13	17	19.5	453	28	AF054044	Rattus norvegicus poly	1.92e+01
c	14	17	19.5	1057	33	SLNPVPOLH	Spodotera litura nucle	1.92e+01
c	15	17	19.5	1667	21	E06147	Promoter of Candida al	1.92e+01
	16	17	19.5	1898	28	MMU17094	Mus musculus ORR1 tran	1.92e+01
c	17	17	19.5	2800	26	HSMYOC1	Homo sapiens myocilin	1.92e+01
	18	17	19.5	3535	15	SPU04047	Streptococcus pneumoni	1.92e+01
	19	17	19.5	4025	28	RNRRNA08	Rat rRNA spacer downst	1.92e+01
c	20	17	19.5	4133	16	AF030369	Streptococcus pneumoni	1.92e+01
c	21	17	19.5	4174	23	YSAAODIA	Candida boidinii metha	1.92e+01
c	22	17	19.5	4346	16	AF030372	Streptococcus pneumoni	1.92e+01
c	23	17	19.5	4981	16	AF030368	Streptococcus pneumoni	1.92e+01
c	24	17	19.5	4981	16	AF030370	Streptococcus pneumoni	1.92e+01
c	25	17	19.5	5300	27	AF007562	Homo sapiens trabecula	1.92e+01
c	26	17	19.5	5437	16	AF030371	Streptococcus pneumoni	1.92e+01
c	27	17	19.5	6798	16	AF030367	Streptococcus pneumoni	1.92e+01
c	28	17	19.5	14056	15	SPDEXCAP	S.pneumoniae dexB, cap	1.92e+01
c	29	17	19.5	14943	15	SPCPS14E	S.pneumoniae cps14 loc	1.92e+01
	30	17	19.5	16621	25	HSPAXIII6	Human paired box homeo	1.92e+01
c	31	17	19.5	17340	16	SPAJ6986	Streptococcus pneumoni	1.92e+01
c	32	17	19.5	23770	15	SPZ83335	S.pneumoniae dexB, cap	1.92e+01
	33	17	19.5	30589	18	CEZK666	Caenorhabditis elegans	1.92e+01
c	34	17	19.5	34657	18	CEK09G1	Caenorhabditis elegans	1.92e+01
	35	17	19.5	35800	27	AC005524	Homo sapiens chromosom	1.92e+01
	36	17	19.5	41155	18	CELF40A3	Caenorhabditis elegans	1.92e+01
	37	17	19.5	45852	26	HSAC002114	Human Cosmid g1940a085	1.92e+01
c	38	17	19.5	47188	26	AC004264	Homo sapiens PAC clone	1.92e+01
	39	17	19.5	70028	23	EPFCPCG	Epifagus virginiana ch	1.92e+01
c	40	17	19.5	70028	23	EPFCPCG	Epifagus virginiana ch	1.92e+01
c	41	17	19.5	78344	23	TM017A05	Arabidopsis thaliana B	1.92e+01
	42	17	19.5	79376	26	HS454G6	Human DNA sequence fro	1.92e+01

c	43	17	19.5	86554	23	ATF24J7	Arabidopsis thaliana D	1.92e+01
c	44	17	19.5	107428	26	HS481A17	Human DNA sequence fro	1.92e+01
c	45	17	19.5	108878	26	AC003092	Human BAC clone RG335O	1.92e+01
c	46	17	19.5	124191	26	AC005153	Homo sapiens PAC clone	1.92e+01
	47	17	19.5	130632	26	AC004547	Homo sapiens PAC clone	1.92e+01
c	48	17	19.5	132981	26	HS82J11	Human DNA sequence fro	1.92e+01
	49	17	19.5	136503	17	AC004866	*** SEQUENCING IN PROG	1.92e+01
c	50	17	19.5	145857	17	AC004033	*** SEQUENCING IN PROG	1.92e+01
c	51	17	19.5	157653	27	AC005146	Homo sapiens 12q24.2 P	1.92e+01
c	52	17	19.5	159479	17	AC005535	*** SEQUENCING IN PROG	1.92e+01
c	53	17	19.5	176433	26	HS71L16	Human DNA sequence fro	1.92e+01
	54	17	19.5	193161	27	AC005500	, complete sequence.	1.92e+01
c	55	17	19.5	210216	17	AC004888	*** SEQUENCING IN PROG	1.92e+01
c	56	16	18.4	69	21	I66508	Sequence 28 from paten	7.93e+01
c	57	16	18.4	142	21	A26063	annealed nucleotide I.	7.93e+01
	58	16	18.4	241	30	DM29G4T	D. melanogaster STS de	7.93e+01
	59	16	18.4	315	30	HSA106VD5	H.sapiens DNA segment	7.93e+01
c	60	16	18.4	320	30	HUMUT7981A	Human STS UT7981, 5' p	7.93e+01
c	61	16	18.4	390	28	MMT17854	Mus musculus mRNA for	7.93e+01
c	62	16	18.4	396	30	HUMUT5360	Human chromosome 8 STS	7.93e+01
	63	16	18.4	430	30	DM151H6S	Drosophila melanogaste	7.93e+01
c	64	16	18.4	478	23	MZEMTHA2	Maize mitochondrial Hi	7.93e+01
	65	16	18.4	620	26	AG000879	Homo sapiens genomic D	7.93e+01
	66	16	18.4	630	26	AG000877	Homo sapiens genomic D	7.93e+01
	67	16	18.4	632	26	AG000878	Homo sapiens genomic D	7.93e+01
	68	16	18.4	668	28	MUSHIS2AR	Mouse replication-depe	7.93e+01
c	69	16	18.4	849	28	MUSIAPFL1	Mus musculus cellular	7.93e+01
c	70	16	18.4	1014	25	HUMPS01	Human S protein-alpha	7.93e+01
c	71	16	18.4	1133	23	SCRFA3	S.cerevisiae gene for	7.93e+01
	72	16	18.4	1243	15	CLOETXDA	Clostridium perfringen	7.93e+01
c	73	16	18.4	1376	21	A22714	gE1 gene.	7.93e+01
	74	16	18.4	1500	19	RABNAP1	Rabbit neutrophil attr	7.93e+01
	75	16	18.4	1631	24	YSCKRE9X	Yeast O-glycoprotein (	7.93e+01
c	76	16	18.4	1839	15	BMU81371	Bacillus methanolicus	7.93e+01
c	77	16	18.4	1959	25	MNKHBD	Spider monkey (A.geoff	7.93e+01
c	78	16	18.4	1976	25	HSDGL1	Human gene for delta-g	7.93e+01
c	79	16	18.4	2149	25	ORAHBD	Orangutan delta globin	7.93e+01
c	80	16	18.4	2210	21	A22715	gE1/partial gE2 gene.	7.93e+01
c	81	16	18.4	2211	28	S75960	Tamm-Horsfall protein	7.93e+01
c	82	16	18.4	2227	28	RATUROMOD	Rat Tamm-Horsfall prot	7.93e+01
c	83	16	18.4	2373	33	SHVVP1CP	Simian hepatitis A vir	7.93e+01
	84	16	18.4	2400	16	LLJ002203	Lactococcus lactis lac	7.93e+01
c	85	16	18.4	2414	25	HSRNANLP	H.sapiens mRNA for nuc	7.93e+01
c	86	16	18.4	2426	33	HPAP1P2A	Hepatitis A Virus; 242	7.93e+01
c	87	16	18.4	2487	21	A22713	gE1/gE2 gene.	7.93e+01
c	88	16	18.4	2583	27	HUMRABCCF	Homo sapiens cellular	7.93e+01
c	89	16	18.4	2625	25	HUMCA11A	Homo sapiens cadherin-	7.93e+01
c	90	16	18.4	2625	21	I34433	Sequence 57 from paten	7.93e+01
c	91	16	18.4	2625	21	I55100	Sequence 57 from paten	7.93e+01
c	92	16	18.4	2625	21	I46831	Sequence 51 from paten	7.93e+01
	93	16	18.4	2668	28	AB015196S2	Mus musculus gene for	7.93e+01

c	94	16	18.4	2923	23	ATHMYB	A.thaliana gene encord	7.93e+01
c	95	16	18.4	3224	21	A22716	structural protein C g	7.93e+01
	96	16	18.4	3251	28	AF037199	Rattus norvegicus zinc	7.93e+01
	97	16	18.4	3370	15	SPDACA0	S.pneumoniae dacA gene	7.93e+01
	98	16	18.4	3428	26	HSAC001157	Homo sapiens (subclone	7.93e+01
c	99	16	18.4	3661	26	HUMOSF4A	Human mRNA for OB-cadh	7.93e+01
c	100	16	18.4	3697	23	ACU94398	Acetabularia cliftonii	7.93e+01
c	101	16	18.4	3712	21	E07382	cDNA encoding human OS	7.93e+01
c	102	16	18.4	3867	26	HUMOSF4B	Human mRNA for OB-cadh	7.93e+01
c	103	16	18.4	3877	25	HSIGC138B	H.sapiens (clone cosl3	7.93e+01
c	104	16	18.4	3914	21	E07383	cDNA encoding human OS	7.93e+01
c	105	16	18.4	3915	15	AF054892	Bacteroides forsythus	7.93e+01
c	106	16	18.4	4484	23	D89341	Arabidopsis thaliana D	7.93e+01
	107	16	18.4	4532	23	LMNSREPRG	L.esculentum microsate	7.93e+01
	108	16	18.4	4771	28	RNSYNP2	R.norvegicus mRNA for	7.93e+01
c	109	16	18.4	4921	24	ATHPAL	Arabidopsis thaliana (	7.93e+01
	110	16	18.4	4935	26	D86969	Human mRNA for KIAA021	7.93e+01
c	111	16	18.4	5031	23	ATHRPB3A	Arabidopsis thaliana R	7.93e+01
	112	16	18.4	5144	15	CPETOXA	C.perfringens plasmid	7.93e+01
	113	16	18.4	5370	24	AF001000	Lycopersicon esculentu	7.93e+01
c	114	16	18.4	5695	28	MMRPS3A	M.musculus RPS3a gene.	7.93e+01
	115	16	18.4	7904	33	PPH16	Human papillomavirus t	7.93e+01
	116	16	18.4	7904	21	I70142	Sequence 1 from patent	7.93e+01
	117	16	18.4	7904	21	I70143	Sequence 2 from patent	7.93e+01
	118	16	18.4	7904	21	I72345	Sequence 1 from patent	7.93e+01
	119	16	18.4	7905	33	HPU89348	Human papillomavirus t	7.93e+01
c	120	16	18.4	9133	23	EGRCPPSAAB	E.gracilis chloroplast	7.93e+01
c	121	16	18.4	12480	33	BVDPP	Bovine viral diarrhea	7.93e+01
c	122	16	18.4	12492	21	E01149	cDNA to genomic RNA of	7.93e+01
	123	16	18.4	13232	18	DDID16466	Cellular slime mold Mi	7.93e+01
c	124	16	18.4	18115	18	CELB0547	Caenorhabditis elegans	7.93e+01
c	125	16	18.4	24957	18	CER74	Caenorhabditis elegans	7.93e+01
	126	16	18.4	26510	18	CELF23H11	Caenorhabditis elegans	7.93e+01
	127	16	18.4	27284	18	CELF18A12	Caenorhabditis elegans	7.93e+01
	128	16	18.4	30601	18	CELC03F11	Caenorhabditis elegans	7.93e+01
c	129	16	18.4	31099	18	CEK03A11	Caenorhabditis elegans	7.93e+01
c	130	16	18.4	31631	18	CELD2062	Caenorhabditis elegans	7.93e+01
	131	16	18.4	31906	18	CELZK813	Caenorhabditis elegans	7.93e+01
c	132	16	18.4	33448	18	CELC34B2	Caenorhabditis elegans	7.93e+01
c	133	16	18.4	33950	18	CEF37B12	Caenorhabditis elegans	7.93e+01
	134	16	18.4	34683	26	HSU09822	Human cosmid clone 26h	7.93e+01
	135	16	18.4	34820	18	CELT24E12	Caenorhabditis elegans	7.93e+01
c	136	16	18.4	35169	18	CELC44E4	Caenorhabditis elegans	7.93e+01
	137	16	18.4	35700	26	HS27C3	Homo sapiens DNA seque	7.93e+01
c	138	16	18.4	35871	18	CELF39E9	Caenorhabditis elegans	7.93e+01
	139	16	18.4	36103	18	CEW03C9	Caenorhabditis elegans	7.93e+01
	140	16	18.4	37443	26	U73627	Human Chromosome 11 Co	7.93e+01
c	141	16	18.4	38926	25	HSF45C1B	Human DNA sequence fro	7.93e+01
	142	16	18.4	39403	18	CICOS46	Ciona intestinalis DNA	7.93e+01
	143	16	18.4	40344	18	CED2023	Caenorhabditis elegans	7.93e+01
c	144	16	18.4	41017	23	CHEGZ	Euglena gracilis Z Chl	7.93e+01

c	145	16	18.4	41125	28	AC002108	Genomic sequence from	7.93e+01
	146	16	18.4	42232	26	AC004078	Human Cosmid g1248a018	7.93e+01
	147	16	18.4	44427	18	CELF41E6	Caenorhabditis elegans	7.93e+01
c	148	16	18.4	45389	18	CELC32E12	Caenorhabditis elegans	7.93e+01
c	149	16	18.4	46610	25	HSU34879	Human 17-beta-hydroxys	7.93e+01
c	150	16	18.4	48710	23	LPJ002236	Lycopersicon pimpinell	7.93e+01
	151	16	18.4	49500	18	AC004360	Drosophila melanogaste	7.93e+01
	152	16	18.4	53042	17	HS361H15	Human DNA sequence ***	7.93e+01
c	153	16	18.4	60029	26	AC003689	Human Chromosome 11 pa	7.93e+01
c	154	16	18.4	60170	33	AF003534	Chilo iridescent virus	7.93e+01
c	155	16	18.4	60560	26	AC003678	Human Chromosome 11 pa	7.93e+01
c	156	16	18.4	63726	26	HSAC002071	Human PAC clone DJ002C	7.93e+01
c	157	16	18.4	66729	20	AF016494	Fugu rubripes gamma-am	7.93e+01
c	158	16	18.4	68316	17	CEH16D04	Caenorhabditis elegans	7.93e+01
	159	16	18.4	71522	23	AB006703	Arabidopsis thaliana g	7.93e+01
c	160	16	18.4	73308	25	HUMHBB	Human beta globin regi	7.93e+01
	161	16	18.4	75307	18	AC004114	Drosophila melanogaste	7.93e+01
	162	16	18.4	80124	18	AC002444	Drosophila melanogaste	7.93e+01
c	163	16	18.4	82415	23	AB005244	Arabidopsis thaliana g	7.93e+01
	164	16	18.4	83698	23	AB010072	Arabidopsis thaliana g	7.93e+01
c	165	16	18.4	84124	17	HSAJ9614	Homo sapiens *** SEQUE	7.93e+01
c	166	16	18.4	84203	24	AC005106	Genomic sequence for A	7.93e+01
c	167	16	18.4	85444	26	AP000023	Homo sapiens genomic D	7.93e+01
	168	16	18.4	85641	26	AC005217	Homo sapiens chromosom	7.93e+01
c	169	16	18.4	85811	26	HS105D12	Human DNA sequence fro	7.93e+01
c	170	16	18.4	87434	23	AB010698	Arabidopsis thaliana g	7.93e+01
	171	16	18.4	90832	27	AC003065	Homo sapiens Chromosom	7.93e+01
	172	16	18.4	95319	17	AC000105	*** SEQUENCING IN PROG	7.93e+01
	173	16	18.4	95968	18	CEY47D3B	Caenorhabditis elegans	7.93e+01
	174	16	18.4	107387	26	HS707K17	Human DNA sequence fro	7.93e+01
c	175	16	18.4	108086	27	HS220J16	Human DNA sequence fro	7.93e+01
c	176	16	18.4	110041	26	HS175E3	Human DNA sequence fro	7.93e+01
	177	16	18.4	110227	17	HS298J15	Human DNA sequence ***	7.93e+01
c	178	16	18.4	113916	27	HUAC003007	Human Chromosome 16 BA	7.93e+01
c	179	16	18.4	114738	23	AC002130	The sequence of BAC F1	7.93e+01
c	180	16	18.4	118153	26	HS249C1	Human DNA sequence fro	7.93e+01
c	181	16	18.4	120193	17	AC004966	*** SEQUENCING IN PROG	7.93e+01
	182	16	18.4	121131	26	HS345B16	Human DNA sequence fro	7.93e+01
	183	16	18.4	123331	26	AC005164	Homo sapiens BAC clone	7.93e+01
c	184	16	18.4	125983	26	HS306D1	Human DNA sequence fro	7.93e+01
	185	16	18.4	128148	28	AC003063	Mus musculus Chromosom	7.93e+01
	186	16	18.4	130491	26	HUAC003977	Homo sapiens Chromosom	7.93e+01
	187	16	18.4	133965	26	AF064861	Homo sapiens PAC 128M1	7.93e+01
c	188	16	18.4	134534	17	HS1163J1	Human DNA sequence ***	7.93e+01
	189	16	18.4	137728	17	AC004860	*** SEQUENCING IN PROG	7.93e+01
c	190	16	18.4	138799	26	AC004056	Homo sapiens chromosom	7.93e+01
c	191	16	18.4	140554	28	AC000399	Genomic sequence from	7.93e+01
c	192	16	18.4	143172	23	CLEGCGA	E.gracilis chloroplast	7.93e+01
	193	16	18.4	143751	26	AC004160	Homo sapiens BAC clone	7.93e+01
	194	16	18.4	145414	26	HS78F24	Homo sapiens DNA seque	7.93e+01
c	195	16	18.4	146360	26	HS1156N12	Human DNA sequence ***	7.93e+01

	196	16	18.4	148204	17	AC004881	*** SEQUENCING IN PROG	7.93e+01
	197	16	18.4	153479	17	AC005083	*** SEQUENCING IN PROG	7.93e+01
c	198	16	18.4	153479	17	AC005083	*** SEQUENCING IN PROG	7.93e+01
	199	16	18.4	153892	17	HS1121J10	Human DNA sequence ***	7.93e+01
c	200	16	18.4	154985	18	AC004641	Drosophila melanogaste	7.93e+01
	201	16	18.4	155622	17	AC004839	*** SEQUENCING IN PROG	7.93e+01
	202	16	18.4	155881	26	AC002465	Human BAC clone RG343P	7.93e+01
c	203	16	18.4	163730	17	AC004852	*** SEQUENCING IN PROG	7.93e+01
c	204	16	18.4	163889	17	AC004958	*** SEQUENCING IN PROG	7.93e+01
c	205	16	18.4	164511	27	AC004894	Homo sapiens PAC clone	7.93e+01
	206	16	18.4	167468	17	AC005033	*** SEQUENCING IN PROG	7.93e+01
	207	16	18.4	170000	17	AC004394	*** SEQUENCING IN PROG	7.93e+01
c	208	16	18.4	172241	26	AC004014	Human BAC clone RG022K	7.93e+01
	209	16	18.4	177380	26	AC002457	Human BAC clone RG060P	7.93e+01
	210	16	18.4	179707	17	HSAC001234	*** SEQUENCING IN PROG	7.93e+01
c	211	16	18.4	180388	25	HUMRETBLAS	Human retinoblastoma s	7.93e+01
c	212	16	18.4	186550	17	CEY44F5	Caenorhabditis elegans	7.93e+01
	213	16	18.4	186609	23	MPOMTCG	Marchantia polymorpha	7.93e+01
c	214	16	18.4	187051	17	AC004688	*** SEQUENCING IN PROG	7.93e+01
c	215	16	18.4	190000	17	AC004580	*** SEQUENCING IN PROG	7.93e+01
	216	16	18.4	197273	26	AC002461	Human BAC clone RG204I	7.93e+01
	217	16	18.4	198499	17	AC004870	*** SEQUENCING IN PROG	7.93e+01
	218	16	18.4	199082	17	AC005632	*** SEQUENCING IN PROG	7.93e+01
	219	16	18.4	199927	17	HS101A4	Human DNA sequence ***	7.93e+01
c	220	16	18.4	200167	26	AC002480	Human BAC clone GS055K	7.93e+01
	221	16	18.4	203789	17	AC004032	*** SEQUENCING IN PROG	7.93e+01
	222	16	18.4	204877	17	AC005375	*** SEQUENCING IN PROG	7.93e+01
	223	16	18.4	207139	17	CEY17G7	Caenorhabditis elegans	7.93e+01
	224	16	18.4	234053	26	AC002429	Human BAC clone GS200K	7.93e+01
	225	16	18.4	239737	17	HS353E16	Human DNA sequence ***	7.93e+01
c	226	16	18.4	244649	17	AC005308	*** SEQUENCING IN PROG	7.93e+01
	227	16	18.4	245101	17	HS437G10	Human DNA sequence ***	7.93e+01
c	228	16	18.4	259474	26	HUAC004605	Homo sapiens Chromosom	7.93e+01
c	229	16	18.4	301692	26	D87675	Homo sapiens DNA for a	7.93e+01
	230	16	18.4	337546	17	CEY47D3	Caenorhabditis elegans	7.93e+01
c	231	15	17.2	224	30	G20957	human STS WI-16705.	3.09e+02
c	232	15	17.2	260	30	G09304	human STS CHLC.ATA5A08	3.09e+02
	233	15	17.2	280	21	A48042	Sequence 109 from Pate	3.09e+02
c	234	15	17.2	298	19	SXCMTRRAH	Sorex cinereus cinereu	3.09e+02
c	235	15	17.2	358	30	G27269	human STS CHLC.GATA181	3.09e+02
	236	15	17.2	367	30	HSA062WE5	H.sapiens (D10S1782) D	3.09e+02
c	237	15	17.2	370	30	G01078	fruit fly STS Dm1736 c	3.09e+02
c	238	15	17.2	374	30	G08683	human STS CHLC.GATA25C	3.09e+02
	239	15	17.2	379	30	G07403	human STS WI-7758, seq	3.09e+02
c	240	15	17.2	387	30	G27280	human STS CHLC.GATA183	3.09e+02
	241	15	17.2	391	33	AF014233	HIV-1 Patient 1256 fro	3.09e+02
	242	15	17.2	406	18	CTKKOCT	C.texile KK-0 mRNA for	3.09e+02
	243	15	17.2	409	30	G09977	human STS CHLC.GATA63E	3.09e+02
	244	15	17.2	425	30	G25635	human STS EST110906.	3.09e+02
	245	15	17.2	433	25	HSC7EX12	H.sapiens C7 gene, exo	3.09e+02
c	246	15	17.2	435	30	HUMUT752	Human chromosome 17 ST	3.09e+02

	247	15	17.2	443	18	AF008988	Plasmodium falciparum	3.09e+02
c	248	15	17.2	464	18	OSU60199	Ophiura signata large	3.09e+02
c	249	15	17.2	507	30	G27775	human STS SHGC-33214.	3.09e+02
	250	15	17.2	523	30	HUM45TS55	Human chromosome 4 seq	3.09e+02
	251	15	17.2	538	20	XLXO20A	X.laevis XO20 mRNA for	3.09e+02
c	252	15	17.2	555	15	MC019	M.capricolum DNA for C	3.09e+02
	253	15	17.2	575	27	HUMYQ15A02	Homo sapiens full leng	3.09e+02
	254	15	17.2	604	25	HUMMDR1A26	Human P-glycoprotein (	3.09e+02
	255	15	17.2	675	25	HSC7EXN12	H.sapiens C7 gene, exo	3.09e+02
c	256	15	17.2	704	23	SRRBCS	Radish mRNA for ribulo	3.09e+02
c	257	15	17.2	736	15	MCAAM	M.capricolum DNA for C	3.09e+02
	258	15	17.2	757	20	XLXO35A	X.laevis XO35 mRNA for	3.09e+02
	259	15	17.2	811	23	RICMT10	Rice mitochondrial DNA	3.09e+02
	260	15	17.2	988	23	CHCRL14	Chlamydomonas reinhard	3.09e+02
	261	15	17.2	996	15	CU8854	C.upsaliensis (LMG 885	3.09e+02
	262	15	17.2	998	26	D78195	Homo sapiens J-chain g	3.09e+02
	263	15	17.2	1047	23	CPHAPSBA	H.annuus chloroplast p	3.09e+02
	264	15	17.2	1055	23	CRECPDL16	C.reinhardtii chloropl	3.09e+02
	265	15	17.2	1163	23	RICCPORF	Rice chloroplast genes	3.09e+02
	266	15	17.2	1292	26	HSAPOA1P	H.sapiens ApoA-1 gene	3.09e+02
	267	15	17.2	1339	20	CCTPMY05	Quail gene for alpha-t	3.09e+02
	268	15	17.2	1351	23	SCYKL056C	S.cerevisiae chromosom	3.09e+02
	269	15	17.2	1409	18	DMZ86118	D.melanogaster mRNA fo	3.09e+02
	270	15	17.2	1440	15	CLOPLC3	Clostridium novyi (str	3.09e+02
	271	15	17.2	1450	25	HUMBTFB	Human BTF3 protein hom	3.09e+02
	272	15	17.2	1542	18	PFCDPK2	P.falciparum CDPK2 gen	3.09e+02
	273	15	17.2	1593	20	XLNOVA	X.laevis mRNA for NOVA	3.09e+02
	274	15	17.2	1603	15	ECOORF2X4X	Escherichia coli wbbK	3.09e+02
c	275	15	17.2	1608	26	AF070620	Homo sapiens clone 246	3.09e+02
c	276	15	17.2	1666	18	EAGTUB	E.sediculatus gamma-tu	3.09e+02
	277	15	17.2	1745	18	DMTEAA	D.melanogaster T-eleme	3.09e+02
	278	15	17.2	1776	28	GPIFCMOX	Cavia porcellus flavin	3.09e+02
	279	15	17.2	1825	18	EHSRPAP	E.histolytica genes en	3.09e+02
	280	15	17.2	1918	24	AF061285	Capsicum annuum sesqui	3.09e+02
c	281	15	17.2	1985	23	YSCUGLSR1	Yeast (S.cerevisiae) L	3.09e+02
c	282	15	17.2	2000	25	BABHEBDPS	monkey (anubis) silent	3.09e+02
c	283	15	17.2	2006	18	MSQTMTATAE	Anopheles albimanus (c	3.09e+02
	284	15	17.2	2137	28	MMUSGAMT2	Mus musculus guanidino	3.09e+02
	285	15	17.2	2148	26	HSU35451	Homo sapiens heterochr	3.09e+02
	286	15	17.2	2267	16	SSY16944	Staphylococcus sp. pla	3.09e+02
c	287	15	17.2	2289	18	CELCPR6G	Caenorhabditis elegans	3.09e+02
c	288	15	17.2	2295	19	OCRAG23UT	O.cuniculus RAG-2 gene	3.09e+02
c	289	15	17.2	2382	26	HSU15173	Homo sapiens BCL2/aden	3.09e+02
	290	15	17.2	2409	28	CRUPGP2A	Chinese hamster P-glyc	3.09e+02
	291	15	17.2	2467	23	LETPS11	L.esculentum TPS11 gen	3.09e+02
	292	15	17.2	2470	25	HUMAPOAI	Human apolipoprotein A	3.09e+02
	293	15	17.2	2616	20	CIIHGAP	Duck embryonic alpha-g	3.09e+02
c	294	15	17.2	2644	24	ATDNALIG	A.thaliana mRNA for DN	3.09e+02
	295	15	17.2	2726	21	I33621	Sequence 1 from patent	3.09e+02
	296	15	17.2	2734	25	HUMCS5	Human chorionic somato	3.09e+02
	297	15	17.2	2735	18	DMU06921	Drosophila melanogaste	3.09e+02

c	298	15	17.2	2745	16	STU25352	Salmonella typhimurium	3.09e+02
	299	15	17.2	2873	25	HUMAPOAIA	Human apolipoprotein A	3.09e+02
	300	15	17.2	2960	23	SCYBR230C	S.cerevisiae chromosom	3.09e+02
c	301	15	17.2	3116	33	SWVRNAP	Satsuma dwarf virus ge	3.09e+02
	302	15	17.2	3176	23	SCCLB2G	S.cerevisiae CLB2 gene	3.09e+02
c	303	15	17.2	3272	18	DMINSCUTE	D.melanogaster mRNA fo	3.09e+02
	304	15	17.2	3325	23	SCYBR231C	S.cerevisiae chromosom	3.09e+02
	305	15	17.2	3356	23	SCU14728	Saccharomyces cerevisi	3.09e+02
	306	15	17.2	3508	28	MUSFSTC6	Mouse fast skeletal tr	3.09e+02
	307	15	17.2	3515	24	CAN5588	Capsicum annuum gene e	3.09e+02
	308	15	17.2	3660	18	SPP3A2	S.purpuratus mRNA for	3.09e+02
c	309	15	17.2	3850	15	MHTHYPHO	M.hominis deo A gene f	3.09e+02
	310	15	17.2	3877	23	SCYKL057C	S.cerevisiae chromosom	3.09e+02
	311	15	17.2	3902	28	MUSYAC11SR	Mouse Y specific regio	3.09e+02
c	312	15	17.2	3948	18	DMU73489	Drosophila melanogaste	3.09e+02
c	313	15	17.2	3977	15	GCGLUABCD	C.glutamicum gluA, gluB	3.09e+02
c	314	15	17.2	4166	25	HUMPROFL02	Human profilaggrin (FL	3.09e+02
	315	15	17.2	4280	28	CRUPGPII	C.griseus P-glycoprote	3.09e+02
	316	15	17.2	4286	28	S67043	oligodendrocyte-myelin	3.09e+02
	317	15	17.2	4309	23	SCYOR354C	S.cerevisiae chromosom	3.09e+02
	318	15	17.2	4378	21	E02326	Multidrug resistance r	3.09e+02
c	319	15	17.2	4447	25	HUMPROFILA	Human profilaggrin gen	3.09e+02
c	320	15	17.2	4519	18	PPSC04089	Plasmodium falciparum	3.09e+02
	321	15	17.2	4646	25	HUMMDR1	Human P-glycoprotein (	3.09e+02
	322	15	17.2	4646	21	I49610	Sequence 2 from patent	3.09e+02
	323	15	17.2	4669	21	I08557	Sequence 3 from Patent	3.09e+02
c	324	15	17.2	4697	15	SPU67397	Synechocystis sp. PCC6	3.09e+02
c	325	15	17.2	4867	20	AF055980	Xenopus laevis insulin	3.09e+02
	326	15	17.2	4894	26	HUMORFKA	Human mRNA for KIAA003	3.09e+02
c	327	15	17.2	5073	24	AF025430	Papaver somniferum ber	3.09e+02
c	328	15	17.2	5159	23	BLYGLB2	Hordeum vulgare 1,3-1,	3.09e+02
c	329	15	17.2	5302	15	BACPST	Bacillus subtilis DNA	3.09e+02
c	330	15	17.2	5311	25	HSTHDB	H.sapiens gene for thi	3.09e+02
	331	15	17.2	5701	23	PMCMSG102B	Pneumocystis carinii g	3.09e+02
	332	15	17.2	5798	23	AF031649	Arabidopsis thaliana n	3.09e+02
	333	15	17.2	5874	24	AF007808	Daucus carota Met2-typ	3.09e+02
c	334	15	17.2	6542	26	AB014524	Homo sapiens mRNA for	3.09e+02
c	335	15	17.2	6750	23	NSD64823	Nicotiana sylvestris a	3.09e+02
	336	15	17.2	7439	16	ECRFBA	Escherichia coli K12 g	3.09e+02
c	337	15	17.2	7654	33	PPH8CG	Human papillomavirus t	3.09e+02
c	338	15	17.2	7876	33	SVBVCOMGN	Strawberry vein bandin	3.09e+02
c	339	15	17.2	8038	28	MMMSH6S03	Mus musculus mismatch	3.09e+02
	340	15	17.2	8216	20	XELFBRNCT	Xenopus laevis fibrone	3.09e+02
	341	15	17.2	8682	23	CIP345RSS	Mesembryanthemum cryst	3.09e+02
c	342	15	17.2	8939	20	XELIP3R	X. laevis mRNA for ino	3.09e+02
	343	15	17.2	9005	23	LEHSF8	L.esculentum Le-hsf8 g	3.09e+02
c	344	15	17.2	9048	21	A58827	Sequence 4 from Patent	3.09e+02
c	345	15	17.2	9763	23	ATZ82992	A.thaliana PCA gene en	3.09e+02
c	346	15	17.2	9763	21	A58824	Sequence 1 from Patent	3.09e+02
	347	15	17.2	10042	15	HPAE000574	Helicobacter pylori se	3.09e+02
	348	15	17.2	10214	16	MCU62057	Mycoplasma capricolum	3.09e+02

c	349	15	17.2	10634	15	AE000877	Methanobacterium therm 3.09e+02
	350	15	17.2	10776	18	CELGLP1A	C.elegans glp-1 gene, 3.09e+02
	351	15	17.2	11268	15	APU38537	Anabaena PCC7120 nifD 3.09e+02
c	352	15	17.2	11406	15	AE000987	Archaeoglobus fulgidus 3.09e+02
c	353	15	17.2	13894	28	RNU68562	Rattus norvegicus chap 3.09e+02
	354	15	17.2	14358	15	HPAE000605	Helicobacter pylori se 3.09e+02
	355	15	17.2	14392	25	HSU33837	Human glycoprotein rec 3.09e+02
	356	15	17.2	15246	33	AF052755	Simian parainfluenza v 3.09e+02
c	357	15	17.2	16032	15	ECAE000294	Escherichia coli K-12 3.09e+02
c	358	15	17.2	17630	25	HUMPROFILX	Human profilaggrin gen 3.09e+02
c	359	15	17.2	18856	15	D90842	E.coli genomic DNA, Ko 3.09e+02
	360	15	17.2	20048	18	CEC33A11	Caenorhabditis elegans 3.09e+02
c	361	15	17.2	20226	15	D90841	E.coli genomic DNA, Ko 3.09e+02
c	362	15	17.2	21836	18	CELZK40	Caenorhabditis elegans 3.09e+02
	363	15	17.2	25164	23	YSCP9642	Saccharomyces cerevisi 3.09e+02
c	364	15	17.2	26039	15	CD26039	C.difficile cdu2, cdu1 3.09e+02
	365	15	17.2	26242	18	CEF02A9	Caenorhabditis elegans 3.09e+02
	366	15	17.2	26802	17	AC005708	*** SEQUENCING IN PROG 3.09e+02
	367	15	17.2	27019	18	CEF47C8	Caenorhabditis elegans 3.09e+02
	368	15	17.2	27263	24	AB016881	Arabidopsis thaliana g 3.09e+02
	369	15	17.2	28600	23	SCXI286K	S.cerevisiae chromosom 3.09e+02
	370	15	17.2	28601	15	AE000784	Borrelia burgdorferi p 3.09e+02
	371	15	17.2	29410	23	YSCLTESPO	Saccharomyces cerevisi 3.09e+02
	372	15	17.2	29952	25	HSV1164A6	Human DNA sequence fro 3.09e+02
	373	15	17.2	30985	24	SPAC9G1	S.pombe chromosome I c 3.09e+02
c	374	15	17.2	31827	17	AB018115	*** SEQUENCING IN PROG 3.09e+02
	375	15	17.2	32087	18	CELF26G5	Caenorhabditis elegans 3.09e+02
c	376	15	17.2	32195	18	AC004154	Drosophila melanogaste 3.09e+02
c	377	15	17.2	33614	18	CELC25B8	Caenorhabditis elegans 3.09e+02
	378	15	17.2	33783	18	CELW09G10	Caenorhabditis elegans 3.09e+02
c	379	15	17.2	33783	18	CELW09G10	Caenorhabditis elegans 3.09e+02
	380	15	17.2	34388	18	CET28A8	Caenorhabditis elegans 3.09e+02
c	381	15	17.2	34541	26	AC005338	Homo sapiens chromosom 3.09e+02
c	382	15	17.2	34875	26	AC005329	Homo sapiens chromosom 3.09e+02
	383	15	17.2	35417	18	CELC24G6	Caenorhabditis elegans 3.09e+02
c	384	15	17.2	35730	26	HSN74G7	Human DNA sequence fro 3.09e+02
c	385	15	17.2	36224	27	AC005788	Homo sapiens chromosom 3.09e+02
c	386	15	17.2	37012	18	CELH17B01	Caenorhabditis elegans 3.09e+02
	387	15	17.2	37416	18	CEF09A5	Caenorhabditis elegans 3.09e+02
	388	15	17.2	38528	26	HSAC001642	Human Chromosome 11 co 3.09e+02
c	389	15	17.2	38570	18	CELC06A8	Caenorhabditis elegans 3.09e+02
c	390	15	17.2	38640.	18	CELF44A2	Caenorhabditis elegans 3.09e+02
	391	15	17.2	38980	18	CELZC13	Caenorhabditis elegans 3.09e+02
	392	15	17.2	39416	18	CELC05E4	Caenorhabditis elegans 3.09e+02
	393	15	17.2	39683	25	HUMN79E2	Human cosmid N79E2, co 3.09e+02
	394	15	17.2	40299	25	HSU59569	Human Xq28 cosmid U21A 3.09e+02
c	395	15	17.2	40469	18	MIPAGEN	Paramecium aurelia mit 3.09e+02
	396	15	17.2	40550	18	CELR12A1	Caenorhabditis elegans 3.09e+02
c	397	15	17.2	40906	18	CELP38E9	Caenorhabditis elegans 3.09e+02
	398	15	17.2	41988	25	HSV362H12	Human DNA sequence fro 3.09e+02
	399	15	17.2	42041	18	CELF36D4	Caenorhabditis elegans 3.09e+02

c	400	15	17.2	42467	18	CELF26F12	Caenorhabditis elegans 3.09e+02
c	401	15	17.2	42734	18	CEDY3	Caenorhabditis elegans 3.09e+02
c	402	15	17.2	43254	18	CELC17E7	Caenorhabditis elegans 3.09e+02
	403	15	17.2	44373	26	HSSRL11M2	Human DNA sequence fro 3.09e+02
	404	15	17.2	44422	18	L32205	Drosophila melanogaste 3.09e+02
c	405	15	17.2	44726	18	CELF09E5	Caenorhabditis elegans 3.09e+02
	406	15	17.2	45966	26	AP000028	Homo sapiens genomic D 3.09e+02
c	407	15	17.2	46288	20	FRU90880	Fugu rubripes zinc fin 3.09e+02
	408	15	17.2	47785	18	CELF45E4	Caenorhabditis elegans 3.09e+02
c	409	15	17.2	47785	18	CELF45E4	Caenorhabditis elegans 3.09e+02
	410	15	17.2	48096	18	CET20G5	Caenorhabditis elegans 3.09e+02
	411	15	17.2	50641	23	AB008270	Arabidopsis thaliana g 3.09e+02
	412	15	17.2	53221	26	AF057177	Homo sapiens T-cell re 3.09e+02
	413	15	17.2	55328	23	PWU02970	Prototheca wickerhamii 3.09e+02
	414	15	17.2	56105	15	SSS56KBPR	S.solfataricus 56 kbp 3.09e+02
c	415	15	17.2	57328	18	PFSC03065	Plasmodium falciparum 3.09e+02
	416	15	17.2	58407	16	MII2CG	Methanococcus jannasch 3.09e+02
	417	15	17.2	59261	24	T12M4	Arabidopsis thaliana c 3.09e+02
	418	15	17.2	59383	18	PFSC03066	Plasmodium falciparum 3.09e+02
	419	15	17.2	59498	18	AC004252	Drosophila melanogaste 3.09e+02
	420	15	17.2	59782	18	PFSC03093	Plasmodium falciparum 3.09e+02
	421	15	17.2	61380	24	AB017066	Arabidopsis thaliana g 3.09e+02
	422	15	17.2	63749	26	AC004694	Homo sapiens BAC clone 3.09e+02
	423	15	17.2	66495	25	HUMGHCSA	Human growth hormone (3.09e+02
c	424	15	17.2	69034	18	AF007261	Reclinomonas americana 3.09e+02
	425	15	17.2	69060	26	AC005196	Homo sapiens chromosom 3.09e+02
c	426	15	17.2	71482	17	HSNS6E8	Human DNA sequence *** 3.09e+02
c	427	15	17.2	71653	26	HS133H11	Human DNA sequence fro 3.09e+02
c	428	15	17.2	71668	26	AC005195	Homo sapiens chromosom 3.09e+02
c	429	15	17.2	72450	17	AC005658	*** SEQUENCING IN PROG 3.09e+02
	430	15	17.2	73076	18	AC003055	Drosophila melanogaste 3.09e+02
	431	15	17.2	74560	26	HS433B8	Human DNA sequence fro 3.09e+02
c	432	15	17.2	74562	26	HS225D2	Human DNA sequence fro 3.09e+02
	433	15	17.2	74717	18	AC004563	Drosophila melanogaste 3.09e+02
	434	15	17.2	74842	24	AB016877	Arabidopsis thaliana g 3.09e+02
	435	15	17.2	76848	17	PFMAL1P2	Plasmodium falciparum 3.09e+02
	436	15	17.2	78181	23	AB011477	Arabidopsis thaliana g 3.09e+02
	437	15	17.2	84194	23	AB006701	Arabidopsis thaliana g 3.09e+02
c	438	15	17.2	84325	24	AB015475	Arabidopsis thaliana g 3.09e+02
c	439	15	17.2	84969	26	AC005178	Homo sapiens chromosom 3.09e+02
c	440	15	17.2	85132	26	U82668	Homo sapiens shox gene 3.09e+02
	441	15	17.2	85230	17	AC004989	*** SEQUENCING IN PROG 3.09e+02
c	442	15	17.2	86147	17	AC004433	*** SEQUENCING IN PROG 3.09e+02
c	443	15	17.2	86374	17	AC005110	*** SEQUENCING IN PROG 3.09e+02
c	444	15	17.2	88099	24	AFI13D4	Arabidopsis thaliana D 3.09e+02
c	445	15	17.2	88128	24	AB017069	Arabidopsis thaliana g 3.09e+02
	446	15	17.2	89016	26	HS86C11	Homo sapiens DNA seque 3.09e+02
c	447	15	17.2	89626	26	AC004776	Homo sapiens chromosom 3.09e+02
	448	15	17.2	90169	26	HS398A12	Human DNA sequence fro 3.09e+02
c	449	15	17.2	90395	17	AC004515	*** SEQUENCING IN PROG 3.09e+02
	450	15	17.2	93932	18	AC004766	Drosophila melanogaste 3.09e+02

	451	15	17.2	95129	26	AC003099	Homo sapiens chromosom 3.09e+02
c	452	15	17.2	95167	23	ATAC003105	Arabidopsis thaliana c 3.09e+02
	453	15	17.2	96675	17	AC004287	*** SEQUENCING IN PROG 3.09e+02
c	454	15	17.2	96792	26	HSU95741	Human chromosome 16p13 3.09e+02
	455	15	17.2	97495	24	ATAC003058	Arabidopsis thaliana c 3.09e+02
c	456	15	17.2	99014	26	HSU78027	Homo sapiens Bruton's 3.09e+02
	457	15	17.2	99033	27	AC005704	Homo sapiens Xp22 bins 3.09e+02
	458	15	17.2	100000	17	AC005543	*** SEQUENCING IN PROG 3.09e+02
	459	15	17.2	100921	27	AC004996	Homo sapiens PAC clone 3.09e+02
c	460	15	17.2	100923	26	HSAC002124	Human BAC clone RG180D 3.09e+02
	461	15	17.2	101311	26	HUAC004513	Homo sapiens Chromosom 3.09e+02
c	462	15	17.2	102703	26	AC004910	Homo sapiens PAC clone 3.09e+02
	463	15	17.2	103735	17	AC004959	*** SEQUENCING IN PROG 3.09e+02
	464	15	17.2	104492	26	AC003985	Human PAC clone DJ0978 3.09e+02
c	465	15	17.2	104597	26	HS393P12	Homo sapiens DNA seque 3.09e+02
c	466	15	17.2	104710	26	AC002488	Homo sapiens chromosom 3.09e+02
	467	15	17.2	106007	23	ATT19K4	Arabidopsis thaliana D 3.09e+02
c	468	15	17.2	106007	23	ATT19K4	Arabidopsis thaliana D 3.09e+02
	469	15	17.2	106184	23	TM018A10	Arabidopsis thaliana B 3.09e+02
	470	15	17.2	106305	23	AC002304	Genomic sequence for A 3.09e+02
c	471	15	17.2	106355	17	AC005309	*** SEQUENCING IN PROG 3.09e+02
	472	15	17.2	106988	26	AC001226	Genomic sequence from 3.09e+02
c	473	15	17.2	107190	27	AC005611	Homo sapiens chromosom 3.09e+02
c	474	15	17.2	107298	25	HSAC000059	Human BAC clone RG128M 3.09e+02
c	475	15	17.2	107557	23	F19P19	Sequence of BAC F19P19 3.09e+02
	476	15	17.2	108260	26	HS179P9	Human DNA sequence fro 3.09e+02
c	477	15	17.2	110000	17	CEY111B2	Caenorhabditis elegans 3.09e+02
c	478	15	17.2	110209	26	HUAC004533	Homo sapiens Chromosom 3.09e+02
c	479	15	17.2	110401	26	AC004068	Homo sapiens chromosom 3.09e+02
c	480	15	17.2	112762	26	HS323B6	Human DNA sequence fro 3.09e+02
c	481	15	17.2	113704	25	HS267P19	Human DNA sequence fro 3.09e+02
	482	15	17.2	116173	17	AC005659	*** SEQUENCING IN PROG 3.09e+02
c	483	15	17.2	116884	17	PFMAL1P1	Plasmodium falciparum 3.09e+02
	484	15	17.2	117252	17	AC005011	*** SEQUENCING IN PROG 3.09e+02
	485	15	17.2	117994	17	HS409O10	Human DNA sequence *** 3.09e+02
	486	15	17.2	118451	27	AC004930	Homo sapiens PAC clone 3.09e+02
c	487	15	17.2	118591	17	AC005024	*** SEQUENCING IN PROG 3.09e+02
	488	15	17.2	118907	17	HS799N4	Human DNA sequence *** 3.09e+02
c	489	15	17.2	118907	17	HS799N4	Human DNA sequence *** 3.09e+02
c	490	15	17.2	119196	27	AC005741	Homo sapiens chromosom 3.09e+02
c	491	15	17.2	119566	25	HS274L7	Human DNA sequence fro 3.09e+02
c	492	15	17.2	120376	26	HS271G9	Human DNA sequence fro 3.09e+02
c	493	15	17.2	122256	17	AC005443	*** SEQUENCING IN PROG 3.09e+02
	494	15	17.2	123234	26	AC002556	Human Chromosome 11 pa 3.09e+02
c	495	15	17.2	123386	24	F12F1	Arabidopsis thaliana c 3.09e+02
	496	15	17.2	123468	17	HS232N11	Human DNA sequence *** 3.09e+02
	497	15	17.2	123491	17	HS126A5	Human DNA sequence *** 3.09e+02
	498	15	17.2	123891	26	AC004549	Homo sapiens BAC clone 3.09e+02
	499	15	17.2	124351	27	AC005185	Homo sapiens Xp22 bins 3.09e+02
c	500	15	17.2	124351	27	AC005185	Homo sapiens Xp22 bins 3.09e+02
c	501	15	17.2	125457	17	AC005045	*** SEQUENCING IN PROG 3.09e+02

c	502	15	17.2	125630	26	HSU91319	Human chromosome 16p13 3.09e+02
c	503	15	17.2	126147	26	HS433M19	Human DNA sequence fro 3.09e+02
c	504	15	17.2	127757	26	AC004468	Homo sapiens Xp22 PAC 3.09e+02
c	505	15	17.2	127816	17	AC004822	*** SEQUENCING IN PROG 3.09e+02
c	506	15	17.2	128473	27	AC004241	Homo sapiens 12q13.1 P 3.09e+02
c	507	15	17.2	128993	17	AC005068	*** SEQUENCING IN PROG 3.09e+02
	508	15	17.2	129169	17	AC005506	*** SEQUENCING IN PROG 3.09e+02
c	509	15	17.2	130001	15	D90915	Synechocystis sp. PCC6 3.09e+02
	510	15	17.2	130064	17	AC002509	*** SEQUENCING IN PROG 3.09e+02
	511	15	17.2	131223	17	AC004853	*** SEQUENCING IN PROG 3.09e+02
	512	15	17.2	132666	18	AC005270	Drosophila melanogaste 3.09e+02
c	513	15	17.2	133469	17	HSAC000383	*** SEQUENCING IN PROG 3.09e+02
c	514	15	17.2	134578	17	AC004824	*** SEQUENCING IN PROG 3.09e+02
	515	15	17.2	134970	26	AC002504	Human Xp22 PAC RPCI1-2 3.09e+02
c	516	15	17.2	135505	27	AC004477	Homo sapiens chromosom 3.09e+02
c	517	15	17.2	136037	26	AC004104	Homo sapiens 12q24 PAC 3.09e+02
c	518	15	17.2	137260	17	PFMAL1P4	Plasmodium falciparum 3.09e+02
c	519	15	17.2	138343	17	AC004907	*** SEQUENCING IN PROG 3.09e+02
c	520	15	17.2	139350	26	HS443E24	Human DNA sequence fro 3.09e+02
	521	15	17.2	139467	15	D90905	Synechocystis sp. PCC6 3.09e+02
c	522	15	17.2	140060	17	AC005101	*** SEQUENCING IN PROG 3.09e+02
c	523	15	17.2	141303	17	AC004943	*** SEQUENCING IN PROG 3.09e+02
	524	15	17.2	142278	26	AC005177	Homo sapiens chromosom 3.09e+02
c	525	15	17.2	142661	17	AC005482	*** SEQUENCING IN PROG 3.09e+02
c	526	15	17.2	143612	17	HS936P19	Human DNA sequence *** 3.09e+02
	527	15	17.2	145076	17	AC005521	*** SEQUENCING IN PROG 3.09e+02
	528	15	17.2	145844	26	AC004931	Homo sapiens PAC clone 3.09e+02
	529	15	17.2	146963	26	HS964D12	Homo sapiens DNA seque 3.09e+02
c	530	15	17.2	148400	26	AC004768	Homo sapiens chromosom 3.09e+02
c	531	15	17.2	148410	17	AC004841	*** SEQUENCING IN PROG 3.09e+02
	532	15	17.2	149566	26	AC003081	Human BAC clone RG009H 3.09e+02
	533	15	17.2	149764	26	HSAC002076	Human BAC clone GS345D 3.09e+02
c	534	15	17.2	150140	17	AC005478	*** SEQUENCING IN PROG 3.09e+02
c	535	15	17.2	151770	26	HSAC002066	Human BAC clone RG030H 3.09e+02
	536	15	17.2	151874	26	AC004594	Homo sapiens PAC clone 3.09e+02
	537	15	17.2	152184	17	CEY70G10	Caenorhabditis elegans 3.09e+02
c	538	15	17.2	152593	23	ATAC002332	Arabidopsis thaliana c 3.09e+02
c	539	15	17.2	154028	26	HSAC002087	Human BAC clone GS113D 3.09e+02
	540	15	17.2	159024	17	HS15D7	Human DNA sequence *** 3.09e+02
c	541	15	17.2	159506	26	HS34I8	Homo sapiens DNA seque 3.09e+02
	542	15	17.2	160048	28	AC002109	Genomic sequence from 3.09e+02
c	543	15	17.2	160850	17	AC004957	*** SEQUENCING IN PROG 3.09e+02
	544	15	17.2	161475	26	HS164C20	Homo sapiens DNA seque 3.09e+02
c	545	15	17.2	161882	17	AC004862	*** SEQUENCING IN PROG 3.09e+02
	546	15	17.2	165228	26	AC005325	Homo sapiens chromosom 3.09e+02
	547	15	17.2	166219	26	AC003084	Human BAC clone RG084D 3.09e+02
	548	15	17.2	171258	17	HS873P14	Human DNA sequence *** 3.09e+02
c	549	15	17.2	172983	26	AC002454	Human BAC clone GS552A 3.09e+02
	550	15	17.2	173494	17	HS662N3	Human DNA sequence *** 3.09e+02
c	551	15	17.2	173640	17	AC005236	*** SEQUENCING IN PROG 3.09e+02
c	552	15	17.2	173882	26	HSAF001550	Homo sapiens chromosom 3.09e+02

	553	15	17.2	174565	17	HS197017	Human DNA sequence *** 3.09e+02
	554	15	17.2	179076	17	AC004890	*** SEQUENCING IN PROG 3.09e+02
c	555	15	17.2	179174	17	AC004796	*** SEQUENCING IN PROG 3.09e+02
c	556	15	17.2	180175	17	AC005105	*** SEQUENCING IN PROG 3.09e+02
	557	15	17.2	181995	17	HS596C15	Human DNA sequence *** 3.09e+02
	558	15	17.2	184490	26	HSU82828	Homo sapiens ataxia te 3.09e+02
	559	15	17.2	185775	25	HUMFMR1S	Homo sapiens fragile X 3.09e+02
	560	15	17.2	185994	17	AC002042	*** SEQUENCING IN PROG 3.09e+02
	561	15	17.2	190248	17	AC004988	*** SEQUENCING IN PROG 3.09e+02
c	562	15	17.2	192056	17	AC004832	*** SEQUENCING IN PROG 3.09e+02
	563	15	17.2	196475	17	AC005386	*** SEQUENCING IN PROG 3.09e+02
	564	15	17.2	196672	18	AC004767	Drosophila melanogaste 3.09e+02
	565	15	17.2	197900	26	AC002531	Homo sapiens chromosom 3.09e+02
	566	15	17.2	198060	17	AC004831	*** SEQUENCING IN PROG 3.09e+02
	567	15	17.2	198570	26	HS143C1	Human DNA sequence fro 3.09e+02
	568	15	17.2	198605	17	AC005003	*** SEQUENCING IN PROG 3.09e+02
c	569	15	17.2	198776	17	AC005075	*** SEQUENCING IN PROG 3.09e+02
	570	15	17.2	199490	17	AC005081	*** SEQUENCING IN PROG 3.09e+02
c	571	15	17.2	200611	23	ATFCA0	Arabidopsis thaliana D 3.09e+02
c	572	15	17.2	204596	17	AC005069	*** SEQUENCING IN PROG 3.09e+02
c	573	15	17.2	209071	18	AC005285	Drosophila melanogaste 3.09e+02
	574	15	17.2	209165	23	ATFCA5	Arabidopsis thaliana D 3.09e+02
c	575	15	17.2	209617	17	HS54B20	Human DNA sequence *** 3.09e+02
c	576	15	17.2	214521	17	HS172B20	Human DNA sequence *** 3.09e+02
	577	15	17.2	216916	17	AC004970	*** SEQUENCING IN PROG 3.09e+02
	578	15	17.2	218074	27	HUAC002044	Human Chromosome 16 BA 3.09e+02
c	579	15	17.2	218074	27	HUAC002044	Human Chromosome 16 BA 3.09e+02
	580	15	17.2	218470	15	BSUB0013	Bacillus subtilis comp 3.09e+02
	581	15	17.2	222922	17	AC005015	*** SEQUENCING IN PROG 3.09e+02
	582	15	17.2	227730	17	CEY56A3	Caenorhabditis elegans 3.09e+02
c	583	15	17.2	227892	17	AC005587	*** SEQUENCING IN PROG 3.09e+02
	584	15	17.2	236281	26	AC004673	Homo sapiens Xp22 BAC 3.09e+02
	585	15	17.2	251969	17	CEY79H2	Caenorhabditis elegans 3.09e+02
c	586	15	17.2	253505	16	AP000007	Pyrococcus horikoshii 3.09e+02
c	587	15	17.2	259202	26	AC002366	Human Xp22 BAC CT-285I 3.09e+02
	588	15	17.2	269078	17	CEY76A2	Caenorhabditis elegans 3.09e+02
c	589	15	17.2	282700	15	BACJH642	Bacillus subtilis DNA, 3.09e+02
	590	15	17.2	302988	17	CEY46G5	Caenorhabditis elegans 3.09e+02
	591	15	17.2	310126	17	CEY48B6	Caenorhabditis elegans 3.09e+02
	592	15	17.2	310213	17	AC004526	*** SEQUENCING IN PROG 3.09e+02
c	593	15	17.2	335867	17	CEY87G2	Caenorhabditis elegans 3.09e+02
	594	15	17.2	336638	17	CEY75B8	Caenorhabditis elegans 3.09e+02
	595	15	17.2	338234	18	DMU31961	Drosophila melanogaste 3.09e+02
	596	14	16.1	205	19	PIGS0082X	Pig microsatellite rep 1.13e+03
c	597	14	16.1	212	18	CMU36333	Cladomyrma maschwitzi 1.13e+03
c	598	14	16.1	236	25	HS8B2R	H.sapiens CpG island D 1.13e+03
c	599	14	16.1	260	26	HSZ85842	H.sapiens BP2P2-Kg6-D0 1.13e+03
	600	14	16.1	278	33	TOHJV1	3' terminal sequence o 1.13e+03
	601	14	16.1	279	21	A48093	Sequence 160 from Pate 1.13e+03
	602	14	16.1	279	15	MASQVMAC1	M.avium sequevar MAC-I 1.13e+03
c	603	14	16.1	280	26	HSZ85898	H.sapiens BP2P3-Kg6-B0 1.13e+03

	604	14	16.1	282	21	A48016	Sequence 83 from Paten 1.13e+03
	605	14	16.1	286	21	A48092	Sequence 159 from Pate 1.13e+03
c	606	14	16.1	287	25	HSIGDRE26	H.sapiens germline imm 1.13e+03
c	607	14	16.1	288	25	HS71C3F	H.sapiens CpG island D 1.13e+03
	608	14	16.1	299	21	A48104	Sequence 171 from Pate 1.13e+03
c	609	14	16.1	324	24	MSU42753	Medicago sativa stress 1.13e+03
	610	14	16.1	331	30	G41428	Z928 Zebrafish AB Dani 1.13e+03
c	611	14	16.1	344	25	HUMARGDRP	Human arginase (ARG1) 1.13e+03
	612	14	16.1	360	18	AF019017	Drosophila melanogaste 1.13e+03
	613	14	16.1	369	30	G08100	human STS CHLC.GATA10C 1.13e+03
c	614	14	16.1	381	25	HSIGHPT11	H.sapiens gene for imm 1.13e+03
c	615	14	16.1	403	25	HSPA9F10	H.sapiens flow-sorted 1.13e+03
c	616	14	16.1	450	18	U91387	Plebeia frontalis Stin 1.13e+03
	617	14	16.1	499	23	AF032694	Oryza sativa NBS-LRR t 1.13e+03
	618	14	16.1	566	28	MMNKRP1AA	M.musculus NKR-P1 34 m 1.13e+03
c	619	14	16.1	572	28	MMSNRPN7	Mus musculus small nuc 1.13e+03
	620	14	16.1	607	15	RPZ82664	R.prowazekii genomic D 1.13e+03
c	621	14	16.1	607	24	AF027589	Hordeum marinum HMAR01 1.13e+03
	622	14	16.1	653	18	DBPTBIFM2	D.bifasciata P-transpo 1.13e+03
	623	14	16.1	663	23	LLU85322	Leucocoprinus longistr 1.13e+03
c	624	14	16.1	693	30	G39245	Z20343 Zebrafish AB Da 1.13e+03
	625	14	16.1	702	23	PCU07059	Pneumocystis carinii c 1.13e+03
	626	14	16.1	737	30	G321116	pig STS PTG82-532R, se 1.13e+03
c	627	14	16.1	771	26	HSSUR2G15	Homo sapiens sulfonylu 1.13e+03
c	628	14	16.1	781	33	AB013231	Hepatitis G virus geno 1.13e+03
	629	14	16.1	816	21	I12428	Sequence 1 from patent 1.13e+03
c	630	14	16.1	822	33	AF044710	Bluetongue virus strai 1.13e+03
c	631	14	16.1	822	33	AF044702	Bluetongue virus strai 1.13e+03
c	632	14	16.1	870	25	HSU34844	Human mercurial-insens 1.13e+03
	633	14	16.1	912	28	S43141	NK1.1 alloantigen=rece 1.13e+03
	634	14	16.1	999	20	DRAJ4863	Danio rerio mRNA for e 1.13e+03
c	635	14	16.1	1044	25	HSIGVKA10	Human germline immunog 1.13e+03
	636	14	16.1	1063	23	CSLCPMATK	Chrysosplenium lowense 1.13e+03
c	637	14	16.1	1082	18	D50345	Carabus vanvolxemi adu 1.13e+03
c	638	14	16.1	1083	18	AB013864	Damaster blaptoides vi 1.13e+03
c	639	14	16.1	1083	18	AB001510	Carabus leptoplesioide 1.13e+03
c	640	14	16.1	1083	18	AB013980	Damaster blaptoides bl 1.13e+03
c	641	14	16.1	1083	18	D50350	Damaster oxuroides mit 1.13e+03
c	642	14	16.1	1083	18	AB017463	Carabus hortensis mito 1.13e+03
c	643	14	16.1	1083	18	D50422	Damaster blaptoides mi 1.13e+03
c	644	14	16.1	1083	18	D50425	Damaster cyanostola mi 1.13e+03
c	645	14	16.1	1083	18	AB013863	Damaster blaptoides ru 1.13e+03
c	646	14	16.1	1083	18	AB013967	Damaster blaptoides bl 1.13e+03
c	647	14	16.1	1083	18	AB013937	Damaster blaptoides ba 1.13e+03
c	648	14	16.1	1083	18	AB013973	Damaster blaptoides br 1.13e+03
c	649	14	16.1	1083	18	AB013951	Damaster blaptoides ox 1.13e+03
c	650	14	16.1	1083	18	AB013942	Damaster blaptoides ox 1.13e+03
c	651	14	16.1	1083	18	AB013871	Damaster blaptoides vi 1.13e+03
c	652	14	16.1	1083	18	AB013955	Damaster blaptoides bl 1.13e+03
c	653	14	16.1	1083	18	AB013870	Damaster blaptoides vi 1.13e+03
c	654	14	16.1	1083	18	AB013977	Damaster blaptoides bl 1.13e+03

c	655	14	16.1	1083	18	AB013969	Damaster blaptoides bl 1.13e+03
c	656	14	16.1	1083	18	AB013971	Damaster blaptoides bl 1.13e+03
c	657	14	16.1	1083	18	AB013869	Damaster blaptoides vi 1.13e+03
c	658	14	16.1	1083	18	AB013939	Damaster blaptoides cy 1.13e+03
c	659	14	16.1	1083	18	D50427	Damaster oxuroides mit 1.13e+03
c	660	14	16.1	1083	18	AB013866	Damaster blaptoides vi 1.13e+03
c	661	14	16.1	1083	18	AB013962	Damaster blaptoides bl 1.13e+03
c	662	14	16.1	1083	18	AB013941	Damaster blaptoides ox 1.13e+03
c	663	14	16.1	1177	20	CHKP2ACS	Chicken mRNA for phosp 1.13e+03
	664	14	16.1	1250	24	AB003047	Chrysosplenium flagell 1.13e+03
c	665	14	16.1	1342	25	HSAT2PRO1	H.sapiens AT2 gene, pr 1.13e+03
c	666	14	16.1	1368	18	DROEMC2	D.melanogaster extrama 1.13e+03
c	667	14	16.1	1447	28	MMCARPMT4	Mus musculus 129SV car 1.13e+03
c	668	14	16.1	1450	25	HSIKLVA26	H.sapiens DNA for IG k 1.13e+03
	669	14	16.1	1473	23	YSCMTCYB1	Yeast (S.cerevisiae) m 1.13e+03
c	670	14	16.1	1522	23	SCYKR055W	S.cerevisiae chromosom 1.13e+03
	671	14	16.1	1561	20	AB012882	Cyprinus carpio mRNA f 1.13e+03
	672	14	16.1	1633	15	CLOETOXIN	Clostridium perfringen 1.13e+03
c	673	14	16.1	1658	28	MMSAGP	M.musculus S-antigen g 1.13e+03
	674	14	16.1	1722	23	YSTMTTGE	Saccharomyces cerevisi 1.13e+03
c	675	14	16.1	1890	23	SCNDI1G	S.cerevisiae NDI1 gene 1.13e+03
	676	14	16.1	1901	21	A21994	Vaccina virus SalI F f 1.13e+03
c	677	14	16.1	1926	18	CEU12762	Caenorhabditis elegans 1.13e+03
c	678	14	16.1	2021	23	HBATPB	H.brasiliensis mRNA fo 1.13e+03
	679	14	16.1	2030	23	S73870	pct1+=res1+/sct1+ homo 1.13e+03
	680	14	16.1	2048	23	YSCHSP150A	Saccharomyces cerevisi 1.13e+03
	681	14	16.1	2081	15	CLOADH	C.acetobutylicum NADPH 1.13e+03
c	682	14	16.1	2232	18	CET51A2C	Caenorhabditis elegans 1.13e+03
	683	14	16.1	2309	24	AF060485	Arabidopsis thaliana M 1.13e+03
c	684	14	16.1	2370	23	LELAP17PR	L.esculentum LAP17.2 g 1.13e+03
	685	14	16.1	2576	25	HSZESTENH	H.sapiens mRNA for enh 1.13e+03
	686	14	16.1	2600	27	HSU61145	Human enhancer of zest 1.13e+03
	687	14	16.1	2600	21	A58329	Sequence 1 from Patent 1.13e+03
	688	14	16.1	2666	27	HSU52965	Human putative transcr 1.13e+03
c	689	14	16.1	2754	33	CLGVP73	Cryptophlebia leucotre 1.13e+03
c	690	14	16.1	2786	23	SCYOR202C	S.cerevisiae chromosom 1.13e+03
c	691	14	16.1	2802	18	OMMIFILPRO	Ommastrephes sloani pa 1.13e+03
c	692	14	16.1	2838	22	BPTWGHOLG	Staphylococcus phage T 1.13e+03
	693	14	16.1	2871	25	HSU44861	Human Down Syndrome re 1.13e+03
c	694	14	16.1	2912	26	HSAC001155	Homo sapiens (subclone 1.13e+03
c	695	14	16.1	2928	25	HUM25RNASE	Human 2-5A-dependent R 1.13e+03
c	696	14	16.1	2936	23	KLRAG3A	K.lactis RAG3 gene. 1.13e+03
	697	14	16.1	3265	18	DROGPAMB	D.melanogaster G prote 1.13e+03
	698	14	16.1	3266	21	I17525	Sequence 3 from patent 1.13e+03
c	699	14	16.1	3280	25	HSU27478	Human angiotensin II t 1.13e+03
	700	14	16.1	3294	23	YSCGTR1	Saccharomyces cerevisi 1.13e+03
	701	14	16.1	3381	26	HSAC001161	Homo sapiens (subclone 1.13e+03
c	702	14	16.1	3463	25	HUMLBR10	Homo sapiens integral 1.13e+03
	703	14	16.1	3465	23	SPGCH1	S.pombe GCS1 gene. 1.13e+03
c	704	14	16.1	3467	15	SSOGPEB	Sulfolobus acidocaldar 1.13e+03
	705	14	16.1	3520	23	PMCMSGBA	Pneumocystis carinii m 1.13e+03

	706	14	16.1	3536	27	U00115	Human zinc-finger prot 1.13e+03
	707	14	16.1	3560	33	NILIZUMOD	Nilaparvata lugens reo 1.13e+03
	708	14	16.1	3591	25	HSLIFR	H.sapiens mRNA for leu 1.13e+03
	709	14	16.1	3720	21	I49950	Sequence 1 from patent 1.13e+03
c	710	14	16.1	4000	23	SCRHO4A	S.cerevisiae RHO4a and 1.13e+03
	711	14	16.1	4129	23	SCCBF2A	S.cerevisiae CBF2 gene 1.13e+03
c	712	14	16.1	4333	26	HUMORFKG1T	Human mRNA for KIAA006 1.13e+03
	713	14	16.1	4504	23	SCYGR140W	S.cerevisiae chromosom 1.13e+03
	714	14	16.1	4647	33	TVU08999	Trichomonas vaginalis 1.13e+03
	715	14	16.1	5087	20	CHKMAFBBZI	Chicken gene for bZip 1.13e+03
	716	14	16.1	5119	28	AF064984	Mus musculus low-densi 1.13e+03
c	717	14	16.1	5253	28	D83697	Rat brain mRNA for neu 1.13e+03
c	718	14	16.1	5297	25	HSU61262	Human neogenin mRNA, c 1.13e+03
	719	14	16.1	5320	27	HSIRF4	Human lymphocyte speci 1.13e+03
c	720	14	16.1	5414	28	MMMEOGEN	M.musculus mRNA for ne 1.13e+03
	721	14	16.1	5898	25	HSHGF	Human mRNA for hepatoc 1.13e+03
	722	14	16.1	6376	26	AB011106	Homo sapiens mRNA for 1.13e+03
c	723	14	16.1	6743	25	HSNADGLCT	H.sapiens Na+-D-glucos 1.13e+03
	724	14	16.1	6909	16	AF013583	Escherichia coli putat 1.13e+03
c	725	14	16.1	7038	28	RATACACA	Rat acetyl-coenzyme A 1.13e+03
c	726	14	16.1	7053	23	SCGTPAP	S.cerevisiae (S288C) I 1.13e+03
	727	14	16.1	7149	18	PFSC03031	Plasmodium falciparum 1.13e+03
c	728	14	16.1	7610	25	HUMTRPM2A1	Human TRPM-2 protein g 1.13e+03
c	729	14	16.1	8158	33	CVU20341	Cassava vein mosaic vi 1.13e+03
	730	14	16.1	9264	33	AF014388	Drosophila C virus str 1.13e+03
	731	14	16.1	9494	18	PFAGAMSPEC	Plasmodium falciparum 1.13e+03
	732	14	16.1	9950	28	MUSUPAA	Mouse Murine urokinase 1.13e+03
c	733	14	16.1	10206	16	AB017355	Streptococcus agalacti 1.13e+03
c	734	14	16.1	10614	28	CRUCH7AHYD	Cricetulus griseus cho 1.13e+03
	735	14	16.1	10901	18	PFSC03070	Plasmodium falciparum 1.13e+03
c	736	14	16.1	11834	15	AE001155	Borrelia burgdorferi (1.13e+03
c	737	14	16.1	12343	15	ECAE000298	Escherichia coli K-12 1.13e+03
	738	14	16.1	12461	23	SC7056	S.cerevisiae chromosom 1.13e+03
c	739	14	16.1	12794	17	CE4R79	Caenorhabditis elegans 1.13e+03
	740	14	16.1	13251	18	CELEGAP9	Caenorhabditis elegans 1.13e+03
	741	14	16.1	13936	24	AB014344	Oryza rufipogon RNA, e 1.13e+03
c	742	14	16.1	14893	23	SCYGR038W	S.cerevisiae chromosom 1.13e+03
c	743	14	16.1	15569	15	AE000733	Aquifex aeolicus secti 1.13e+03
c	744	14	16.1	16556	18	PTU19464	Paramecium tetraurelia 1.13e+03
	745	14	16.1	16624	20	POU62532	Polypterus ornatipinni 1.13e+03
	746	14	16.1	16646	23	CUCACC1	Zucchini 1-aminocyclop 1.13e+03
	747	14	16.1	16840	20	XMU17896	Xiphophorus maculatus 1.13e+03
c	748	14	16.1	16966	23	SCYGR161C	S.cerevisiae chromosom 1.13e+03
c	749	14	16.1	20089	23	SC8142B	S.cerevisiae chromosom 1.13e+03
	750	14	16.1	20518	18	PFSC03030	Plasmodium falciparum 1.13e+03
c	751	14	16.1	21202	18	PFSC04088	Plasmodium falciparum 1.13e+03
	752	14	16.1	23000	23	SPBC20F10	S.pombe chromosome II 1.13e+03
	753	14	16.1	23223	23	YSCL9233	Saccharomyces cerevisi 1.13e+03
	754	14	16.1	23536	18	CEF57B7	Caenorhabditis elegans 1.13e+03
	755	14	16.1	24281	18	CEC06H2	Caenorhabditis elegans 1.13e+03
	756	14	16.1	24667	28	AF001295	Homo sapiens tyrosinas 1.13e+03

	757	14	16.1	24784	18	CELF29B9	Caenorhabditis elegans 1.13e+03
c	758	14	16.1	26209	18	CEY39E4A	Caenorhabditis elegans 1.13e+03
c	759	14	16.1	27451	18	CEP53F8	Caenorhabditis elegans 1.13e+03
c	760	14	16.1	27874	28	MMU71085	Mus musculus insulin-1 1.13e+03
c	761	14	16.1	28142	25	HSL58B6	Human DNA sequence fro 1.13e+03
	762	14	16.1	30320	18	CEC05B5	Caenorhabditis elegans 1.13e+03
c	763	14	16.1	30850	18	CEK02B9	Caenorhabditis elegans 1.13e+03
	764	14	16.1	30953	23	SPAC4D7	S.pombe chromosome I c 1.13e+03
c	765	14	16.1	31197	17	CEBA2	*** SEQUENCING IN PROG 1.13e+03
	766	14	16.1	32043	18	CEC06C3	Caenorhabditis elegans 1.13e+03
c	767	14	16.1	32329	23	SPBC3D6	S.pombe chromosome II 1.13e+03
c	768	14	16.1	33183	18	CELM01E11	Caenorhabditis elegans 1.13e+03
	769	14	16.1	33206	25	HUMBFXIII	Human factor XIII b su 1.13e+03
	770	14	16.1	33429	18	CEF09F3	Caenorhabditis elegans 1.13e+03
c	771	14	16.1	33450	17	CEH27I04	Caenorhabditis elegans 1.13e+03
	772	14	16.1	33466	18	CEEC47	Caenorhabditis elegans 1.13e+03
c	773	14	16.1	34509	18	CEP57G12	Caenorhabditis elegans 1.13e+03
	774	14	16.1	35234	25	HSUB2213	Human chromosome 10 cl 1.13e+03
	775	14	16.1	35473	25	HSL30G1	Human DNA sequence fro 1.13e+03
	776	14	16.1	35654	18	CER05D7	Caenorhabditis elegans 1.13e+03
c	777	14	16.1	35677	26	AC005361	Homo sapiens chromosom 1.13e+03
	778	14	16.1	35820	18	CELT21E12	Caenorhabditis elegans 1.13e+03
	779	14	16.1	36000	18	CEEK632	Caenorhabditis elegans 1.13e+03
	780	14	16.1	36170	18	CER10E4	Caenorhabditis elegans 1.13e+03
c	781	14	16.1	36470	15	G00014	Mycobacterium leprae c 1.13e+03
	782	14	16.1	36471	26	DR7021	Human (lambda) DNA for 1.13e+03
c	783	14	16.1	36552	18	CELF26A1	Caenorhabditis elegans 1.13e+03
c	784	14	16.1	36558	25	HSU73645	Human chromosome 11 46 1.13e+03
	785	14	16.1	36922	18	PFSC03026	Plasmodium falciparum 1.13e+03
c	786	14	16.1	37013	24	SPAC3A12	S.pombe chromosome I c 1.13e+03
	787	14	16.1	37121	18	CELT19B4	Caenorhabditis elegans 1.13e+03
	788	14	16.1	37248	18	CEF46F2	Caenorhabditis elegans 1.13e+03
c	789	14	16.1	37321	27	AC005498	Homo sapiens chromosom 1.13e+03
	790	14	16.1	37652	25	HSU85A3	Human DNA sequence fro 1.13e+03
	791	14	16.1	37750	18	CEEC504	Caenorhabditis elegans 1.13e+03
	792	14	16.1	38103	18	CELR09A1	Caenorhabditis elegans 1.13e+03
c	793	14	16.1	38190	17	AC004193	*** SEQUENCING IN PROG 1.13e+03
	794	14	16.1	38345	25	HSU81830	Human cosmid LL12NC01- 1.13e+03
	795	14	16.1	38374	26	HSU95090	Homo sapiens chromosom 1.13e+03
	796	14	16.1	38886	18	CELF46H5	Caenorhabditis elegans 1.13e+03
	797	14	16.1	39174	23	YSCD9798	Saccharomyces cerevisi 1.13e+03
	798	14	16.1	39240	18	AF052831	Trypanosoma cruzi CL B 1.13e+03
c	799	14	16.1	39269	26	AC004171	Homo sapiens clone UWG 1.13e+03
c	800	14	16.1	39283	18	CET04A11	Caenorhabditis elegans 1.13e+03
	801	14	16.1	39496	18	CELEEED8	Caenorhabditis elegans 1.13e+03
c	802	14	16.1	40251	17	AB018108	*** SEQUENCING IN PROG 1.13e+03
c	803	14	16.1	40594	18	CER03A10	Caenorhabditis elegans 1.13e+03
	804	14	16.1	40596	18	CELC36C5	Caenorhabditis elegans 1.13e+03
	805	14	16.1	40753	17	AC005715	*** SEQUENCING IN PROG 1.13e+03
	806	14	16.1	40881	26	AC004679	Homo sapiens chromosom 1.13e+03
	807	14	16.1	41122	26	HSAC000109	Human Cosmid g0771a222 1.13e+03

	808	14	16.1	41494	18	CEF22D6	Caenorhabditis elegans 1.13e+03
	809	14	16.1	41495	26	AC004224	Homo sapiens chromosom 1.13e+03
	810	14	16.1	41752	18	CELC15H9	Caenorhabditis elegans 1.13e+03
c	811	14	16.1	42393	27	AC005780	Homo sapiens chromosom 1.13e+03
c	812	14	16.1	43105	26	AC002988	Human DNA from chromos 1.13e+03
	813	14	16.1	43116	17	AC005710	*** SEQUENCING IN PROG 1.13e+03
	814	14	16.1	43310	26	AC004597	Homo sapiens chromosom 1.13e+03
c	815	14	16.1	43788	18	CEC10C6	Caenorhabditis elegans 1.13e+03
	816	14	16.1	44957	26	AC005372	Homo sapiens chromosom 1.13e+03
c	817	14	16.1	48338	18	CET02D9	Caenorhabditis elegans 1.13e+03
	818	14	16.1	49457	17	CEC47A4	Caenorhabditis elegans 1.13e+03
c	819	14	16.1	52283	33	CVGRI90	Cowpox virus strain GR 1.13e+03
c	820	14	16.1	56583	26	HS347M6	Human DNA sequence fro 1.13e+03
	821	14	16.1	57031	23	AB012244	Arabidopsis thaliana g 1.13e+03
	822	14	16.1	60023	26	AC000386	Human Chromosome 15 pa 1.13e+03
c	823	14	16.1	60050	18	AC002441	Drosophila melanogaste 1.13e+03
c	824	14	16.1	63155	20	AP026198	Fugu rubripes neural c 1.13e+03
c	825	14	16.1	68889	23	AB010694	Arabidopsis thaliana g 1.13e+03
c	826	14	16.1	69822	18	AC003121	Drosophila melanogaste 1.13e+03
	827	14	16.1	71618	23	AB012240	Arabidopsis thaliana g 1.13e+03
	828	14	16.1	72245	26	HS93C23	Human DNA sequence fro 1.13e+03
	829	14	16.1	73303	17	AC005526	*** SEQUENCING IN PROG 1.13e+03
	830	14	16.1	75551	18	AC005111	Drosophila melanogaste 1.13e+03
c	831	14	16.1	77380	23	AB010077	Arabidopsis thaliana g 1.13e+03
	832	14	16.1	79323	18	AC002442	Drosophila melanogaste 1.13e+03
	833	14	16.1	79793	17	AC004834	*** SEQUENCING IN PROG 1.13e+03
	834	14	16.1	79976	23	AB008264	Arabidopsis thaliana g 1.13e+03
	835	14	16.1	80101	18	AC004115	Drosophila melanogaste 1.13e+03
c	836	14	16.1	81421	17	CEC53C7	Caenorhabditis elegans 1.13e+03
	837	14	16.1	81522	18	AC004298	Drosophila melanogaste 1.13e+03
	838	14	16.1	82073	26	HS172K10	Homo sapiens DNA seque 1.13e+03
	839	14	16.1	84132	26	AC005348	Homo sapiens chromosom 1.13e+03
c	840	14	16.1	85139	18	AC001657	Drosophila melanogaste 1.13e+03
c	841	14	16.1	85214	24	ATAC0053396	Arabidopsis thaliana c 1.13e+03
	842	14	16.1	85458	17	HS917N8	Human DNA sequence *** 1.13e+03
	843	14	16.1	85567	24	AB016878	Arabidopsis thaliana g 1.13e+03
	844	14	16.1	85661	24	AB013394	Arabidopsis thaliana g 1.13e+03
c	845	14	16.1	86065	23	AB005243	Arabidopsis thaliana g 1.13e+03
	846	14	16.1	87630	24	AB016889	Arabidopsis thaliana g 1.13e+03
	847	14	16.1	87978	26	AC003971	Human BAC 367D17 from 1.13e+03
c	848	14	16.1	88987	26	AC004541	Homo sapiens PAC clone 1.13e+03
	849	14	16.1	89603	27	HSU63834	Human KIT protein and 1.13e+03
c	850	14	16.1	90082	26	AF045555	Homo sapiens wbscr1 (W 1.13e+03
	851	14	16.1	93911	17	HSAC000002	*** SEQUENCING IN PROG 1.13e+03
c	852	14	16.1	94343	26	AC005255	Homo sapiens chromosom 1.13e+03
	853	14	16.1	95417	24	ATT16H5	Arabidopsis thaliana D 1.13e+03
c	854	14	16.1	96372	17	AC004713	*** SEQUENCING IN PROG 1.13e+03
c	855	14	16.1	96574	23	ATT18B16	Arabidopsis thaliana D 1.13e+03
c	856	14	16.1	98178	17	AC005300	*** SEQUENCING IN PROG 1.13e+03
c	857	14	16.1	98387	26	AC003077	Human BAC clone GS034D 1.13e+03
	858	14	16.1	98957	24	ATAC004136	Arabidopsis thaliana c 1.13e+03

c	859	14	16.1	100639	17	AC002417	*** SEQUENCING IN PROG 1.13e+03
	860	14	16.1	100753	17	AC003102	*** SEQUENCING IN PROG 1.13e+03
	861	14	16.1	102431	26	AC005213	Homo sapiens chromosom 1.13e+03
c	862	14	16.1	102742	17	HS184J9	Human DNA sequence *** 1.13e+03
c	863	14	16.1	103566	26	AC003087	Human BAC clone RG119H 1.13e+03
	864	14	16.1	104609	17	AC005683	*** SEQUENCING IN PROG 1.13e+03
	865	14	16.1	106039	26	AC004159	Homo sapiens PAC clone 1.13e+03
c	866	14	16.1	107889	26	AC004519	Homo sapiens BAC clone 1.13e+03
c	867	14	16.1	107897	26	AP000057	Homo sapiens genomic D 1.13e+03
c	868	14	16.1	108260	26	HS179P9	Human DNA sequence fro 1.13e+03
c	869	14	16.1	108390	17	HS312L17	Human DNA sequence *** 1.13e+03
	870	14	16.1	109155	26	HS106B9	Human DNA sequence fro 1.13e+03
	871	14	16.1	110000	17	CEY102A5	Caenorhabditis elegans. 1.13e+03
	872	14	16.1	110000	17	CEY102G3	Caenorhabditis elegans 1.13e+03
c	873	14	16.1	110000	17	CEY39E4	Caenorhabditis elegans 1.13e+03
c	874	14	16.1	110000	17	HSY73BF9	Human DNA sequence *** 1.13e+03
	875	14	16.1	110608	17	HS283E3	Human DNA sequence *** 1.13e+03
	876	14	16.1	111062	17	HSAC001237	*** SEQUENCING IN PROG 1.13e+03
	877	14	16.1	111118	25	HSAC000112	Human PAC clone DJ149P 1.13e+03
c	878	14	16.1	112018	17	HS128O17	Human DNA sequence *** 1.13e+03
c	879	14	16.1	113241	26	HUAC002425	Homo sapiens Chromosom 1.13e+03
c	880	14	16.1	114617	23	ATP13C5	Arabidopsis thaliana D 1.13e+03
c	881	14	16.1	114825	17	HS671O14	Human DNA sequence *** 1.13e+03
	882	14	16.1	115359	24	ATAC003028	Arabidopsis thaliana c 1.13e+03
	883	14	16.1	115602	17	HS111BD24	Human DNA sequence *** 1.13e+03
c	884	14	16.1	115743	27	HS468H17	Human DNA sequence fro 1.13e+03
c	885	14	16.1	115851	23	ATAC002505	Arabidopsis thaliana c 1.13e+03
	886	14	16.1	116271	17	HS394I7	Human DNA sequence *** 1.13e+03
	887	14	16.1	116884	17	PFMAL1P1	Plasmodium falciparum 1.13e+03
c	888	14	16.1	117215	17	AC003950	*** SEQUENCING IN PROG 1.13e+03
	889	14	16.1	117911	26	HS117P19	Human DNA sequence fro 1.13e+03
c	890	14	16.1	118086	23	ATF6H11	Arabidopsis thaliana D 1.13e+03
	891	14	16.1	118591	17	AC005024	*** SEQUENCING IN PROG 1.13e+03
	892	14	16.1	119453	26	AC004006	Human PAC clone DJ0740 1.13e+03
c	893	14	16.1	120375	17	AC004442	*** SEQUENCING IN PROG 1.13e+03
c	894	14	16.1	120955	27	HUAC002310	Human Chromosome 16 BA 1.13e+03
c	895	14	16.1	122000	17	AC004050	*** SEQUENCING IN PROG 1.13e+03
c	896	14	16.1	122322	26	AC002448	Human PAC clone DJ0068 1.13e+03
c	897	14	16.1	123336	26	AC003075	Human PAC clone DJ0658 1.13e+03
	898	14	16.1	123360	24	ATAC003040	Arabidopsis thaliana c 1.13e+03
c	899	14	16.1	123891	26	AC004549	Homo sapiens BAC clone 1.13e+03
	900	14	16.1	124010	26	HS340G1	Human DNA sequence fro 1.13e+03
c	901	14	16.1	125040	17	AC004922	*** SEQUENCING IN PROG 1.13e+03
c	902	14	16.1	125469	15	SYCSLRA	Synechocystis sp. PCC6 1.13e+03
c	903	14	16.1	125590	26	AC002382	Human BAC clone RG022J 1.13e+03
	904	14	16.1	125930	17	HS124E22	Human DNA sequence *** 1.13e+03
c	905	14	16.1	129722	26	HS15O5	Homo sapiens DNA seque 1.13e+03
c	906	14	16.1	130223	26	AF064863	Homo sapiens chromosom 1.13e+03
	907	14	16.1	131757	26	AC004540	Homo sapiens PAC clone 1.13e+03
	908	14	16.1	131929	17	HS239K6	Human DNA sequence *** 1.13e+03
	909	14	16.1	133469	17	HSAC000383	*** SEQUENCING IN PROG 1.13e+03

	910	14	16.1	133798	17	AC004950	*** SEQUENCING IN PROG 1.13e+03
c	911	14	16.1	134199	15	SYCSLRF	Synechocystis sp. PCC6 1.13e+03
	912	14	16.1	134845	26	HS863K19	Human DNA sequence fro 1.13e+03
c	913	14	16.1	134929	27	AC004655	Homo sapiens Xp22-140- 1.13e+03
	914	14	16.1	135648	26	AC004069	Homo sapiens chromosom 1.13e+03
c	915	14	16.1	135648	26	AC004069	Homo sapiens chromosom 1.13e+03
	916	14	16.1	136335	24	AC005275	Arabidopsis thaliana B 1.13e+03
	917	14	16.1	136823	26	AC003083	Human BAC clone RG060N 1.13e+03
c	918	14	16.1	137836	17	AC004837	*** SEQUENCING IN PROG 1.13e+03
c	919	14	16.1	138706	17	AC004946	*** SEQUENCING IN PROG 1.13e+03
c	920	14	16.1	139259	18	CEY51A2D	Caenorhabditis elegans 1.13e+03
	921	14	16.1	139958	27	HUAC002519	Human Chromosome 16 BA 1.13e+03
	922	14	16.1	141132	17	HS192P9	Human DNA sequence *** 1.13e+03
c	923	14	16.1	141589	26	HS117P20	Homo sapiens DNA seque 1.13e+03
c	924	14	16.1	141878	26	AC000004	Genomic sequence from 1.13e+03
	925	14	16.1	142640	17	AC004936	*** SEQUENCING IN PROG 1.13e+03
c	926	14	16.1	143029	26	AC003074	Human PAC clone DJ0596 1.13e+03
	927	14	16.1	143092	18	CEY17G7B	Caenorhabditis elegans 1.13e+03
	928	14	16.1	143427	26	AC002432	Human BAC clone RG317G 1.13e+03
	929	14	16.1	143819	26	AC002994	Homo sapiens chromosom 1.13e+03
c	930	14	16.1	144227	26	HS119E23	Homo sapiens DNA seque 1.13e+03
	931	14	16.1	145721	17	HS112L6	Human DNA sequence *** 1.13e+03
c	932	14	16.1	148883	25	HS358H7	Human DNA sequence fro 1.13e+03
	933	14	16.1	149409	26	AC004081	Homo sapiens PAC clone 1.13e+03
c	934	14	16.1	149698	26	HS76C18	Human DNA sequence fro 1.13e+03
c	935	14	16.1	150000	17	AC004704	*** SEQUENCING IN PROG 1.13e+03
	936	14	16.1	150221	26	AC004593	Homo sapiens PAC clone 1.13e+03
	937	14	16.1	152107	17	HS811H13	Human DNA sequence *** 1.13e+03
	938	14	16.1	152392	17	HS1018D12	Human DNA sequence *** 1.13e+03
c	939	14	16.1	152592	17	HS396D17	Human DNA sequence *** 1.13e+03
c	940	14	16.1	153356	17	AC004995	*** SEQUENCING IN PROG 1.13e+03
c	941	14	16.1	153687	17	HS441J1	Human DNA sequence *** 1.13e+03
	942	14	16.1	154152	17	AC005057	*** SEQUENCING IN PROG 1.13e+03
	943	14	16.1	154619	16	D90917	Synechocystis sp. PCC6 1.13e+03
c	944	14	16.1	154821	17	AC005341	*** SEQUENCING IN PROG 1.13e+03
	945	14	16.1	154942	17	AC005227	*** SEQUENCING IN PROG 1.13e+03
	946	14	16.1	155529	17	HS738P11	Human DNA sequence *** 1.13e+03
	947	14	16.1	156399	26	AC002456	Human BAC clone RG013L 1.13e+03
	948	14	16.1	156994	17	HS1018K9	Human DNA sequence *** 1.13e+03
c	949	14	16.1	157248	26	AC002312	Human Chromosome 15 pa 1.13e+03
c	950	14	16.1	157838	26	HUAC004626	Homo sapiens Chromosom 1.13e+03
	951	14	16.1	159479	17	AC005535	*** SEQUENCING IN PROG 1.13e+03
	952	14	16.1	160378	26	HS93H18	Human DNA sequence fro 1.13e+03
	953	14	16.1	162524	17	HS498I24	Human DNA sequence *** 1.13e+03
	954	14	16.1	163730	17	AC004852	*** SEQUENCING IN PROG 1.13e+03
	955	14	16.1	163889	17	AC004958	*** SEQUENCING IN PROG 1.13e+03
	956	14	16.1	164536	17	HS134N8	Human DNA sequence *** 1.13e+03
	957	14	16.1	165491	26	AC000385	Human Chromosome 11 pa 1.13e+03
	958	14	16.1	165590	17	HS782D21	Human DNA sequence *** 1.13e+03
c	959	14	16.1	166126	26	AC004142	Homo sapiens BAC clone 1.13e+03
c	960	14	16.1	168056	26	AC004458	Homo sapiens BAC clone 1.13e+03

c	961	14	16.1	171905	26	HSAC002080	Human BAC clone RG253B	1.13e+03
c	962	14	16.1	172106	17	CEY67D11	Caenorhabditis elegans	1.13e+03
c	963	14	16.1	174311	18	AC004758	Drosophila melanogaste	1.13e+03
	964	14	16.1	175120	26	AC004687	Homo sapiens chromosom	1.13e+03
c	965	14	16.1	175339	17	AC005772	*** SEQUENCING IN PROG	1.13e+03
c	966	14	16.1	175859	26	AC004491	Homo sapiens BAC clone	1.13e+03
	967	14	16.1	177455	17	AC004980	*** SEQUENCING IN PROG	1.13e+03
	968	14	16.1	179861	17	CEY65A5	Caenorhabditis elegans	1.13e+03
c	969	14	16.1	181331	17	AC005230	*** SEQUENCING IN PROG	1.13e+03
c	970	14	16.1	181925	24	YUP8H12R	Arabidopsis thaliana c	1.13e+03
c	971	14	16.1	182123	17	AC002399	*** SEQUENCING IN PROG	1.13e+03
c	972	14	16.1	183909	17	AC005528	*** SEQUENCING IN PROG	1.13e+03
c	973	14	16.1	185994	17	AC002042	*** SEQUENCING IN PROG	1.13e+03
	974	14	16.1	187051	17	AC004688	*** SEQUENCING IN PROG	1.13e+03
	975	14	16.1	188682	17	AC005050	*** SEQUENCING IN PROG	1.13e+03
	976	14	16.1	189476	17	AC004976	*** SEQUENCING IN PROG	1.13e+03
	977	14	16.1	190665	17	HS495010	Human DNA sequence ***	1.13e+03
	978	14	16.1	191737	33	VACCG	Vaccinia virus, comple	1.13e+03
c	979	14	16.1	192335	17	AC005042	*** SEQUENCING IN PROG	1.13e+03
c	980	14	16.1	195142	26	AC004554	Homo sapiens Xp22 BAC	1.13e+03
	981	14	16.1	197391	17	AC005140	*** SEQUENCING IN PROG	1.13e+03
c	982	14	16.1	199500	17	AC005384	*** SEQUENCING IN PROG	1.13e+03
	983	14	16.1	200134	17	AC004806	*** SEQUENCING IN PROG	1.13e+03
c	984	14	16.1	206419	23	ATAP21	Arabidopsis thaliana D	1.13e+03
c	985	14	16.1	210954	26	AC003684	Homo sapiens Xp22 BAC	1.13e+03
c	986	14	16.1	219447	25	HUMPLNG6PD	Homo sapiens chromosom	1.13e+03
	987	14	16.1	220000	17	AC005798	*** SEQUENCING IN PROG	1.13e+03
c	988	14	16.1	226916	27	AC005066	Homo sapiens BAC clone	1.13e+03
	989	14	16.1	234542	27	HUAC002041	Human Chromosome 16 BA	1.13e+03
	990	14	16.1	238438	17	AC005046	*** SEQUENCING IN PROG	1.13e+03
	991	14	16.1	244362	17	HS618F1	Human DNA sequence ***	1.13e+03
	992	14	16.1	250529	26	HUAE000658	Homo sapiens T-cell re	1.13e+03
c	993	14	16.1	263113	17	CEY48C3	Caenorhabditis elegans	1.13e+03
c	994	14	16.1	267156	26	U66059	Human germline T-cell	1.13e+03
	995	14	16.1	291015	17	CEY59A8	Caenorhabditis elegans	1.13e+03
c	996	14	16.1	292720	17	CEY47H10	Caenorhabditis elegans	1.13e+03
c	997	14	16.1	292828	17	AC004801	*** SEQUENCING IN PROG	1.13e+03
	998	14	16.1	328808	17	CEY44A6	Caenorhabditis elegans	1.13e+03
c	999	14	16.1	328808	17	CEY44A6	Caenorhabditis elegans	1.13e+03
	1000	14	16.1	330836	17	HS780M13	Human DNA sequence ***	1.13e+03

ALIGNMENTS

```
RESULT    1
LOCUS       G05576          307 bp    DNA             STS       04-JUN-1996
DEFINITION  human STS WI-8386.
ACCESSION   G05576
NID         g858821
KEYWORDS    STS sequence; primer; sequence tagged site.
SOURCE      human STSs derived from sequences in dBEST and the Unigene
```



```
            collection.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 307)
  AUTHORS   Hudson,T.
  TITLE     Whitehead Institute/MIT Center for Genome Research; Physically
            Mapped ESTs
  JOURNAL   Unpublished (1995)
REFERENCE   2  (bases 1 to 307)
  AUTHORS   Hudson,T.
  TITLE     Whitehead Institute/MIT Center for Genome Research; Physically
            Mapped STSs
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Thomas Hudson
            Whitehead Institute/MIT Center for Genome Research
            Whitehead Institute for Biomedical Research
            9 Cambridge Center, Cambridge MA 02142 USA
            Tel: 617 252 1900
            Fax: 617 252 1902
            Email: thudson@genome.wi.mit.edu

            Primer A: TTTTCCTTTCTAAAACCCTCACC
            Primer B: CCTCTGCTGACACCTTGGAT
            STS size: 257
            PCR Profile:
                  Presoak:
                  Denaturation:
                  Annealing: 56 degrees C
                  Polymerization:
                  PCR Cycles: 35
                  Thermal Cycler:
            Protocol:
                Template: 10 ng
                Primer: each 5 pM
                dNTPs: each 4 nM
                Taq Polymerase: 0.025 units/ul
                Total Vol: 20 ul

            Buffer:
                MgCl2: 1.5 mM
                KCl: 50 mM
                Tris-HCL: 10 mM
                pH: 9.3

            Prepared with primer pairs derived from T57478 -- dbEST.
FEATURES            Location/Qualifiers
     source         1..307
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
```



```
                    /map="710_G_9; 772_H_1; 882_B_3; 914_H_6; 920_C_7;
                    939_B_6; 676.8 cR from top of Chr1 linkage group"
     STS            41. .297
     primer_bind    41. .63
     primer_bind    complement(278. .297)
BASE COUNT      102 a     75 c     45 g     84 t      1 others
ORIGIN

  Query Match           87.4%;  Score 76;  DB 30;  Length 307;
  Best Local Similarity  98.7%;  Pred. No. 7.13e-50;
  Matches    76;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db      1 AGATAGATNTTATTAAGACAAATTACACATTTTACTATATTTTTCCTTTCTAAAACCCTC 60
          ||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     77 AGATAGATTTTATTAAGACAAATTACACATTTTACTATATTTTTCCTTTCTAAAACCCTC 18

Db     61 ACCATCAAAGATTGATC 77
          |||||||||||||||||
Cp     17 ACCATCAAAGATTGATC 1


RESULT    2
LOCUS       T25C13     120067 bp    DNA             PLN       15-SEP-1998
DEFINITION  Arabidopsis thaliana BAC T25C13.
ACCESSION   AF080121
NID         g3600062
KEYWORDS    .
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Charophyta/Embryophyta group;
            Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta;
            Magnoliophyta; eudicotyledons; Rosidae; Capparales; Brassicaceae;
            Arabidopsis.
REFERENCE   1  (bases 1 to 120067)
  AUTHORS   Washington University Genome Sequencing Center.
  TITLE     The A. thaliana Genome Sequencing Project
  JOURNAL   Unpublished (1997)
REFERENCE   2  (bases 1 to 120067)
  AUTHORS   Ozersky,P.
  TITLE     The sequence of A. thaliana T25C13
  JOURNAL   Unpublished (1998)
REFERENCE   3  (bases 1 to 120067)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUL-1998) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     Submitted by:
                  Genome Sequencing Center
                  Department of Genetics, Washington University,
                  St. Louis, MO 63108, USA
                  e-mail: rwilson@watson.wustl.edu
```



```
            MAPPING: Clones were assigned to the YAC map by hybridization by
            M. Lodhi, Cold Spring Harbor Laboratories, and fingerprinted
            by M. Marra, WashU, to pick the best candidates for sequencing.


            NOTICE:  This sequence may not be the entire insert of this clone.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we provide a small overlap between
            neighboring submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded or sequenced with an alternate
            chemistry; an attempt was made to resolve all sequencing problems,
            such as compressions and repeats; all regions were covered by
            sequence from more than one subclone

                        NEIGHBORING COSMID INFORMATION:

            The 5' clone is F7N22; 3' clone is T26D22, 200 bp overlap.  Actual
            end is at 6603 of T26D22.

            The assembly of the tandem region from base 48777 to 73335 cannot
            be guaranteed. The region contains sequence which is single
            stranded, and of low quality. Restriction digest with HINDIII shows
            a 25 kb band, which is consistent with the submitted sequence.
FEATURES            Location/Qualifiers
     source         1. .120067
                    /organism="Arabidopsis thaliana"
                    /cultivar="Columbia"
                    /db_xref="taxon:3702"
                    /clone="T25C13"
                    /chromosome="V"
                    /map="unknown"
BASE COUNT    38749 a   20572 c   20644 g   40102 t
ORIGIN

  Query Match           24.1%;  Score 21;  DB 24;  Length 120067;
  Best Local Similarity 100.0%;  Pred. No. 4.12e-02;
  Matches    21;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  25237 AGAAAGGAAAAATATAGTAAA 25257
          |||||||||||||||||||||
Qy     27 AGAAAGGAAAAATATAGTAAA 47


RESULT    3
LOCUS       AP000027    48690 bp    DNA             PRI       13-MAY-1998
```

```
DEFINITION  Homo sapiens genomic DNA, chromosome 21q11.1, segment 5/5, complete
            sequence.
ACCESSION   AP000027
NID         g3132337
KEYWORDS    HTG.
SOURCE      Homo sapiens DNA, clone:72018t2f_1.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 48690)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Shiba,T. and Sakaki,Y.
  TITLE     Homo sapiens genomic DNA, chromosome 21q
  JOURNAL   Published Only in DataBase (1998) In press
REFERENCE   2  (bases 1 to 48690)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Shiba,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-MAY-1998) to the DDBJ/EMBL/GenBank databases.
            Masahira Hattori, Kitasato University, Department of Science, JST
            Sequencing Laboratory; Kitasato 1-15-1, Sagamihara 228, Japan
            (E-mail:hattori@hgc.ims.u-tokyo.ac.jp, Tel:0427-78-9732,
            Fax:0427-78-9561)
FEATURES             Location/Qualifiers
     source          1. .48690
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /clone="72018t2f_1"
                     /map="21q11.1"
BASE COUNT    15983 a   8815 c   8835 g  15057 t
ORIGIN

  Query Match           21.8%;  Score 19;  DB 26;  Length 48690;
  Best Local Similarity 100.0%;  Pred. No. 9.68e-01;
  Matches    19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  16922 AATATAGTAAAATGTGTAA 16940
          ||||||||||||||||||
Qy     37 AATATAGTAAAATGTGTAA 55


RESULT    4
LOCUS       AC005727   121567 bp    DNA             HTG       30-SEP-1998
DEFINITION  *** SEQUENCING IN PROGRESS *** Arabidopsis thaliana 'IGF' BAC
            'F8N16' genomic sequence near marker 'mi54'; HTGS phase 1, 20
            unordered pieces.
ACCESSION   AC005727
NID         g3687220
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Charophyta/Embryophyta group;
```



```
            Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta;
            Magnoliophyta; eudicotyledons; Rosidae; Capparales; Brassicaceae;
            Arabidopsis.
REFERENCE   1  (bases 1 to 121567)
  AUTHORS   Rounsley,S.D., Lin,X., Ketchum,K.A., Crosby,M.L., Brandon,R.C.,
            Mason,T.M., Kerlavage,A.R., Adams,M.D., Somerville,C.R. and
            Venter,J.C.
  TITLE     Arabidopsis thaliana 'IGF' BAC 'F8N16' genomic sequence near marker
            'mi54'
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 121567)
  AUTHORS   Rounsley,S.D. and Lin,X.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-SEP-1998) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, rounsley@tigr.org
COMMENT     ***                                                          ***
            *** WARNING: Phase 1 High Throughput Genome Sequence         ***
            ***                                                          ***
            * This sequence is unfinished. It consists of 20 contigs for
            * which the order is not known; their order in this record is
            * arbitrary. In some cases, the exact lengths of the gaps
            * between the contigs are also unknown; these gaps are  presented
            * as runs of N as a convenience only. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
            *      1     3549: contig of 3549 bp in length
            *   3550     3599: gap of unknown length
            *   3600     5677: contig of 2078 bp in length
            *   5678     5727: gap of unknown length
            *   5728    10365: contig of 4638 bp in length
            *  10366    10415: gap of unknown length
            *  10416    14129: contig of 3714 bp in length
            *  14130    14179: gap of unknown length
            *  14180    22523: contig of 8344 bp in length
            *  22524    22573: gap of unknown length
            *  22574    28785: contig of 6212 bp in length
            *  28786    28835: gap of unknown length
            *  28836    31353: contig of 2518 bp in length
            *  31354    31403: gap of unknown length
            *  31404    38837: contig of 7434 bp in length
            *  38838    38887: gap of unknown length
            *  38888    42411: contig of 3524 bp in length
            *  42412    42461: gap of unknown length
            *  42462    46530: contig of 4069 bp in length
            *  46531    46580: gap of unknown length
            *  46581    50163: contig of 3583 bp in length
            *  50164    50213: gap of unknown length
            *  50214    59373: contig of 9160 bp in length
            *  59374    59423: gap of unknown length
            *  59424    61779: contig of 2356 bp in length
            *  61780    61829: gap of unknown length
```



```
            *  61830    67008: contig of 5179 bp in length
            *  67009    67058: gap of unknown length
            *  67059    79413: contig of 12355 bp in length
            *  79414    79463: gap of unknown length
            *  79464    97979: contig of 18516 bp in length
            *  97980    98029: gap of unknown length
            *  98030   104243: contig of 6214 bp in length
            *  104244   104293: gap of unknown length
            *  104294   107411: contig of 3118 bp in length
            *  107412   107461: gap of unknown length
            *  107462   115768: contig of 8307 bp in length
            *  115769   115818: gap of unknown length
            *  115819   121567: contig of 5749 bp in length.
FEATURES             Location/Qualifiers
     source          1. .121567
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /clone="F8N16"
BASE COUNT    39064 a  20921 c  20933 g  39692 t    957 others
ORIGIN

  Query Match           21.8%;  Score 19;  DB 17;  Length 121567;
  Best Local Similarity 100.0%;  Pred. No. 9.68e-01;
  Matches    19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  76766 TTACACATTTTACTATATT 76784
          ||||||||||||||||||
Cp     55 TTACACATTTTACTATATT 37


RESULT    5
LOCUS       HS12803    124515 bp    DNA             HTG       02-OCT-1997
DEFINITION  Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone 12803;
            HTGS phase 1.
ACCESSION   Z98742
NID         g2465029
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 124515)
  AUTHORS   Phillips,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-1997) Wellcome Trust Genome Campus, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquires:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 4, 1997 this sequence version replaced gi:2342575.
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and
```



```
            the release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated
            with foreign sequence from E.coli, yeast, vector, phage etc. Order
            of segments is not known; 800 n's separate segments. Unfinished
            sequence: dJ12803  Contig_ID: 01586  Length: 60169 bp Unfinished
            sequence: dJ12803  Contig_ID: 00746  Length: 804 bp Unfinished
            sequence: dJ12803  Contig_ID: 02675  Length: 764 bp Unfinished
            sequence: dJ12803  Contig_ID: 00270  Length: 7709 bp Unfinished
            sequence: dJ12803  Contig_ID: 02449  Length: 22818 bp Unfinished
            sequence: dJ12803  Contig_ID: 02304  Length: 28251 bp.
            ***                                                          ***
            *** WARNING: Phase 1 High Throughput Genome Sequence         ***
            ***                                                          ***
            * This sequence is unfinished. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
FEATURES             Location/Qualifiers
     source          1. .124515
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="12803"
                     /chromosome="6"
BASE COUNT    35048 a  25304 c  25414 g  34720 t    4029 others
ORIGIN

  Query Match           21.8%;  Score 19;  DB 17;  Length 124515;
  Best Local Similarity 100.0%;  Pred. No. 9.68e-01;
  Matches    19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  16728 ATTTTACTATATTTTCCT 16746
          ||||||||||||||||||
Cp     49 ATTTTACTATATTTTCCT 31


RESULT    6
LOCUS       HUAC004158 180551 bp    DNA             PRI       24-MAR-1998
DEFINITION  Homo sapiens Chromosome 16 BAC clone CIT987SK-A-10F4, complete
            sequence.
ACCESSION   AC004158
NID         g2982170
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 180551)
  AUTHORS   Adams,M.D., Loftus,B.J., Zhou,L., Crosby,M., Fuhrmann,J.,
            Mason,T.M., Brandon,R., Kim,U.J., Kerlavage,A.R. and Venter,J.C.
  TITLE     Homo sapiens Chromosome 16 BAC clone CIT987SK-A-10F4
  JOURNAL   Unpublished
```

```
  REFERENCE   2  (bases 1 to 180551)
    AUTHORS   Adams,M.D. and Loftus,B.J.
    TITLE     Direct Submission
    JOURNAL   Submitted (19-FEB-1998) The Institute for Genomic Research, 9712
              Medical Center Dr, Rockville, MD 20850, USA, Email:
              bjloftus@tigr.org
  REFERENCE   3  (bases 1 to 180551)
    AUTHORS   Adams,M.D. and Loftus,B.J.
    TITLE     Direct Submission
    JOURNAL   Submitted (24-MAR-1998) The Institute for Genomic Research, 9712
              Medical Center Dr., Rockville, MD 20850, USA
  COMMENT     On Mar 24, 1998 this sequence version replaced gi:2924795.
              Address all correspondence to: Mark Adams The Institute for Genomic
              Research 9712 Medical Center Dr, Rockville, MD 20850, USA e-mail
              address: mdadams@tigr.org.   The orientation of the sequence is
              from SP6 end to T7 end. Genes were identified by a combination of
              five methods including: XGRAIL (available by anonymous ftp from
              arthur.epm.ornl.gov), Genefinder (Phil Green, University of
              Washington), Genscan  (Chris Burge,
              http://gnomic.stanford.Edu/~chris/GENSCANW.html)searches of the
              complete sequence against a peptide database,and the Human gene
              Index database at TIGR (http://www.tigr.org/tdb/hgi/hgi.html).
              Genes without pepetide homolgy having spliced EST hits are termed
              'Unknown gene product'. Genes encoding tRNAs are predicted by
              tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
  FEATURES             Location/Qualifiers
       source          1. .180551
                       /organism="Homo sapiens"
                       /db_xref="taxon:9606"
                       /chromosome="16"
                       /map="16q22.2"
                       /clone="A-10F4"
       STS             1331. .1462
                       /note="32826, A002D07, Chr. -, Homo sapiens"
                       /db_xref="dbSTS:G19948"
       STS             2359. .2536
                       /note="7821, STS1-cSRL-31b6-uA/cSRL-31b6-uZ, Chr. -, Homo
                       sapiens"
                       /db_xref="dbSTS:G02335"
       STS             24885. .25016
                       /note="32826, A002D07, Chr. -, Homo sapiens"
                       /db_xref="dbSTS:G19948"
       STS             125974. .126123
                       /note="10000, WI-4008, Chr. 16, Homo sapiens"
                       /db_xref="dbSTS:G04514"
  BASE COUNT    57546 a   32350 c   33107 g   57548 t
  ORIGIN

    Query Match           21.8%;  Score 19;  DB 26;  Length 180551;
    Best Local Similarity 100.0%;  Pred. No. 9.68e-01;
    Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```



```
Db  47263 ACATTTTACTATATTTTTC 47281
          |||||||||||||||||||
Cp     51 ACATTTTACTATATTTTTC 33


  RESULT    7
  LOCUS       MMU17092     1903 bp    DNA              ROD       18-DEC-1994
  DEFINITION  Mus musculus ORR1a transposon-like element, clone ORR-F.
  ACCESSION   U17092
  NID         g602951
  KEYWORDS    .
  SOURCE      house mouse.
    ORGANISM  Mus musculus
              Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
              Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;
              Murinae; Mus.
  REFERENCE   1  (bases 1 to 1903)
    AUTHORS   Smit,A.F.A. and Riggs,A.D.
    TITLE     Analysis of transposable elements in rodents related to THE1 in
              primates
    JOURNAL   Unpublished
  REFERENCE   2  (bases 1 to 1903)
    AUTHORS   Smit,A.F.A.
    TITLE     Direct Submission
    JOURNAL   Submitted (10-NOV-1994) Arian F.A. Smit, Molecular Biotechnology,
              University of Washington, Seattle, WA 98195, USA
  FEATURES             Location/Qualifiers
       source          1. .1903
                       /organism="Mus musculus"
                       /strain="C57"
                       /db_xref="taxon:10090"
                       /clone="ORR-F"
       LTR             <1. .44
       repeat_region   <1. .>1903
                       /standard_name="ORR1a transposon-like element"
                       /rpt_type=dispersed
                       /rpt_family="MaLR"
       CDS             90. .545
                       /note="has similarity with putative THE-1 product"
                       /pseudo
       repeat_region   605. .710
                       /rpt_type=dispersed
                       /rpt_family="B1"
       misc_feature    845. .1390
                       /note="similar to MT transposon-like element"
       LTR             1819. .>1903
  BASE COUNT      568 a    305 c    528 g    500 t      2 others
  ORIGIN

    Query Match           20.7%;  Score 18;  DB 28;  Length 1903;
```



```
    Best Local Similarity 100.0%;  Pred. No. 4.41e+00;
    Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

  Db    463 TTAGAAAGGAAAAATATA 480
            ||||||||||||||||||
  Qy     25 TTAGAAAGGAAAAATATA 42


  RESULT    8
  LOCUS       G06425       2438 bp    DNA              STS       19-OCT-1995
  DEFINITION  human STS WI-7161.
  ACCESSION   G06425
  NID         g859670
  KEYWORDS    STS sequence; primer; sequence tagged site.
  SOURCE      human STSs derived from sequences in dbEST and the Unigene
              collection.
    ORGANISM  Homo sapiens
              Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
              Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Choanata;
              Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates;
              Catarrhini; Hominidae; Homo.
  REFERENCE   1  (bases 1 to 2438)
    AUTHORS   Hudson,T.
    TITLE     Whitehead Institute/MIT Center for Genome Research; Physically
              Mapped ESTs
    JOURNAL   Unpublished (1995)
  COMMENT
              Contact: Thomas Hudson
              Whitehead Institute/MIT Center for Genome Research
              Whitehead Institute for Biomedical Research
              9 Cambridge Center, Cambridge MA 02142 USA
              Tel: 617 252 1900
              Fax: 617 252 1902
              Email: thudson@genome.wi.mit.edu

              Primer A: AGGACATGACCAGCAGCTG
              Primer B: CACTGTACGGGGAGTGGG
              STS size: 345
              PCR Profile:
                   Presoak:
                   Denaturation:
                   Annealing: 56 degrees C
                   Polymerization:
                   PCR Cycles: 35
                   Thermal Cycler:
              Protocol:
                  Template: 10 ng
                  Primer: each 5 pM
                  dNTPs: each 4 nM
                  Taq Polymerase: 0.025 units/ul
                  Total Vol: 20 ul
```



```
              Buffer:
                  MgCl2: 1.5 mM
                  KCl: 50 mM
                  Tris-HCL: 10 mM
                  pH: 9.3

              Prepared with primer pairs derived from M35878 -- Unigene.
  FEATURES             Location/Qualifiers
       source          1. .2438
                       /organism="Homo sapiens"
       STS             1003. .1347
       primer_bind     1003. .1021
       primer_bind     complement(1330. .1347)
  BASE COUNT      680 a    494 c    543 g    636 t     85 others
  ORIGIN

    Query Match           20.7%;  Score 18;  DB 30;  Length 2438;
    Best Local Similarity 100.0%;  Pred. No. 4.41e+00;
    Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

  Db    657 ATTTTATTAAGACAAATT 674
            ||||||||||||||||||
  Cp     71 ATTTTATTAAGACAAATT 54


  RESULT    9
  LOCUS       HUMIBP3     10884 bp    DNA              PRI       08-NOV-1994
  DEFINITION  Human insulin-like growth factor-binding protein-3 gene, complete
              cds, clone HL1006d.
  ACCESSION   M35878 J05537 J05538 M35879 M35880 M35881 M35882 M35883 M35884
              M35885 M35886 M36121 M36122
  NID         g184522
  KEYWORDS    insulin-like growth factor binding protein 3.
  SOURCE      Homo sapiens DNA.
    ORGANISM  Homo sapiens
              Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
              Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE   1  (bases 1 to 10884)
    AUTHORS   Cubbage,M.L., Suwanichkul,A. and Powell,D.R.
    TITLE     Insulin-like growth factor binding protein-3. Organization of the
              human chromosomal gene and demonstration of promoter activity
    JOURNAL   J. Biol. Chem. 265 (21), 12642-12649 (1990)
    MEDLINE   90324259
  COMMENT     Draft entry and computer-readable sequence for [J. Biol. Chem.
              (1990) In press] kindly submitted
              by D.R.Powell, 03-JUL-1990
              The sequence presented here appears in Figures 2 and 3 of ref. [J.
              Biol. Chem. (1990) In press].
  FEATURES             Location/Qualifiers
       source          1. .10884
```

```
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_type="leukocyte"
                     /map="Unassigned"
     promoter        102. .1905
                     /gene="IGFBP3"
                     /note="G00-126-724"
     protein_bind    1808. .1821
                     /gene="IGFBP3"
                     /note="G00-126-724"
                     /bound_moiety="Sp1 and AP-2"
     TATA_signal     1876. .1881
                     /gene="IGFBP3"
                     /note="G00-126-724"
     exon            1906. .2440
                     /partial
                     /gene="IGFBP3"
                     /note="G00-126-724"
                     /number=1
     sig_peptide     2038. .2118
                     /gene="IGFBP3"
                     /note="G00-126-724"
     gene            join(2038. .2440,5726. .5952,6497. .6616,8212. .8337)
                     /gene="IGFBP3"
     CDS             join(2038. .2440,5726. .5952,6497. .6616,8212. .8337)
                     /gene="IGFBP3"
                     /note="The AAs encoded by bases 5728-5736 and 5917-5926
                     may be ASN-linked glycosylation sites.; insulin-like
                     precursor"
                     /codon_start=1
                     /db_xref="GDB:G00-126-724"
                     /product="growth factor-binding protein-3"
                     /db_xref="PID:g386791"
                     /translation="MQRARPTLWAAALTLLVLLRGPPVARAGASSGGLGPVVRCEPCD
                     ARALQCAPPPAVCAELVREPGCGCCLTCALSEGQPCGITTERCGSGLRCQPSPDEAR
                     PLQALLDGRGLCVNASAVSRLRAYLLPAPPAPGNASESEEDRSAGSVESPSVSSTHRV
                     SDPRFHPLHSKIIIIKKGHAKDSQRYKVDYESQSTDTQNFSSESKRETEYGPCRREME
                     DTLNHLKFLNVLSPRGVHIPNCDKKGFYKKKQCRPSKGRKRGFCWCVDKYGQPLPGYT
                     TKGKEDVHCYSMQSK"
     mat_peptide     join(2119. .2440,5726. .5952,6497. .6616,8212. .8334)
                     /gene="IGFBP3"
                     /note="G00-126-724"
                     /product="growth factor-binding protein-3"
     intron          2441. .5725
                     /gene="IGFBP3"
                     /note="G00-126-724"
                     /number=1
     exon            5726. .5952
                     /gene="IGFBP3"
                     /note="G00-126-724"
                     /number=2
```



```
     intron          5953. .6496
                     /gene="IGFBP3"
                     /note="IGFBP-3 intron B; G00-126-724"
     exon            6497. .6616
                     /gene="IGFBP3"
                     /note="insulin-like growth factor-binding protein-3
                     precursor; G00-126-724"
                     /number=3
     intron          6617. .8211
                     /gene="IGFBP3"
                     /note="IGFBP-3 intron C; G00-126-724"
     exon            8212. .10775
                     /gene="IGFBP3"
                     /note="insulin-like growth factor-binding protein-3
                     precursor; G00-126-724"
                     /number=4
     polyA_signal    10751. .10756
                     /gene="IGFBP3"
                     /note="G00-126-724"
BASE COUNT     2796 a   2578 c   2737 g   2773 t
ORIGIN

  Query Match           20.7%;  Score 18;  DB 25;  Length 10884;
  Best Local Similarity 100.0%;  Pred. No. 4.41e+00;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   8994 ATTTTATTAAGACAAATT 9011
         ||||||||||||||||||
Cp     71 ATTTTATTAAGACAAATT 54


RESULT   10
LOCUS       AC002315   110079 bp    DNA             ROD       20-FEB-1998
DEFINITION  Mouse BAC-146N21 Chromosome X contains iduronate-2-sulfatase gene;
            complete sequence.
ACCESSION   AC002315
NID         g2258164
KEYWORDS    HTG.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 110079)
  AUTHORS   Timms,K.M., Huckett,L., Edwards,F.J., Lu,J., Muzny,D.M., Miller,W.
            and Gibbs,R.A.
  TITLE     Cross-species sequence comparison of the IDS region
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 110079)
  AUTHORS   Muzny,D., Ansari-Lari,M.A., Timms,K.M., Yu,W., Dugan,S., Lu,J.,
            Shen,Y., Rowland,K., Liu,W., Perez,L., Ding,Y., Haywood,M.,
            Jain,A., Leal,B., Logan,O., Nguyen,V., Savage,L., Shen,H.,
```



```
            Worley,K., Chen,E., Forcum,J., Arenson,A.D., Chiu,M.W.,
            Gorrell,J.H., Brundage,E., Di,W., Chinault,C., Nelson,D. and
            Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 110079)
  AUTHORS   Chiu,M.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JUL-1997) Molecular and Human Genetics, Baylor
            College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT     Sequencing is completed to a minimum standard of double strand
            coverage with a minimum of 2 clones and 2 reads with no ambiguities
            or 2 chemistries with a minimum of 2 clones and 3 reads with no
            ambiguities. If the sequence quality does not meet this standard,
            it will be indicated in the annotation.

            The repeat regions shown were identified using RepeatMasker by
            Adrian Smit.

            Sequence similarities were identified using Powerblast by Jinghui
            Zhang.

            Exon/Intron boundaries of identified genes were chosen if there
            were canonical splice junctions that maintained sequence continuity
            across the splice junctions.
FEATURES             Location/Qualifiers
     source          1. .110079
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="X"
                     /clone="BAC-146N21"
     repeat_region   244. .561
                     /rpt_family="LTR/MaLR"
     repeat_region   567. .2305
                     /rpt_family="LTR/MaLR"
     repeat_region   2306. .2646
                     /rpt_family="LTR/MaLR"
     repeat_region   complement(2864. .3101)
                     /rpt_family="LTR/Retroviral"
     repeat_region   complement(3812. .3847)
                     /rpt_family="Simple_repeat"
     repeat_region   complement(4982. .5064)
                     /rpt_family="Simple_repeat"
     repeat_region   complement(5065. .5184)
                     /rpt_family="LINE/L1"
     repeat_region   complement(5257. .5457)
                     /rpt_family="SINE/B2"
     repeat_region   6904. .6953
                     /rpt_family="Simple_repeat"
     repeat_region   complement(7980. .8199)
                     /rpt_family="SINE/B2"
```



```
     repeat_region   complement(10494. .10582)
                     /rpt_family="SINE/Alu"
     repeat_region   complement(10633. .10861)
                     /rpt_family="LTR/MaLR"
     repeat_region   complement(10892. .11185)
                     /rpt_family="SINE/B4"
     repeat_region   complement(11193. .11366)
                     /rpt_family="LTR/MaLR"
     repeat_region   11629. .11718
                     /rpt_family="Simple_repeat"
     repeat_region   complement(11839. .11971)
                     /rpt_family="SINE/B4"
     repeat_region   complement(12138. .12171)
                     /rpt_family="Low_complexity"
     repeat_region   complement(12276. .12396)
                     /rpt_family="SINE/B2"
     repeat_region   14061. .14118
                     /rpt_family="Simple_repeat"
     unsure          14096. .14220
                     /note="Single strand coverage only"
     repeat_region   14179. .14201
                     /rpt_family="Low_complexity"
     gene            join(14201. .14287,15145. .15281,15929. .16106,17952. .18040,
                     18934. .19134,20646. .20816,25808. .25934,27982. .28155)
                     /gene="IDS"
     repeat_region   complement(14424. .14541)
                     /rpt_family="Simple_repeat"
     repeat_region   complement(14783. .14914)
                     /rpt_family="Simple_repeat"
     repeat_region   17009. .17084
                     /rpt_family="SINE/MIR"
     repeat_region   complement(17555. .17697)
                     /rpt_family="SINE/Alu"
     repeat_region   complement(17620. .17719)
                     /rpt_family="LTR/MaLR"
     repeat_region   complement(17821. .17869)
                     /rpt_family="SINE/MIR"
     repeat_region   complement(18187. .18314)
                     /rpt_family="SINE/Alu"
     repeat_region   18371. .18499
                     /rpt_family="SINE/B4"
     repeat_region   19306. .19411
                     /rpt_family="SINE/Alu"
     repeat_region   complement(19902. .20388)
                     /rpt_family="Other/MER21_gro"
     repeat_region   complement(21582. .21661)
                     /rpt_family="Simple_repeat"
     repeat_region   complement(21662. .21761)
                     /rpt_family="scRNA"
     repeat_region   complement(21684. .21761)
                     /rpt_family="SINE/ID"
```

```
     repeat_region    21812. .21849
                      /rpt_family="DNA/MER1_type"
     repeat_region    complement(23849. .23940)
                      /rpt_family="SINE/Alu"
     repeat_region    25152. .25220
                      /rpt_family="LTR/MaLR"
     repeat_region    complement(26165. .26376)
                      /rpt_family="SINE/B2"
     repeat_region    complement(26379. .26577)
                      /rpt_family="SINE/B2"
     repeat_region    26995. .27125
                      /rpt_family="SINE/B2"
     repeat_region    27322. .27505
                      /rpt_family="SINE/B2"
     repeat_region    complement(27695. .27734)
                      /rpt_family="Low_complexity"
     repeat_region    29292. .30131
                      /rpt_family="LINE/L1"
     repeat_region    30204. .30233
                      /rpt_family="Simple_repeat"
     repeat_region    complement(31179. .31235)
                      /rpt_family="Simple_repeat"
     repeat_region    31236. .31266
                      /rpt_family="Simple_repeat"
     repeat_region    31579. .31703
                      /rpt_family="Simple_repeat"
     repeat_region    31705. .32083
                      /rpt_family="LINE/L1"
     repeat_region    33920. .33985
                      /rpt_family="Low_complexity"
     repeat_region    34396. .34534
                      /rpt_family="Simple_repeat"
     repeat_region    complement(34591. .34664)
                      /rpt_family="DNA/MER1_type"
     repeat_region    complement(34824. .34916)
                      /rpt_family="Simple_repeat"
     repeat_region    complement(35493. .35633)
                      /rpt_family="LINE/L1"
     repeat_region    complement(36159. .36380)
                      /rpt_family="SINE/B2"
     repeat_region    complement(36381. .36489)
                      /rpt_family="LTR/MaLR"
     repeat_region    36490. .36804
                      /rpt_family="LTR/MaLR"
     repeat_region    36803. .38233
                      /rpt_family="LTR/MaLR"
     repeat_region    38236. .38585
                      /rpt_family="LTR/MaLR"
     repeat_region    complement(38986. .39048)
                      /rpt_family="SINE/Alu"
     repeat_region    39226. .39295
```



```
                      /rpt_family="Low_complexity"
     repeat_region    complement(39309. .39438)
                      /rpt_family="SINE/Alu"
     repeat_region    complement(39439. .39929)
                      /rpt_family="LINE/L1"
     repeat_region    complement(40689. .40779)
                      /rpt_family="SINE/Alu"
     repeat_region    complement(41671. .42216)
                      /rpt_family="LTR/MaLR"
     repeat_region    complement(42325. .42428)
                      /rpt_family="Simple_repeat"
     repeat_region    42539. .42710
                      /rpt_family="Simple_repeat"
     repeat_region    42898. .42980
                      /rpt_family="DNA/MER1_type"
     repeat_region    complement(43056. .43106)
                      /rpt_family="Low_complexity"
     repeat_region    43286. .43340
                      /rpt_family="Low_complexity"
     repeat_region    43406. .43513
                      /rpt_family="SINE/Alu"
     repeat_region    43947. .44060
                      /rpt_family="Simple_repeat"
     repeat_region    46601. .46704
                      /rpt_family="DNA/MER1_type"
     repeat_region    46977. .47046
                      /rpt_family="Simple_repeat"
     repeat_region    48901. .49040
...
Note: remainder of annotations omitted.

  Query Match           20.7%;  Score 18;  DB 28;  Length 110079;
  Best Local Similarity 100.0%;  Pred. No. 4.4le+00;
  Matches    18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    985 TTAGAAAGGAAAAATATA 1002
          ||||||||||||||||||
Qy     25 TTAGAAAGGAAAAATATA 42


RESULT   11
LOCUS       AC002406   194985 bp    DNA              ROD       21-MAR-1998
DEFINITION  Mouse chromosome X BAC B178A13 (Research Genetics mouse BAC
            library) complete sequence.
ACCESSION   AC002406
NID         g2981248
KEYWORDS    HTG.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
```



```
  REFERENCE   1  (bases 1 to 194985)
    AUTHORS   Muzny,D., Arenson,A.D., Brundage,E., Carvelli,K., Chen,E., Chen,J.,
              Di,W., Ding,Y., Dugan,S., Durbin,J., Forcum,J., Ganesh,R.,
              Garcia,C., Goodman,M., Gorrell,J.H., Haywood,M., Jackson,L.,
              Jin,S., Kampal,R., Karpathy,S., Leal,B., Li,Y., Liu,W., Logan,O.,
              Lu,J., Ly,T., Martinez,C., Oswal,G., Perez,L., Rashid,N.D.,
              Rowland,K., Savage,L., Scherer,S.S., Shen,H., Timms,K.M., Todd,J.,
              Vo,Q., Worley,K.C., Yu,W., Chinault,C., Nelson,D. and Gibbs,R.A.
    TITLE     Direct Submission
    JOURNAL   Unpublished
  REFERENCE   2  (bases 1 to 194985)
    AUTHORS   Chiu,M.W.
    TITLE     Direct Submission
    JOURNAL   Submitted (11-AUG-1997) Molecular and Human Genetics, Baylor
              College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
  REFERENCE   3  (bases 1 to 194985)
    AUTHORS   Worley,K.C.
    TITLE     Direct Submission
    JOURNAL   Submitted (21-MAR-1998) Human Genome Sequencing Center, Department
              of Molecular and Human Genetics, Baylor College of Medicine, One
              Baylor Plaza, Houston, TX 77030, USA
  COMMENT     On Mar 23, 1998 this sequence version replaced gi:2909685.
              Sequencing is completed to a minimum standard of double strand
              coverage with a minimum of 2 clones and 2 reads with no ambiguities
              or 2 chemistries with a minimum of 2 clones and 3 reads with no
              ambiguities. If the sequence quality does not meet this standard,
              it will be indicated in the annotation.

              The repeat regions shown were identified using RepeatMasker by
              Adrian Smit.

              Sequence similarities were identified using Powerblast by Jinghui
              Zhang.

              Exon/Intron boundaries of identified genes were chosen if there
              were canonical splice junctions that maintained sequence continuity
              across the splice junctions.
  FEATURES             Location/Qualifiers
       source          1. .194985
                       /organism="Mus musculus"
                       /db_xref="taxon:10090"
                       /clone="B178A13"
                       /chromosome="X"
                       /clone_lib="Research Genetics mouse BAC library"
       repeat_region   226. .352
                       /rpt_family="MSTD"
       repeat_region   562. .599
                       /rpt_family="(TA)n"
       repeat_region   complement(863. .1868)
                       /rpt_family="L1_MM"
       repeat_region   complement(2661. .2921)
```



```
                       /rpt_family="B4A"
       repeat_region   complement(5160. .5372)
                       /rpt_family="B3"
       repeat_region   complement(5456. .5705)
                       /rpt_family="L1MD3"
       repeat_region   complement(5766. .5807)
                       /rpt_family="(TAAA)n"
       repeat_region   complement(5900. .6083)
                       /rpt_family="L1MD3"
       repeat_region   complement(6433. .6546)
                       /rpt_family="(TAAAA)n"
       repeat_region   complement(6547. .6575)
                       /rpt_family="AT_rich"
       repeat_region   complement(6576. .6776)
                       /rpt_family="L1MD2"
       repeat_region   complement(6808. .6947)
                       /rpt_family="B1-F"
       repeat_region   complement(7019. .7348)
                       /rpt_family="L1MD2"
       repeat_region   7396. .7541
                       /rpt_family="B1_MM"
       repeat_region   7542. .7587
                       /rpt_family="(CAAA)n"
       repeat_region   complement(7915. .8149)
                       /rpt_family="L1M4"
       repeat_region   8166. .8512
                       /rpt_family="L1M4"
       repeat_region   complement(8730. .9162)
                       /rpt_family="L1_MM"
       repeat_region   9149. .9265
                       /rpt_family="L1_MM"
       repeat_region   complement(9454. .9606)
                       /rpt_family="RLTR10"
       repeat_region   complement(9734. .10146)
                       /rpt_family="L1_MM"
       repeat_region   11635. .11661
                       /rpt_family="POLY_A"
       repeat_region   complement(11830. .11966)
                       /rpt_family="RLTR10"
       repeat_region   11967. .12203
                       /rpt_family="L1_MM"
       repeat_region   12575. .12708
                       /rpt_family="B1-F"
       repeat_region   12711. .12772
                       /rpt_family="(CAAAA)n"
       repeat_region   complement(13570. .13598)
                       /rpt_family="AT_rich"
       repeat_region   14287. .14359
                       /rpt_family="MIR"
       repeat_region   14418. .15428
                       /rpt_family="L1"
```

```
     repeat_region   15490. .15596
                     /rpt_family="L1"
     repeat_region   15828. .16199
                     /rpt_family="Lx4"
     repeat_region   complement(16753. .16830)
                     /rpt_family="MTE"
     repeat_region   17198. .17275
                     /rpt_family="MLT1A1"
     repeat_region   17311. .17443
                     /rpt_family="MLT1A2"
     repeat_region   complement(19559. .19879)
                     /rpt_family="MTE"
     repeat_region   complement(19895. .19963)
                     /rpt_family="MTE"
     repeat_region   19964. .26215
                     /rpt_family="L1P"
     repeat_region   complement(26283. .26511)
                     /rpt_family="L1"
     repeat_region   26468. .27422
                     /rpt_family="L1_MM"
     repeat_region   complement(27624. .27755)
                     /rpt_family="B1-F"
     repeat_region   27759. .28051
                     /rpt_family="MTC"
     repeat_region   28280. .28490
                     /rpt_family="L1"
     repeat_region   28491. .33172
                     /rpt_family="L1_MM"
     misc_feature    33178. .33672
                     /note="Region: Similar to U29402|MMU29402 Mus musculus
                     acidic ribosomal phosphoprotein P1 mRNA, complete cds"
     repeat_region   33682. .35079
                     /rpt_family="L1_MM"
     repeat_region   37207. .37292
                     /rpt_family="RSINE1"
     repeat_region   complement(37310. .37338)
                     /rpt_family="(GA)n"
     repeat_region   37354. .37474
                     /rpt_family="(CATA)n"
     repeat_region   complement(37574. .37728)
                     /rpt_family="RSINE1"
     repeat_region   complement(38061. .38179)
                     /rpt_family="(TA)n"
     repeat_region   39320. .39492
                     /rpt_family="B4"
     repeat_region   39603. .39661
                     /rpt_family="(CA)n"
     repeat_region   complement(40026. .40452)
                     /rpt_family="Lx2B"
     repeat_region   40580. .40624
                     /rpt_family="(CA)n"
```



```
     repeat_region   complement(40951. .41040)
                     /rpt_family="MLT2E"
     repeat_region   41143. .41538
                     /rpt_family="MTA"
     repeat_region   complement(42185. .42262)
                     /rpt_family="(CAG)n"
     repeat_region   complement(43208. .43318)
                     /rpt_family="(CAGA)n"
     repeat_region   complement(43486. .43974)
                     /rpt_family="L1"
     repeat_region   43954. .44894
                     /rpt_family="L1_MM"
     repeat_region   45112. .45198
                     /rpt_family="(CAA)n"
     repeat_region   complement(45459. .45479)
                     /rpt_family="AT_rich"
     repeat_region   46721. .46836
                     /rpt_family="(GAA)n"
     repeat_region   46914. .47150
                     /rpt_family="MTD"
     repeat_region   47213. .47296
                     /rpt_family="PB1D10"
     repeat_region   50241. .50324
                     /rpt_family="MTC"
     repeat_region   50328. .51555
                     /rpt_family="L1_MM"
     repeat_region   51568. .52378
                     /rpt_family="L1_MM"
     repeat_region   52399. .52434
                     /rpt_family="POLY_A"
     repeat_region   52446. .52503
                     /rpt_family="L1_MM"
     repeat_region   52520. .52838
                     /rpt_family="MTC"
     repeat_region   53200. .53245
                     /rpt_family="(CAAAA)n"
     repeat_region   complement(53474. .53534)
...
Note: remainder of annotations omitted.

  Query Match          20.7%;  Score 18;  DB 28;  Length 194985;
  Best Local Similarity 100.0%;  Pred. No. 4.4le+00;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  66755 TTAGAAAGGAAAAATATA 66772
          ||||||||||||||||||
Qy     25 TTAGAAAGGAAAAATATA 42


RESULT   12
LOCUS       HS357K22   199496 bp    DNA             HTG       22-SEP-1998
```



```
DEFINITION  Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone
            357K22; HTGS phase 1.
ACCESSION   AL022720
NID         g3646072
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 199496)
  AUTHORS   Heath,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-SEP-1998) Wellcome Trust Genome Campus, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquires:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 24, 1998 this sequence version replaced gi:3581799.
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and the release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc. Order of segments is not known; 800 n's separate
            segments. Unfinished: dJ357K22  Contig_ID: 01322  acc=AL022720
            Length: 1334 bp Unfinished: dJ357K22  Contig_ID: 02131
            acc=AL022720  Length: 1461 bp Unfinished: dJ357K22  Contig_ID:
            00190  acc=AL022720  Length: 1141 bp Unfinished: dJ357K22
            Contig_ID: 01246  acc=AL022720  Length: 1398 bp Unfinished:
            dJ357K22  Contig_ID: 01266  acc=AL022720  Length: 1535 bp
            Unfinished: dJ357K22  Contig_ID: 01249  acc=AL022720  Length: 1482
            bp Unfinished: dJ357K22  Contig_ID: 01290  acc=AL022720  Length:
            1308 bp Unfinished: dJ357K22  Contig_ID: 01927  acc=AL022720
            Length: 145619 bp Unfinished: dJ357K22  Contig_ID: 02277
            acc=AL022720  Length: 1211 bp Unfinished: dJ357K22  Contig_ID:
            02288  acc=AL022720  Length: 1644 bp Unfinished: dJ357K22
            Contig_ID: 02264  acc=AL022720  Length: 1255 bp Unfinished:
            dJ357K22  Contig_ID: 02257  acc=AL022720  Length: 1734 bp
            Unfinished: dJ357K22  Contig_ID: 02291  acc=AL022720  Length: 1842
            bp Unfinished: dJ357K22  Contig_ID: 02295  acc=AL022720  Length:
            1611 bp Unfinished: dJ357K22  Contig_ID: 02296  acc=AL022720
            Length: 1409 bp Unfinished: dJ357K22  Contig_ID: 02300
            acc=AL022720  Length: 1493 bp Unfinished: dJ357K22  Contig_ID:
            00634  acc=AL022720  Length: 1484 bp Unfinished: dJ357K22
            Contig_ID: 02301  acc=AL022720  Length: 1424 bp Unfinished:
            dJ357K22  Contig_ID: 02302  acc=AL022720  Length: 1524 bp
            Unfinished: dJ357K22  Contig_ID: 02256  acc=AL022720  Length: 1033
            bp Unfinished: dJ357K22  Contig_ID: 02252  acc=AL022720  Length:
            1393 bp Unfinished: dJ357K22  Contig_ID: 01983  acc=AL022720
            Length: 1237 bp Unfinished: dJ357K22  Contig_ID: 02237
            acc=AL022720  Length: 1051 bp Unfinished: dJ357K22  Contig_ID:
            02240  acc=AL022720  Length: 1645 bp Unfinished: dJ357K22
            Contig_ID: 02224  acc=AL022720  Length: 2028 bp.
```



```
            ***                                                    ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                    ***
            * This sequence is unfinished. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
FEATURES              Location/Qualifiers
     source           1. .199496
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="X"
                      /clone="J357K22"
BASE COUNT    56004 a  36575 c  32647 g  54864 t  19406 others
ORIGIN

  Query Match          20.7%;  Score 18;  DB 17;  Length 199496;
  Best Local Similarity 100.0%;  Pred. No. 4.4le+00;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  18472 ATTTTATTAAGACAAATT 18489
          ||||||||||||||||||
Cp     71 ATTTTATTAAGACAAATT 54


RESULT   13
LOCUS       AF054044      453 bp    DNA             ROD       15-APR-1998
DEFINITION  Rattus norvegicus polymorphic marker D11UIA5 sequence.
ACCESSION   AF054044
NID         g3300022
KEYWORDS    .
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 453)
  AUTHORS   Walder,R.Y., Garrett,M.R., McClain,A.M., Beck,G.E., Brennan,T.M.H.,
            Kramer,N.A., Kanis,A.B., Mark,A.L., Rapp,J.P. and Sheffield,V.C.
  TITLE     High Quality Polymorphic Markers for the Rat Genome Developed from
            Marker-Selected Libraries
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 453)
  AUTHORS   Walder,R.Y., Garrett,M.R., McClain,A.M., Beck,G.E., Brennan,T.M.H.,
            Kramer,N.A., Kanis,A.B., Mark,A.L., Rapp,J.P. and Sheffield,V.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-MAR-1998) Internal Medicine and Pediatrics, The
            University of Iowa, 440 EMRB, Iowa City, IA 52242, USA
FEATURES              Location/Qualifiers
     source           1. .453
                      /organism="Rattus norvegicus"
                      /db_xref="taxon:10116"
                      /clone="RUIGATAC03"
```

```
BASE COUNT       89 a     99 c     76 g    169 t     20 others
ORIGIN

  Query Match           19.5%;  Score 17;  DB 28;  Length 453;
  Best Local Similarity  94.4%;  Pred. No. 1.92e+01;
  Matches    17;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db    305 TTAATAAAATCTATCTAT 322
          ||||||||||||||| |
Qy     62 TTAATAAAATCTATCTNT 79


RESULT   14
LOCUS       SLNPVPOLH     1057 bp    DNA              VRL       01-APR-1996
DEFINITION  Spodotera litura nuclear polyhydrosis virus polh gene for
            polyhedrin.
ACCESSION   X94437
NID         g1246844
KEYWORDS    polh gene; polyhedrin.
SOURCE      Spodoptera litura nuclear polyhedrosis virus.
  ORGANISM  Spodoptera litura nuclear polyhedrosis virus
            Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
            Nucleopolyhedrovirus.
REFERENCE   1  (bases 1 to 1057)
  AUTHORS   Bansal,A., Bansal,O.B., Behera,A.K. and Das,R.H.
  TITLE     Identification and nucleotide sequencing of the polyhedrin gene of
            Spodotera litura nuclear polyhedrosis virus
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1057)
  AUTHORS   Das,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JAN-1996) R.H. Das, Centre for Biochemical Technology
            CSIR, Genetic Engineering Division, Delhi University Campus, Mall
            Road, Delhi 110007, INDIA
  REMARK    Revised by submittor 26-JAN-96
FEATURES             Location/Qualifiers
     source          1. .1057
                     /organism="Spodoptera litura nuclear polyhedrosis virus"
                     /db_xref="taxon:46242"
                     /map="0.8% genome"
     misc_feature    240. .251
                     /note="conserved 12-mer"
     misc_signal     242. .245
                     /note="taag"
     gene            286. .1026
                     /gene="polh"
     CDS             286. .1026
                     /gene="polh"
                     /codon_start=1
                     /product="polyhedrin"
                     /db_xref="PID:e218453"
```



```
                     /db_xref="PID:g1246845"
                     /translation="MYTPYSYNPSLGRTYVYDNKFYKNLGSVIKNAKRKEHLVHHEIE
                     ERTLDPLERYVVAEDPFLGPGKNQKLTLFKEIRLVKPDTMKLVVNWSGKEFLRETWTR
                     FMEDSFPSVNDQEIMDVFLVINMRPTRPNRCYRFLAQHALRCDPDYVPHEVIRIVEPS
                     YVGSNNEYRISLAKKGGGCPVMNLHSEYTHSFEEFINRVIWENFYKPIVYVGTDSGEE
                     EEILLEVALVFKIKEFAPDASLYNGPAY"
     polyA_signal    1027. .1032
BASE COUNT       337 a    253 c    203 g    264 t
ORIGIN

  Query Match           19.5%;  Score 17;  DB 33;  Length 1057;
  Best Local Similarity 100.0%;  Pred. No. 1.92e+01;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    250 TTTTTCCTTTCTAAAAC 266
          |||||||||||||||||
Cp     38 TTTTTCCTTTCTAAAAC 22


RESULT   15
LOCUS       E06147       1667 bp    DNA              PAT       26-NOV-1996
DEFINITION  Promoter of Candida alcohol oxidase gene.
ACCESSION   E06147
NID         g2174334
KEYWORDS    JP 1993344895-A/1.
SOURCE      Candida boidinii.
  ORGANISM  Candida boidinii
            Eukaryotae; mitochondrial eukaryotes; Fungi; Ascomycota;
            Hemiascomycetes; Saccharomycetales; Candidaceae; Candida.
REFERENCE   1  (bases 1 to 1667)
  AUTHORS   Yasuyoshi,S., Yoshiki,T., Yuji,S., Hirato,K. and Haruyo,H. .
  TITLE     NEW VECTOR HAVING PROMOTER INDUCIBLE BY ADDITION OF METHANOL AND/OR
            GLYCEROL
  JOURNAL   Patent: JP 1993344895-A 1 27-DEC-1993;
            SUNTORY LTD
COMMENT     OS   Candida boidinii
            PN   JP 1993344895-A/1
            PD   27-DEC-1993
            PF   26-FEB-1993 JP 1993038035
            PR   28-FEB-1992 JP 92P    43361
            PI   SAKAI YASUYOSHI, TANI YOSHIKI, SHIBANO YUJI, KONDO HIRATO, PI
             HATANAKA HARUYO
            PC   C12N15/81,C12N1/19,C12N9/00,C12N9/02,C12N9/08, PC
            C12N9/12//C12N15/11,
            PC   C12N15/53,C12N15/54,(C12N1/19,C12R1:72),(C12N15/11,C12R1:72),
            PC   (C12N15/53,
            PC   C12R1:01),(C12N15/54,C12R1:865);
            CC   strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
```



```
            CC   *source: strain=S2 AOU-1;
            CC   *source: clone=pMOX330;
            FH   Key             Location/Qualifiers
            FH
            FT   promoter        1. .1667
            FT                   /note='promoter of Candida alcohol oxidase  FT
                                 gene'
            FT   TATA_signal     1329. .1335.
FEATURES             Location/Qualifiers
     source          1. .1667
                     /organism="Candida boidinii"
                     /db_xref="taxon:5477"
BASE COUNT      546 a    276 c    204 g    641 t
ORIGIN

  Query Match           19.5%;  Score 17;  DB 21;  Length 1667;
  Best Local Similarity 100.0%;  Pred. No. 1.92e+01;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    802 CACATTTTACTATATTT 818
          |||||||||||||||||
Cp     52 CACATTTTACTATATTT 36


RESULT   16
LOCUS       MMU17094     1898 bp    DNA              ROD       18-DEC-1994
DEFINITION  Mus musculus ORR1 transposon-like element.
ACCESSION   U17094
NID         g602953
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;
            Murinae; Mus.
REFERENCE   1  (bases 1 to 1898)
  AUTHORS   Smit,A.F.A. and Riggs,A.D.
  TITLE     Analysis of transposable elements in rodents related to THE1 in
            primates
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1898)
  AUTHORS   Smit,A.F.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-NOV-1994) Arian F.A. Smit, Molecular Biotechnology,
            University of Washington, Seattle, WA 98195, USA
FEATURES             Location/Qualifiers
     source          1. .1898
                     /organism="Mus musculus"
                     /strain="C57"
                     /db_xref="taxon:10090"
                     /clone="ORR-10"
```



```
     LTR             <1. .44
     repeat_region   <1. .>1898
                     /standard_name="ORR1 transposon-like element"
                     /rpt_type=dispersed
                     /rpt_family="MaLR"
     repeat_region   586. .691
                     /rpt_type=dispersed
                     /rpt_family="B1"
     misc_feature    825. .1379
                     /note="similar to MT transposon-like element"
     LTR             1806. .>1898
BASE COUNT      574 a    307 c    513 g    498 t      6 others
ORIGIN

  Query Match           19.5%;  Score 17;  DB 28;  Length 1898;
  Best Local Similarity 100.0%;  Pred. No. 1.92e+01;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    444 TTAGAAAGGAAAAATAT 460
          . |||||||||||||||||
Qy     25 TTAGAAAGGAAAAATAT 41


RESULT   17
LOCUS       HSMYOC1      2800 bp    DNA              PRI       21-APR-1998
DEFINITION  Homo sapiens myocilin (GLC1A) gene, promoter and exon 1.
ACCESSION   AF049791
NID         g3065670
KEYWORDS    .
SEGMENT     1 of 3
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2800)
  AUTHORS   Fingert,J.H., Ying,L., Swiderski,R.E., Nystuen,A.M., Arbour,N.C.,
            Alward,W.L.M., Sheffield,V.C. and Stone,E.M.
  TITLE     Characterization and comparison of the human and mouse GLC1A
            glaucoma genes
  JOURNAL   Genome Res. (1998) In press
REFERENCE   2  (bases 1 to 2800)
  AUTHORS   Fingert,J.H., Ying,L., Swiderski,R.E., Nystuen,A.M., Arbour,N.C.,
            Alward,W.L.M., Sheffield,V.C. and Stone,E.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-FEB-1998) Ophthalmology, University of Iowa, 200
            Hawkins Drive, Iowa City, IA 52242, USA
FEATURES             Location/Qualifiers
     source          1. .2800
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
```

```
                     /map="1q21-q31"
     promoter        1. .1905
                     /gene="GLC1A"
     exon            1906. .2509
                     /gene="GLC1A"
                     /number=1
BASE COUNT      781 a    588 c    673 g    758 t
ORIGIN

  Query Match           19.5%;  Score 17;  DB 26;  Length 2800;
  Best Local Similarity 100.0%;  Pred. No. 1.92e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    182 TACTATATTTTTCCTTT 198
          |||||||||||||||||
Cp     45 TACTATATTTTTCCTTT 29


RESULT   18
LOCUS       SPU004047     3535 bp    DNA             BCT       01-OCT-1994
DEFINITION  Streptococcus pneumoniae SSZ dextran glucosidase gene and insertion
            sequence IS1202 transposase gene, complete cds.
ACCESSION   U04047
NID         g506632
KEYWORDS    .
SOURCE      Streptococcus pneumoniae.
  ORGANISM  Streptococcus pneumoniae
            Eubacteria; Firmicutes; Low G+C gram-positive bacteria;
            Streptococcaceae; Streptococcus.
REFERENCE   1  (bases 1 to 3535)
  AUTHORS   Morona,J.K., Guidolin,A., Morona,R., Hansman,D. and Paton,J.C.
  TITLE     Isolation, characterization, and nucleotide sequence of IS1202, an
            insertion sequence of Streptococcus pneumoniae
  JOURNAL   J. Bacteriol. 176 (14), 4437-4443 (1994)
  MEDLINE   94292475
REFERENCE   2  (bases 1 to 3535)
  AUTHORS   Paton,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-DEC-1993) James C. Paton, Women's and Children's
            Hospital, Microbiology, King William Rd, North Adelaide, South
            Australia 5006, Australia
FEATURES             Location/Qualifiers
     source          1. .3535
                     /organism="Streptococcus pneumoniae"
                     /strain="SSZ, serotype 19F"
                     /db_xref="taxon:1313"
     repeat_unit     237. .263
                     /rpt_type=direct
     source          264. .2010
                     /organism="Streptococcus pneumoniae"
                     /insertion_seq="IS1202"
```



```
     repeat_unit     264. .286
                     /rpt_type=inverted
     -35_signal      295. .300
     -10_signal      321. .326
     RBS             337. .342
     CDS             354. .1751
                     /note="located in IS1202"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="transposase"
                     /db_xref="PID:g551466"
                     /translation="MNETKKYLVIKAIAQGKKTKKRACVELNLSERQINRPLLAYQQK
                     GKEAFRHGNRNRKPKHAIPDEIKERILKKYLSYETTYKPNVLHPCELLAEEGIKLSDT
                     TVRKILYKKNILSPKSHRKTKKRVREQAKLNLNQPLDNPILPTAKDFLEDPKKVHPSR
                     PRKKFAGELIQMDASPHAWFGPETTNLHLAIDDASGNILGAYFDKQETLNAYYHVLEQ
                     ILANHGIPLQMKTDKRTVFTYQASNSKKMEDDSYTQFGYACHQLGILLETTSIPQAKG
                     RVERLNQTLQSRLPIELERNKIHTLEEANTFLLSYIQTFNEQPGNKTKLSVFEEAPNP
                     SERNLILARLAERVVDSGHHIRFQNRCYIPTEQGKEVYFIRKTKALVLKAFDGDIYLN
                     IADKIYHTKELLDHELYSKNFEQEPEQKKERRKYIPPQTHPWKLTSFKQYLHKNKKDY
                     EEFTSEEIHSPQLQV"
     stem_loop       1883. .1908
     repeat_unit     1988. .2010
                     /rpt_type=inverted
     repeat_unit     2011. .2037
                     /rpt_type=direct
     CDS             complement(3019. .3504)
                     /note="dextran glucosidase; homology to C-terminal portion
                     of S. mutans dexB gene product, PIR Accession Number
                     A37231"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /db_xref="PID:g551467"
                     /translation="MRVRLLKKGVPIEEIMDSIRVIGRDNARTPMQWDESKNAGPSTG
                     QPWLAVNPNYEMINVQEALANPDSIPYTYQKLVQIRKENSWLIRADFELLDTADKVFA
                     YIRKDGDRRFLVVANLSNEEQDLTVEGKVKSVLIENTLAQEVFEKQILVPWDAFCVEL
                     L"
BASE COUNT     1206 a    712 c    616 g    1001 t
ORIGIN

  Query Match           19.5%;  Score 17;  DB 15;  Length 3535;
  Best Local Similarity 100.0%;  Pred. No. 1.92e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2064 AAAAATATAGTAAAATG 2080
          |||||||||||||||||
Qy     34 AAAAATATAGTAAAATG 50


RESULT   19
```



```
LOCUS       RNRRNA08     4025 bp    DNA             ROD       18-NOV-1986
DEFINITION  Rat rRNA spacer downstream of 28S rRNA gene 3'end.
ACCESSION   X03695
NID         g57151
KEYWORDS    identifier sequence; repetitive sequence; ribosomal DNA.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;
            Murinae; Rattus.
REFERENCE   1  (bases 1 to 4025)
  AUTHORS   Yavachev,L.P., Georgiev,O.I., Braga,E.A., Avdonina,T.A.,
            Bogomolova,A.E., Zhurkin,V.B., Nosikov,V.V. and Hadjiolov,A.A.
  TITLE     Nucleotide sequence analysis of the spacer regions flanking the rat
            rRNA transcription unit and identification of repetitive elements
  JOURNAL   Nucleic Acids Res. 14 (6), 2799-2810 (1986)
  MEDLINE   86176766
FEATURES             Location/Qualifiers
     source          1. .4025
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
     misc_feature    <1. .99
                     /note="28S rRNA gene"
     repeat_region   3704. .3718
                     /note="direct repeat"
     misc_feature    3719. .3817
                     /note="identifier sequence"
     repeat_region   3819. .3833
                     /note="direct repeat"
BASE COUNT      842 a   1083 c    880 g   1220 t
ORIGIN

  Query Match           19.5%;  Score 17;  DB 28;  Length 4025;
  Best Local Similarity 94.4%;  Pred. No. 1.92e+01;
  Matches   17;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db    963 TTAATAAAATCTATCTAT 980
          |||||||||||||||| |
Qy     62 TTAATAAAATCTATCTNT 79


RESULT   20
LOCUS       AF030369     4133 bp    DNA             BCT       16-JUL-1998
DEFINITION  Streptococcus pneumoniae strain SP-VA92 alpha, 1-6-glucosidase
            (dexB) gene, partial cds; putative regulatory protein (cpsA) and
            CpsB (cpsB) genes, complete cds; and putative chain length
            terminator (cpsC) gene, partial cds.
ACCESSION   AF030369
NID         g2804745
KEYWORDS    .
SOURCE      Streptococcus pneumoniae.
```



```
  ORGANISM  Streptococcus pneumoniae
            Eubacteria; Firmicutes; Clostridium group firmicutes;
            Streptococcaceae; Streptococcus.
REFERENCE   1  (bases 1 to 4133)
  AUTHORS   Coffey,T.J., Enright,M.C., Daniels,M., Morona,J.K., Morona,R.,
            Hryniewicz,W., Paton,J.C. and Spratt,B.G.
  TITLE     Recombinational exchanges at the capsular polysaccharide
            biosynthetic locus lead to frequent serotype changes among natural
            isolates of Streptococcus pneumoniae
  JOURNAL   Mol. Microbiol. 27 (1), 73-83 (1998)
  MEDLINE   98125733
REFERENCE   2  (bases 1 to 4133)
  AUTHORS   Coffey,T.J., Daniels,M. and Spratt,B.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-OCT-1997) The Wellcome Trust Centre for the
            Epidemiology of Infectious Disease, Department of Zoology,
            University of Oxford, South Parks Road, Oxford, Oxfordshire OX1
            3PS, UK
REFERENCE   3  (bases 1 to 4133)
  AUTHORS   Coffey,T.J., Daniels,M. and Spratt,B.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-1998) The Wellcome Trust Centre for the
            Epidemiology of Infectious Disease, Department of Zoology,
            University of Oxford, South Parks Road, Oxford, Oxfordshire OX1
            3PS, UK
  REMARK    Sequence update by submitter
FEATURES             Location/Qualifiers
     source          1. .4133
                     /organism="Streptococcus pneumoniae"
                     /strain="SP-VA92"
                     /serotype="19F"
                     /db_xref="taxon:1313"
                     /note="clinical isolate from Spain"
     gene            <1. .439
                     /gene="dexB"
     CDS             <1. .439
                     /gene="dexB"
                     /codon_start=2
                     /transl_table=11
                     /product="alpha, 1-6-glucosidase"
                     /db_xref="PID:g2804746"
                     /translation="DSIRVIGRDNARTPMQWDESKNAGFSTDQPWLAVNPNYEMINVQ
                     EALANPDSIPYTYQKLVQIRKENSWLIRADFELLDTADKVFAYIRKDGDRRFLVVANL
                     SNEEQDLTVEGKVKSVLIENTLAQEVFEKQILVPWDAFCVELL"
     gene            1505. .2950
                     /gene="cpsA"
     CDS             1505. .2950
                     /gene="cpsA"
                     /codon_start=1
                     /transl_table=11
                     /product="putative regulatory protein"
```

```
                     /db_xref="PID:g2804747"
                     /translation="MSRRFKKSRSQKVKRSVNIVLLTIYLLLVCFLLFLIFKYNILAF
                     RYLNLVVTALVLLVALVGLLLIIYKKAEKFTIFLLVFSILVSSVSLFAVQQFVGLTNR
                     LNATSNYSEYSISVAVLADSDIENVTQLTSVTAPTGTDNENIQKLLADIKSSQNTDLT
                     VNQSSSYLAAYKSLIAGETKAIVLNSVFENIISSEYPDYASKIKKIYTKGFTKVEAP
                     KTSKNQSFNIYVSGIDTYGPISSVSRSDVNILMTVNRDTKKILLTTTPRDATVPIADG
                     GNNQKDKLTHAGIYGVDSSIHTLENLYGVDINYYVRLNFTSPLKMIDLLGGVDVHNDQ
                     EFSALHGKFHFPVGNVHLDSEQALGFVRERYSLADGDRDRGRNQQKVIVAIIQKLTST
                     EVLKNYSSILQGLQDSLQTNMPIETMIDLVNTQLESGGNYKVNSQDLKGTGRMDLPSY
                     AMPDINLYVMEIDDSSLAVVKAAIQDVMEGR"
     gene            2952. .3683
                     /gene="cpsB"
     CDS             2952. .3683
                     /gene="cpsB"
                     /codon_start=1
                     /transl_table=11
                     /product="CpsB"
                     /db_xref="PID:g2804748"
                     /translation="MIDIHSHIVFDVDDGPKSREESKALLAESYRQGVRTIVSTSHRR
                     KGMPETPEEKIAENFLQVREIAKEVADDLVIAYGAEIYYTLDALEKLEKKEIPTLNDS
                     RYALIEFSMHTSYRQIHTGLSNILMLGITPVIAHIERYDALENNEKRVRELIDMGCYT
                     QINSYHVSKPKFFGEKYKFMKKRARYFLERDLVHVVASDMHNLDSRPPYMQQAYDIIA
                     KKYGAKKAKELFVDNPRKIIMDQLI"
     gene            3692. .>4133
                     /gene="cpsC"
     CDS             3692. .>4133
                     /gene="cpsC"
                     /codon_start=1
                     /transl_table=11
                     /product="putative chain length terminator"
                     /db_xref="PID:g2804749"
                     /translation="MKEQNTLEIDVLQLFRALWKRKLVILLVAIITSSVAPAYSTFVI
                     KPEPTSMTRIYVVNRDQGEKSGLTNQDLQAGSSLVKDYREIILSQDVLEEVVSDLKLD
                     LTPKDLANKIKVTVPVDTRIVSVSVSDRVPEEASRIANSLREVAA"
BASE COUNT     1314 a    708 c    834 g   1277 t
ORIGIN

  Query Match           19.5%;  Score 17;  DB 16;  Length 4133;
  Best Local Similarity 100.0%;  Pred. No. 1.92e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1382 CATTTTACTATATTTTT 1398
          |||||||||||||||||
Cp     50 CATTTTACTATATTTTT 34


RESULT   21
LOCUS       YSAAOD1A     4174 bp    DNA             PLN       17-JUN-1992
DEFINITION  Candida boidinii methanol oxidase (AOD1) gene, complete cds.
ACCESSION   M81702
NID         g170819
```



```
KEYWORDS    AOD1 gene; alcohol oxidase; methanol oxidase.
SOURCE      Candida boidinii (strain S2 AOU-1) DNA.
  ORGANISM  Candida boidinii
            Eukaryotae; mitochondrial eukaryotes; Fungi; Ascomycota;
            Hemiascomycetes; Saccharomycetales; Candidaceae; Candida.
REFERENCE   1  (bases 1 to 4174)
  AUTHORS   Sakai,Y. and Tani,Y.
  TITLE     Cloning and sequencing of the alcohol oxidase-encoding gene (A)D1)
            from the formaldehyde-producing asporogenous methyklotrophic yeast,
            Candida boidinii S2
  JOURNAL   Gene 114, 67-73 (1992)
  MEDLINE   92267386
FEATURES             Location/Qualifiers
     source          1. .4174
                     /organism="Candida boidinii"
                     /strain="S2 AOU-1"
                     /db_xref="taxon:5477"
     TATA_signal     1329. .1335
                     /gene="AOD1"
                     /note="putative"
     gene            1329. .4124
                     /gene="AOD1"
     TATA_signal     1365. .1371
                     /gene="AOD1"
                     /note="putative"
     CDS             1668. .3659
                     /gene="AOD1"
                     /EC_number="1.1.3.13"
                     /codon_start=1
                     /product="methanol oxidase"
                     /db_xref="PID:g170820"
                     /translation="MAIPEEFDVIVCGGGSTGCVIAGRLANVDENLKVLLIENGENNL
                     NNPWVYLPGIYPRNMRLDSKTATPYNSRPSKHLNGRRAIVPQANILGGGGSSINFMMYT
                     RASASDYDDWESEGWTTDELLPLMKKFETYQRPCNNRDVHGFDGPIKVSFGNYTYPQC
                     QDFLRACETQGIPYVDDLEDLKTSHGAEQWLKWINRDFGRRSDTAHAFIHSTMRNKEN
                     LFLMTNTKVDEVIIEDGRAVAVETVPSKPIGDSKVSRTFKARKQIVVSCGTVSSPMVL
                     QRSGIGEPSKLRAAGVKPIVELPGVGRNFQDHFCIFVPYRIKQDSESFDAFVSGDKEA
                     QKSAFDQWYATGAGPLATNGIEAGVKIRPTEAELATADKAFQQGWESYFENKPDKPLM
                     HYSVISGFFGDHTRLPPGKYMTMFHFLEYPFSRGWLHISSDDPYAAPDFDPGFMNDDR
                     DMWPMVWAFKKSRETARRMECFAGEPTAFHPHYKVDSPARALEQSAEDTKKVAGPLHL
                     TANLYHGSWSTPIGEADKHDPNHVTSSHINVYSKDIQYTKEDDEAIENYIKEHAETTW
                     HCLGTNSMAPREGNKNAPEGGVLDPRLNVHGVKGLKVADLSVCPDNVGCNTFSTALTI
                     GEKAAVLVAEDLGYSGGSELDMEVPQHKLKTYEQTGAARY"
     terminator      4036. .4075
                     /gene="AOD1"
     polyA_signal    4119. .4124
                     /gene="AOD1"
BASE COUNT     1366 a    640 c    627 g   1541 t
ORIGIN

  Query Match           19.5%;  Score 17;  DB 23;  Length 4174;
```



```
  Best Local Similarity 100.0%;  Pred. No. 1.92e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    802 CACATTTTACTATATTT 818
          |||||||||||||||||
Cp     52 CACATTTTACTATATTT 36


RESULT   22
LOCUS       AF030372     4346 bp    DNA             BCT       16-JUL-1998
DEFINITION  Streptococcus pneumoniae strain SP-VA96 alpha, 1-6-glucosidase
            (dexB) gene, partial cds; putative regulatory protein (cpsA) and
            CpsB (cpsB) genes, complete cds; and putative chain length
            regulator (cpsC) gene, partial cds.
ACCESSION   AF030372
NID         g2804762
KEYWORDS    .
SOURCE      Streptococcus pneumoniae.
  ORGANISM  Streptococcus pneumoniae
            Eubacteria; Firmicutes; Cloatridium group firmicutes;
            Streptococcaceae; Streptococcus.
REFERENCE   1  (bases 1 to 4346)
  AUTHORS   Coffey,T.J., Enright,M.C., Daniels,M., Morona,J.K., Morona,R.,
            Hryniewicz,W., Paton,J.C. and Spratt,B.G.
  TITLE     Recombinational exchanges at the capsular polysaccharide
            biosynthetic locus lead to frequent serotype changes among natural
            isolates of Streptococcus pneumoniae
  JOURNAL   Mol. Microbiol. 27 (1), 73-83 (1998)
  MEDLINE   98125733
REFERENCE   2  (bases 1 to 4346)
  AUTHORS   Coffey,T.J., Daniels,M. and Spratt,B.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-OCT-1997) The Wellcome Trust Centre for the
            Epidemiology of Infectious Disease, Department of Zoology,
            University of Oxford, South Parks Road, Oxford, Oxfordshire OX1
            3PS, UK
REFERENCE   3  (bases 1 to 4346)
  AUTHORS   Coffey,T.J., Daniels,M. and Spratt,B.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-1998) The Wellcome Trust Centre for the
            Epidemiology of Infectious Disease, Department of Zoology,
            University of Oxford, South Parks Road, Oxford, Oxfordshire OX1
            3PS, UK
  REMARK    Sequence update by submitter
FEATURES             Location/Qualifiers
     source          1. .4346
                     /organism="Streptococcus pneumoniae"
                     /strain="SP-VA96"
                     /serotype="19F"
                     /db_xref="taxon:1313"
                     /note="clinical isolate from Spain"
```



```
     gene            <1. .439
                     /gene="dexB"
     CDS             <1. .439
                     /gene="dexB"
                     /codon_start=2
                     /transl_table=11
                     /product="alpha, 1-6-glucosidase"
                     /db_xref="PID:g2804763"
                     /translation="DSIRVIGRDNARTPMQWDESKNAGFSTDQPWLAVNPNYEMINVQ
                     EALANPDSIFTTYQKLVQIRKENSWLIRADFELLDTADKVFAYIRKDGDRRFLVVANL
                     SNEEQDLTVEGKVKSVLIENTLAQEVFEKQILVPWDAFCVELL"
     gene            1505. .2950
                     /gene="cpsA"
     CDS             1505. .2950
                     /gene="cpsA"
                     /codon_start=1
                     /transl_table=11
                     /product="putative regulatory protein"
                     /db_xref="PID:g2804764"
                     /translation="MSRRFKKSRSQKVKRSVNIVLLTIYLLLVCFLLFLIFKYNILAF
                     RYLNLVVTALVLLVALVGLLLIIYKKAEKFTIFLLVFSILVSSVSLFAVQQFVGLTNR
                     LNATSNYSEYSISVAVLADSDIENVTQLTSVTAPTGTDNENIQKLLADIKSSQNTDLT
                     VNQSSSYLAAYKSLIAGETKAIVLNSVFENIISSEYPDYASKIKKIYTKGFTKVEAP
                     KTSKNQSFNIYVSGIDTYGPISSVSRSDVNILMTVNRDTKKILLTTTPRDATVPIADG
                     GNNQKDKLTHAGIYGVDSSIHTLENLYGVDINYYVRLNFTSPLKMIDLLGGVDVHNDQ
                     EFSALHGKFHFPVGNVHLDSEQALGFVRERYSLADGDRDRGRNQQKVIVAIIQKLTST
                     EVLKNYSSILQGLQDSLQTNMPIETMIDLVNTQLESGGNYKVNSQDLKGTGRMDLPSY
                     AMPDINLYVMEIDDSSLAVVKAAIQDVMEGR"
     gene            2952. .3683
                     /gene="cpsB"
     CDS             2952. .3683
                     /gene="cpsB"
                     /codon_start=1
                     /transl_table=11
                     /product="CpsB"
                     /db_xref="PID:g2804765"
                     /translation="MIDIHSHIVFDVDDGPKSREESKALLAESYRQGVRTIVSTSHRR
                     KGMPETPEEKIAENFLQVREIAKEVADDLVIAYGAEIYYTLDALEKLEKKEIPTLNDS
                     RYALIEFSMHTSYRQIHTGLSNILMLGITPVIAHIERYDALENNEKRVRELIDMGCYT
                     QINSYHVSKPKFFGEKYKFMKKRARYFLERDLVHVVASDMHNLDSRPPYMQQAYDIIA
                     KKYGAKKAKELFVDNPRKIIMDQLI"
     gene            3692. .>4346
                     /gene="cpsC"
     CDS             3692. .>4346
                     /gene="cpsC"
                     /codon_start=1
                     /transl_table=11
                     /product="putative chain length regulator"
                     /db_xref="PID:g2804766"
                     /translation="MKEQNTLEIDVLQLFRALWKRKLVILLVAIITSSVAPAYSTFVI
                     KPEPTSMTRIYVVNRDQGEKSGLTNQDLQAGSSLVKDYREIILSQDVLEEVVSDLKLD
```

```
                     LTPKDLANKIKVTVPVDTRIVSVSVSDRVPEEASRIANSLREVAAQKIISITRVSDVT
                     TLEEARPATSPSSPNIKRSTLIGFLAGVIGTSVIVLILELLDTRVKRPKDIEDTLQMT
BASE COUNT      1379 a    748 c    881 g   1338 t
ORIGIN

  Query Match           19.5%;  Score 17;  DB 16;  Length 4346;
  Best Local Similarity 100.0%;  Pred. No. 1.92e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1382 CATTTTACTATATTTTT 1398
         |||||||||||||||||
Cp     50 CATTTTACTATATTTTT 34


RESULT   23
LOCUS       AF030368     4981 bp    DNA             BCT       16-JUL-1998
DEFINITION  Streptococcus pneumoniae strain SP-496 alpha, 1-6-glucosidase
            (dexB) gene, partial cds; putative regulatory protein (cpsA), CpsB
            (cpsB), and putative chain length regulator (cpsC) genes, complete
            cds; and CpsD (cpsD) gene, partial cds.
ACCESSION   AF030368
NID         g2804739
KEYWORDS    .
SOURCE      Streptococcus pneumoniae.
  ORGANISM  Streptococcus pneumoniae
            Eubacteria; Firmicutes; Clostridium group firmicutes;
            Streptococcaceae; Streptococcus.
REFERENCE   1  (bases 1 to 4981)
  AUTHORS   Coffey,T.J., Enright,M.C., Daniels,M., Morona,J.K., Morona,R.,
            Hryniewicz,W., Paton,J.C. and Spratt,B.G.
  TITLE     Recombinational exchanges at the capsular polysaccharide
            biosynthetic locus lead to frequent serotype changes among natural
            isolates of Streptococcus pneumoniae
  JOURNAL   Mol. Microbiol. 27 (1), 73-83 (1998)
  MEDLINE   98125733
REFERENCE   2  (bases 1 to 4981)
  AUTHORS   Coffey,T.J., Daniels,M. and Spratt,B.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-OCT-1997) The Wellcome Trust Centre for the
            Epidemiology of Infectious Disease, Department of Zoology,
            University of Oxford, South Parks Road, Oxford, Oxfordshire OX1
            3PS, UK
REFERENCE   3  (bases 1 to 4981)
  AUTHORS   Coffey,T.J., Daniels,M. and Spratt,B.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-1998) The Wellcome Trust Centre for the
            Epidemiology of Infectious Disease, Department of Zoology,
            University of Oxford, South Parks Road, Oxford, Oxfordshire OX1
            3PS, UK
  REMARK    Sequence update by submitter
```



```
FEATURES             Location/Qualifiers
     source          1. .4981
                     /organism="Streptococcus pneumoniae"
                     /strain="SP-496"
                     /serotype="19F"
                     /db_xref="taxon:1313"
                     /note="clinical isolate from Spain"
     gene            <1. .439
                     /gene="dexB"
     CDS             <1. .439
                     /gene="dexB"
                     /codon_start=2
                     /transl_table=11
                     /product="alpha, 1-6-glucosidase"
                     /db_xref="PID:g2804740"
                     /translation="DSIRVIGRDNARTPMQWDESKNAGFSTGQPWLAVNPNYEMINVQ
                     EALANPDSIFTTYQKLVQIRKENSWLIRADFELLDTADKVFAYIRKDGDRRFLVVANL
                     SNEEQDLTVEGKVKSVLIENTLAQEVFEKQILVPWDAFCVELL"
     gene            1505. .2950
                     /gene="cpsA"
     CDS             1505. .2950
                     /gene="cpsA"
                     /codon_start=1
                     /transl_table=11
                     /product="putative regulatory protein"
                     /db_xref="PID:g2804741"
                     /translation="MSRRFKKSRSQKVKRSVNIVLLTIYLLLVCFLLFLIFKYNILAF
                     RYLNLVVTALVLLVALVGLLLIIYKKAEKFTIFLLVFSILVSSVSLFAVQQFVGLTNR
                     LNATSNYSEYSISVAVLADSDIENVTQLTSVTAPTGTDNENIQKLLADIKSSQNTDLT
                     VNQSSSYLAAYKSLIAGETKAIVLNSVFENIIESEYPDYASKIKKIYTKGFTKKVEAP
                     KTSKNQSFNIYVSGIDTYGPISSVSRSDVNILMTVNRDTKKILLTTTPRDAYVPIADG
                     GNNQKDKLTHAGIYGVDSSIHTLENLYGVDINYYVRLNFTSFLKMIDLLGGVDVHNDQ
                     EFSALHGKFHFPVGNVHLDSEQALGFVRERYSLADGDRDRGRNQQKVIVAIIQKLTST
                     EVLENYSSILQGLQDSLQTNMPIETMIDLVNTQLESGGNYKVNSQDLKGTGRMDLPSY
                     AMPDSNLYVMEIDDSSLAVVKAAIQDVMEGR"
     gene            2952. .3683
                     /gene="cpsB"
     CDS             2952. .3683
                     /gene="cpsB"
                     /codon_start=1
                     /transl_table=11
                     /product="CpsB"
                     /db_xref="PID:g2804742"
                     /translation="MIDIHSHIVFDVDDGPKSREESKALLAESYRQGVRTIVSTSHRR
                     KGMFETPEEKIAENFLQVREIAKEVADDLVIAYGAEIYYTLDALEKLEKKEIPTLNDS
                     RYALIEFSMHTSYRQIHTGLSNILMLGITPVIAHIERYDALENNEKRVRELIDMGCYT
                     QINSYHVSKPKFFGEKYKFMKKRARYFLERDLVHVVASDMHNLDSRPPYMQQAYDIIA
                     KKYGAKKAKELFVDNPRKIIMDQLI"
     gene            3692. .4384
                     /gene="cpsC"
     CDS             3692. .4384
```



```
                     /gene="cpsC"
                     /codon_start=1
                     /transl_table=11
                     /product="putative chain length regulator"
                     /db_xref="PID:g2804743"
                     /translation="MKEQNTLEIDVLQLFRALWKRKLVILLVAIITSSVAFAYSTFVI
                     KPEPTSMTRIYVVNRDQGEKSGLTNQDLQAGSSLVKDYREIILSQDVLEEVVSDLKLD
                     LTPKDLANKIKVTVPVDTRIVSVSVSDRVPEEASRIANSLREVAAQKIISITRVSDVT
                     TLEEARPATSPSSPNIKRSTLIGFLAGVIGTSVIVLILELLDTRVKRPKDIEDTLQMT
                     LLGIVPNLNKLK"
     gene            4394. .>4981
                     /gene="cpsD"
     CDS             4394. .>4981
                     /gene="cpsD"
                     /codon_start=1
                     /transl_table=11
                     /product="CpsD"
                     /db_xref="PID:g2804744"
                     /translation="MPTLEIAQKKLEFIKKAEEYYNALCTNIQLSGDKLKVISVTSVN
                     PGEGKTTTSVNIARSFARAGYKTLLIDGDTRNSVMSGFFKSREKITGLTEFLSGTADL
                     SHGLCDTNIENLFVVQSGTVSPNPTALLQSKNFNDMIETLRKYFDYIIVDTAPIGIVI
                     DAAIITQKCDASILVTATGEVNKRDVQKAKQQLEQT"
BASE COUNT      1606 a    845 c   1009 g   1521 t
ORIGIN

  Query Match           19.5%;  Score 17;  DB 16;  Length 4981;
  Best Local Similarity 100.0%;  Pred. No. 1.92e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1382 CATTTTACTATATTTTT 1398
         |||||||||||||||||
Cp     50 CATTTTACTATATTTTT 34


RESULT   24
LOCUS       AF030370     4981 bp    DNA             BCT       16-JUL-1998
DEFINITION  Streptococcus pneumoniae strain SP-GA71 alpha, 1-6-glucosidase
            (dexB) gene, partial cds; putative regulatory protein (cpsA), CpsB
            (cpsB), and putative chain length regulator (cpsC) genes, complete
            cds; and CpsD (cpsD) gene, partial cds.
ACCESSION   AF030370
NID         g2804750
KEYWORDS    .
SOURCE      Streptococcus pneumoniae.
  ORGANISM  Streptococcus pneumoniae
            Eubacteria; Firmicutes; Clostridium group firmicutes;
            Streptococcaceae; Streptococcus.
REFERENCE   1  (bases 1 to 4981)
  AUTHORS   Coffey,T.J., Enright,M.C., Daniels,M., Morona,J.K., Morona,R.,
            Hryniewicz,W., Paton,J.C. and Spratt,B.G.
  TITLE     Recombinational exchanges at the capsular polysaccharide
```



```
            biosynthetic locus lead to frequent serotype changes among natural
            isolates of Streptococcus pneumoniae
  JOURNAL   Mol. Microbiol. 27 (1), 73-83 (1998)
  MEDLINE   98125733
REFERENCE   2  (bases 1 to 4981)
  AUTHORS   Coffey,T.J., Daniels,M. and Spratt,B.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-OCT-1997) The Wellcome Trust Centre for the
            Epidemiology of Infectious Disease, Department of Zoology,
            University of Oxford, South Parks Road, Oxford, Oxfordshire OX1
            3PS, UK
REFERENCE   3  (bases 1 to 4981)
  AUTHORS   Coffey,T.J., Daniels,M. and Spratt,B.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-1998) The Wellcome Trust Centre for the
            Epidemiology of Infectious Disease, Department of Zoology,
            University of Oxford, South Parks Road, Oxford, Oxfordshire OX1
            3PS, UK
  REMARK    Sequence update by submitter
FEATURES             Location/Qualifiers
     source          1. .4981
                     /organism="Streptococcus pneumoniae"
                     /strain="SP-GA71"
                     /serotype="19F"
                     /db_xref="taxon:1313"
                     /note="clinical isolate from Spain"
     gene            <1. .439
                     /gene="dexB"
     CDS             <1. .439
                     /gene="dexB"
                     /codon_start=2
                     /transl_table=11
                     /product="alpha, 1-6-glucosidase"
                     /db_xref="PID:g2804751".
                     /translation="DSIRVIGRDNARTPMQWDESKNAGFSTGQPWLAVNPNYEMINVQ
                     EALANPDSIFTTYQKLVQIRKENSWLIRADFELLDTADKVFAYIRKDGDRRFLVVANL
                     SNEEQDLTVEGKVKSVLIENTLAQEVFEKQILVPWDAFCVELL"
     gene            1505. .2950
                     /gene="cpsA"
     CDS             1505. .2950
                     /gene="cpsA"
                     /codon_start=1
                     /transl_table=11
                     /product="putative regulatory protein"
                     /db_xref="PID:g2804752"
                     /translation="MSRRFKKSRSQKVKRSVNIVLLTIYLLLVCFLLFLIFKYNILAF
                     RYLNLVVTALVLLVALVGLLLIIYKKAEKFTIFLLVFSILVSSVSLFAVQQFVGLTNR
                     LNATSNYSEYSISVAVLADSDIENVTQLTSVTAPTGTDNENIQKLLADIKSSQNTDLT
                     VNQSSSYLAAYKSLIAGETKAIVLNSVFENIIESEYPDYASKIKKIYTKGFTKKVEAP
                     KTSKNQSFNIYVSGIDTYGPISSVSRSDVNILMTVNRDTKKILLTTTPRDAYVPIADG
                     GNNQKDKLTHAGIYGVDSSIHTLENLYGVDINYYVRLNFTSFLKMIDLLGGVDVHNDQ
```

```
                     EFSALHGKFHFPVGNVHLDSEQALGFVRERYSLADGDRDRGRNQQKVIVAIIQKLTST
                     EVLKNYSSILQGLQDSLQTNMPIETMIDLVNTQLESGGNYKVNSQDLKGTGRMDLPSY
                     AMPDSNLYVMEIDDSSLAVVKAAIQDVMEGR"
     gene            2952. .3683
                     /gene="cpsB"
     CDS             2952. .3683
                     /gene="cpsB"
                     /codon_start=1
                     /transl_table=11
                     /product="CpsB"
                     /db_xref="PID:g2804753"
                     /translation="MIDIHSHIVPDVDDGPKSREESKALLAESYRQGVRTIVSTSHRR
                     KGMFETPEEKIAENFLQVREIAKEVADDLVIAYGAEIYYTLDALEKLEKKEIPTLNDS
                     RYALIEFSMHTSYRQIHTGLSNILMLGITPVIAHIERYDALENNEKRVRELIDMGCYT
                     QINSYHVSKPKFPGEKYKFMKKRARYFLERDLVHVVASDMHNLDSRPPYMQQAYDIIA
                     KKYGAKKAKELFVDNPRKIIMDQLI"
     gene            3692. .4384
                     /gene="cpsC"
     CDS             3692. .4384
                     /gene="cpsC"
                     /codon_start=1
                     /transl_table=11
                     /product="putative chain length regulator"
                     /db_xref="PID:g2804754"
                     /translation="MKEQNTLEIDVLQLFRALWKRKLVILLVAIITSSVAFAYSTFVI
                     KPEFTSMTRIYVVNRDQEKSGLTNQDLQAGSSLVKDYREIILSQDVLEEVVSDLKLD
                     LTPKDLANKIKVTVPVDTRIVSVSVSDRVPEEASRIANSLREVAAQKIISITRVSDVT
                     TLEEARPATSPSSPNIKRSTLIGFLAGVIGTSVIVLILELLDTRVKRPKDIEDTLQMT
                     LLGIVPNLNKLK"
     gene            4394. .>4981
                     /gene="cpsD"
     CDS             4394. .>4981
                     /gene="cpsD"
                     /codon_start=1
                     /transl_table=11
                     /product="CpsD"
                     /db_xref="PID:g2804755"
                     /translation="MPTLEIAQKKLEPIKKAEEYYNALCTNIQLSGDKLKVISVTSVN
                     PGEGKTTTSVNIARSFARAGYKTLLIDGDTRNSVMSGFFKSREKITGLTEFLSGTADL
                     SHGLCDTNIENLFVVQSGTVSPNPTALLQSKNFNDMIETLRKYFDYIIVDTAPIGIVI
                     DAAIITQKCDASILVTATGEVNKRDVQKAKQQLEQT"
BASE COUNT     1606 a    845 c   1009 g   1521 t
ORIGIN

  Query Match           19.5%;  Score 17;  DB 16;  Length 4981;
  Best Local Similarity 100.0%;  Pred. No. 1.92e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1382 CATTTTACTATATTTTT 1398
           |||||||||||||||||
Cp     50 CATTTTACTATATTTTT 34
```



```
RESULT   25
LOCUS       AF007562     5300 bp    DNA             PRI       19-MAR-1998
DEFINITION  Homo sapiens trabecular meshwork inducible glucocorticoid response
            protein (TIGR) gene, promoter region and partial mRNA sequence.
ACCESSION   AF007562
NID         g2970123
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5300)
  AUTHORS   Nguyen,T.D., Chen,P., Huang,W.D., Chen,H., Johnson,D. and
            Polansky,J.R.
  TITLE     Gene structure and properties of TIGR, an olfactomedin-related
            glycoprotein cloned from glucocorticoid-induced trabecular meshwork
            cells
  JOURNAL   J. Biol. Chem. 273 (11), 6341-6350 (1998)
  MEDLINE   98165818
REFERENCE   2  (bases 1 to 5300)
  AUTHORS   Nguyen,T.D., Chen,P., Chen,H. and Polansky,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUN-1997) Ophthalmology, University of California San
            Francisco, 10 Kirkham Street, San Francisco, CA 94143-0730, USA
FEATURES             Location/Qualifiers
     source          1. .5300
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     gene            1. .>5300
                     /gene="TIGR"
     promoter        1. .5271
                     /gene="TIGR"
     mRNA            5272. .>5300
                     /gene="TIGR"
                     /note="additional 3' sequence deposited as TIGR mRNA with
                     GenBank Accession Number U85257"
                     /product="trabecular meshwork inducible glucocorticoid
                     response protein"
BASE COUNT     1483 a   1152 c   1236 g   1429 t
ORIGIN

  Query Match           19.5%;  Score 17;  DB 27;  Length 5300;
  Best Local Similarity 100.0%;  Pred. No. 1.92e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   3613 TACTATATTTTTCCTTT 3629
           |||||||||||||||||
Cp     45 TACTATATTTTTCCTTT 29
```



```
RESULT   26
LOCUS       AF030371     5437 bp    DNA             BCT       16-JUL-1998
DEFINITION  Streptococcus pneumoniae strain PO-329 alpha, 1-6-glucosidase
            (dexB) gene, partial cds; putative regulatory protein (cpsA), CpsB
            (cpsB), and putative chain length regulator (cpsC) genes, complete
            cds; and CpsD (cpsD) gene, partial cds.
ACCESSION   AF030371
NID         g2804756
KEYWORDS    .
SOURCE      Streptococcus pneumoniae.
  ORGANISM  Streptococcus pneumoniae
            Eubacteria; Firmicutes; Clostridium group firmicutes;
            Streptococcaceae; Streptococcus.
REFERENCE   1  (bases 1 to 5437)
  AUTHORS   Coffey,T.J., Enright,M.C., Daniels,M., Morona,J.K., Morona,R.,
            Hryniewicz,W., Paton,J.C. and Spratt,B.G.
  TITLE     Recombinational exchanges at the capsular polysaccharide
            biosynthetic locus lead to frequent serotype changes among natural
            isolates of Streptococcus pneumoniae
  JOURNAL   Mol. Microbiol. 27 (1), 73-83 (1998)
  MEDLINE   98125733
REFERENCE   2  (bases 1 to 5437)
  AUTHORS   Coffey,T.J., Daniels,M. and Spratt,B.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-OCT-1997) The Wellcome Trust Centre for the
            Epidemiology of Infectious Disease, Department of Zoology,
            University of Oxford, South Parks Road, Oxford, Oxfordshire OX1
            3PS, UK
REFERENCE   3  (bases 1 to 5437)
  AUTHORS   Coffey,T.J., Daniels,M. and Spratt,B.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-1998) The Wellcome Trust Centre for the
            Epidemiology of Infectious Disease, Department of Zoology,
            University of Oxford, South Parks Road, Oxford, Oxfordshire OX1
            3PS, UK
  REMARK    Sequence update by submitter
FEATURES             Location/Qualifiers
     source          1. .5437
                     /organism="Streptococcus pneumoniae"
                     /strain="PO-329"
                     /serotype="19F"
                     /db_xref="taxon:1313"
                     /note="clinical isolate from Poland"
     gene            <1. .439
                     /gene="dexB"
     CDS             <1. .439
                     /gene="dexB"
                     /codon_start=2
                     /transl_table=11
                     /product="alpha, 1-6-glucosidase"
```



```
                     /db_xref="PID:g2804757"
                     /translation="DSIRVIGRDNARTPMQWDESKNAGFSTGQPWLAVNPNYEMINVQ
                     EALANPDSIFYTTQKLVQIRKENSWLIRADFELLDTADKVFAYIRKDGDRRFLVVANL
                     SNEEQDLTVEGKVKSVLIENTLAQEVFEKQILVPWDAFCVELL"
     misc_feature    976. .1080
                     /note="indel"
     misc_feature    1392. .1744
                     /note="indel"
     gene            1960. .3405
                     /gene="cpsA"
     CDS             1960. .3405
                     /gene="cpsA"
                     /codon_start=1
                     /transl_table=11
                     /product="putative regulatory protein"
                     /db_xref="PID:g2804758"
                     /translation="MSRRFKKSRSQKVKRSVNIVLLTIYLLLVCFLLFLIFKYNILAF
                     RYLNLVVTALVLLVALVGLLLIIYKKAEKFTIFLLVFSILVSSVSLFAVQQFVGLTNR
                     LNATSNYSEYSISVAVLADSDIENVTQLTSVTAPTGTDNENIQKLLADIKSSQNTDLT
                     VNQSSSYLAAYKSLIAGETKAIVLNSVFENIIESEYPDYASKIKKIYTKGFTKKVEAP
                     KTSKNQSFNIYVSGIDTTGPISSVSRSDVNILMTVNRDTKKILLTTTPRDAYVPIADG
                     GNNQKDKLTHAGIYGVDSSIHTLENLYGVDINYYVRLNFTSFLKMIDLLGGVDVHNDQ
                     EFSALHGKFHFPVGNVHLDSEQALGFVRERYSLADGDRDRGRNQQKVIVAIIQKLTST
                     EVLKNYSSILQGLQDSLQTNMPIETMIDLVNTQLESGGNYKVNSQDLKGTGRMDLPSY
                     AMPDSNLYVMEIDDSSLAVVKAAIQDVMEGR"
     gene            3407. .4138
                     /gene="cpsB"
     CDS             3407. .4138
                     /gene="cpsB"
                     /codon_start=1
                     /transl_table=11
                     /product="CpsB"
                     /db_xref="PID:g2804759"
                     /translation="MIDIHSHIVPDVDDGPKSREESKALLAESYRQGVRTIVSTSHRR
                     KGMFETPEEKIAENFLQVREIAKEVADDLVIAYGAEIYYTLDALEKLEKKEIPTLNDS
                     RYALIEFSMHTSYRQIHTGLSNILMLGITPVIAHIERYDALENNEKRVRELIDMGCYT
                     QINSYHVSKPKFPGEKYKFMKKRARYFLERDLVHVVASDMHNLDSRPPYMQQAYDIIA
                     KKYGAKKAKELFVDNPRKIIMDQLI"
     gene            4147. .4839
                     /gene="cpsC"
     CDS             4147. .4839
                     /gene="cpsC"
                     /codon_start=1
                     /transl_table=11
                     /product="putative chain length regulator"
                     /db_xref="PID:g2804760"
                     /translation="MKEQNTLEIDVLQLFRALWKRKLVILLVAIITSSVAFAYSTFVI
                     KPEFTSMTRIYVVNRDQEKSGLTNQDLQAGSSLVKDYREIILSQDVLEEVVSDLKLD
                     LTPKDLANKIKVTVPVDTRIVSVSVSDRVPEEASRIANSLREVAAQKIISITRVSDVT
                     TLEEARPATSPSSPNIKRSTLIGFLAGVIGTSVIVLILELLDTRVKRPKDIEDTLQMT
                     LLGIVPNLNKLK"
```

```
     gene            4849. .>5437
                     /gene="cpsD"
     CDS             4849. .>5437
                     /gene="cpsD"
                     /codon_start=1
                     /transl_table=11
                     /product="CpsD"
                     /db_xref="PID:g2804761"
                     /translation="MPTLEIAQKKLEFVKKAEEYYNALCTNIQLSGDKLKVISVTSVN
                     PGEGKTTTSVNIARSFARAGYKTLLIDGDIRNSVMSGFFKSREKITGLTEFLSGTADL
                     SHGLCDTNIENLFVVQSGTVSPNPTALLQSKNFNDMIETLRKYFDYIIVDTAPIGIVI
                     DAAIITQKCDASILVTATGEVNRRDVQKAKQQLEQT"
BASE COUNT     1721 a    916 c   1099 g   1700 t      1 others
ORIGIN

  Query Match           19.5%;  Score 17;  DB 16;  Length 5437;
  Best Local Similarity 100.0%;  Pred. No. 1.92e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1837 CATTTTACTATATTTTT 1853
           |||||||||||||||||
Cp     50 CATTTTACTATATTTTT 34


RESULT  27
LOCUS       AF030367     6798 bp    DNA             BCT       16-JUL-1998
DEFINITION  Streptococcus pneumoniae strain NCTC11906 alpha, 1-6-glucosidase
            (dexB) gene, partial cds; maturase-related protein, putative
            regulatory protein (cpsA), CpsB (cpsB), and putative chain length
            regulator (cpsC) genes, complete cds; and CpsD (cpsD) gene, partial
            cds.
ACCESSION   AF030367
NID         g2804732
KEYWORDS    .
SOURCE      Streptococcus pneumoniae.
  ORGANISM  Streptococcus pneumoniae
            Eubacteria; Firmicutes; Clostridium group firmicutes;
            Streptococcaceae; Streptococcus.
REFERENCE   1  (bases 1 to 6798)
  AUTHORS   Coffey,T.J., Enright,M.C., Daniels,M., Morona,J.K., Morona,R.,
            Hryniewicz,W., Paton,J.C. and Spratt,B.G.
  TITLE     Recombinational exchanges at the capsular polysaccharide
            biosynthetic locus lead to frequent serotype changes among natural
            isolates of Streptococcus pneumoniae
  JOURNAL   Mol. Microbiol. 27 (1), 73-83 (1998)
  MEDLINE   98125733
REFERENCE   2  (bases 1 to 6798)
  AUTHORS   Coffey,T.J., Daniels,M. and Spratt,B.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-OCT-1997) The Wellcome Trust Centre for the
            Epidemiology of Infectious Disease, Department of Zoology,
```



```
            University of Oxford, South Parks Road, Oxford, Oxfordshire OX1
            3PS, UK
REFERENCE   3  (bases 1 to 6798)
  AUTHORS   Coffey,T.J., Daniels,M. and Spratt,B.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-1998) The Wellcome Trust Centre for the
            Epidemiology of Infectious Disease, Department of Zoology,
            University of Oxford, South Parks Road, Oxford, Oxfordshire OX1
            3PS, UK
  REMARK    Sequence update by submitter
FEATURES            Location/Qualifiers
     source         1. .6798
                    /organism="Streptococcus pneumoniae"
                    /strain="NCTC11906"
                    /serotype="19F"
                    /db_xref="taxon:1313"
                    /note="from National Collection of Type Cultures"
     gene           <1. .439
                    /gene="dexB"
     CDS            <1. .439
                    /gene="dexB"
                    /codon_start=2
                    /transl_table=11
                    /product="alpha, 1-6-glucosidase"
                    /db_xref="PID:g2804733"
                    /translation="DSIRVIGRDNARTPMQWDESKNAGFSTGQPWLAVNPNYEMINVQ
                    EALANPDSIFTTYQKLVQIRKENSWLIRADFELLDTADKVFAYIRKDGDRRFLVVANL
                    SNEEQDLTVEGKVKSVLIENTLAQEVPEKQILVPWDAPCVELL"
     CDS            1412. .2689
                    /note="similar to the maturase of group II introns"
                    /codon_start=1
                    /transl_table=11
                    /product="maturase-related protein"
                    /db_xref="PID:g2804734"
                    /translation="MSKLLDKILSRENMLEAYNQVKSNKGSAGIDGMTIEEMDNYLRQ
                    NWRLTKELIKQRKYKPQPVLKVEIPKPDGGIRQLGIPTVMDRMIQQAIVQVMSPICEP
                    HFSDTSYGFRPNRSCEKAIMKLLEYLNDGYEWIVDIDLEKFFDTVPQDRLMSLVHNII
                    EDGDTESLIRKYLHSGVIINGQRYRTLVGTPQGGNLSPLLSNIMLNELDKELEKRGLR
                    FVRYADDCVITVGSEAAAKRVMYSVSRFIEKRLGLKVNMTKTKITRPRELKYLGFGFW
                    KSSDGWKSRPHQDSVRRFKLKLKKLTQRKWSIDLTRRIEQLNLSIRGWINYFSLGNMK
                    SIVASIDERLRTRLRMIIWKQWKKKSRRLWGLLKLGVPKWIADKVSGWGDHYQLVAQK
                    SVLKRAISKPVLERKGLVSCLDYYLERHALKVS"
     misc_feature   3689. .4152
                    /note="indel"
     gene           3774. .5219
                    /gene="cpsA"
     CDS            3774. .5219
                    /gene="cpsA"
                    /codon_start=1
                    /transl_table=11
                    /product="putative regulatory protein"
```



```
                    /db_xref="PID:g2804735"
                    /translation="MSRRFKKSRSQKVKRSVNIVLLTIYLLLVCFLLFLIFKYNILAF
                    RYLNLVVTALVLLVALVGLLLIIFKKAEKFTIFLLVFSILVSSVSLSFAVQQFVGLTNR
                    LWATSNYSEYSISVAVLADSDIENVTQLTSVTAPTGTDNENIQKLLADIKSSQNTDLT
                    VNQSSSYLAAYKSLIAGETKAIVLNSVFENIIESEYPDTASKIKKIYTKGFTKKVEAP
                    KTSKNQSFNIYVSGIDTTGPISSVSRSDVNILMTVNRDTKKILLTTTPRDAYVPIADG
                    GNNQKDKLTHAGIYGVDSSIHTLENLYGVDINYYVRLNFTSPLKMIDLLGGVDVHNDQ
                    EFSALHGKFHFPVGNVHLDSEQALGFVRERYSLADGDRDRGRNQQKVIVAIIQKLTST
                    EVLKNYSSILQGLQDSLQTNMPIETMIDLVNTQLESGGNYKVNSQDLKGTGRMDLPSY
                    AMPDSNLYVMEIDDSSLAVVKAAIQDVMEGR"
     gene           5221. .5952
                    /gene="cpsB"
     CDS            5221. .5952
                    /gene="cpsB"
                    /codon_start=1
                    /transl_table=11
                    /product="CpsB"
                    /db_xref="PID:g2804736"
                    /translation="MIDIHSHIVFDVDDGPKSREESKALLAESYRQGVRTIVSTSHRR
                    KGMFETPEEKIAENFLQVREIAKEVADDLVIAYGAEIYYTLDALEKLEKKEIPTLNDS
                    RYALIEFSMHTSYRQIHTGLSNILMLGITPVIAHIERYDALENNEKRVRELIDMGCYT
                    QINSIHVSKPKFPGEKYKFMKKRARYFLERDLVHVVASDMHNLDSRPPYMQQAYDIIA
                    KKYGAKKVKELFVDNPRKIIMDQLI"
     gene           5961. .6653
                    /gene="cpsC"
     CDS            5961. .6653
                    /gene="cpsC"
                    /codon_start=1
                    /transl_table=11
                    /product="putative chain length regulator"
                    /db_xref="PID:g2804737"
                    /translation="MKEQNTLEIDVLQLFRALWKRELVILLVAIITSSVAFAYSTFVI
                    KPEFTSMTRIVVNRNQGEKSGLTNQDLQAGSSLVKDYREIILSQDVLEEVVSDLKLD
                    LTPKDLANKIKVTVPVDTRIVSVSVSDRVPEEASRIANSLREVVAAQKIISITRVSDVT
                    TLEEARPATSPSSPNIKRSTLIGFLAGVIGTSVIVLILELLDTRVKRPKDIEDTLWMT
                    LLGIVPNLNKLK"
     gene           6663. .>6798
                    /gene="cpsD"
     CDS            6663. .>6798
                    /gene="cpsD"
                    /codon_start=1
                    /transl_table=11
                    /product="CpsD"
                    /db_xref="PID:g2804738"
                    /translation="MPTLEIAQKKLEFVKKAEEYYNALCTNIQLSGDKLKVISVTSVN
                    P"
BASE COUNT     2166 a   1167 c   1455 g   2010 t
ORIGIN

  Query Match           19.5%;  Score 17;  DB 16;  Length 6798;
  Best Local Similarity 100.0%;  Pred. No. 1.92e+01;
```



```
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   3651 CATTTTACTATATTTTT 3667
           |||||||||||||||||
Cp     50 CATTTTACTATATTTTT 34


RESULT  28
LOCUS       SPDEXCAP    14056 bp    DNA             BCT       04-NOV-1996
DEFINITION  S.pneumoniae dexB, cap3A, cap3B and cap3C genes and orfs.
ACCESSION   Z47210
NID         g1658316
KEYWORDS    alpha,1-6-glucosidase; cap3A gene; cap3B gene; cap3C gene;
            cellubiuronic acid synthase; dexB gene; glucosephosphate
            uridylyltransferase; uridine diphosphate glucose dehydrogenase.
SOURCE      Streptococcus pneumoniae.
  ORGANISM  Streptococcus pneumoniae
            Eubacteria; Firmicutes; Low G+C gram-positive bacteria;
            Streptococcaceae; Streptococcus.
REFERENCE   1  (bases 1 to 14056)
  AUTHORS   Garcia,E., Arrecubieta,C., Munoz,R., Mollerach,M. and Lopez,R.
  TITLE     A functional analysis of the Streptococcus pneumoniae genes
            involved in the synthesis of type 1 and type 3 capsular
            polysaccharides
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 9704)
  AUTHORS   Arrecubieta,C., Garcia,E. and Lopez,R.
  TITLE     Sequence and transcriptional analysis of a DNA region involved in
            the production of capsular polysaccharide in Streptococcus
            pneumoniae type 3
  JOURNAL   Gene 167 (1-2), 1-7 (1995)
  MEDLINE   96144241
REFERENCE   3  (bases 5911 to 7095)
  AUTHORS   Arrecubieta,C., Lopez,R. and Garcia,E.
  TITLE     Molecular characterization of cap3A, a gene from the operon
            required for the synthesis of the capsule of Streptococcus
            pneumoniae type 3: sequencing of mutations responsible for the
            unencapsulated phenotype and localization of the capsular cluster
            on the pneumococcal chromosome
  JOURNAL   J. Bacteriol. 176 (20), 6375-6383 (1994)
  MEDLINE   95014083
REFERENCE   4  (bases 1 to 14056)
  AUTHORS   Garcia,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JAN-1995) Garcia E., Centro de Investigaciones
            Biologicas. CSIC., Microbiologia Molecular, Velazquez, 144, Madrid,
            Madrid, SPAIN, 28006
  REMARK    Revised by [5]
REFERENCE   5  (bases 1 to 14056)
  AUTHORS   Garcia,E.
  TITLE     Direct Submission
```

```
  JOURNAL   Submitted (11-SEP-1996) Garcia E., Centro de Investigaciones
            Biologicas. CSIC., Microbiologia Molecular, Velazquez, 144, Madrid,
            Madrid, SPAIN, 28006
COMMENT     On Nov 5, 1996 this sequence version replaced gi:695741.
FEATURES             Location/Qualifiers
     source          1. .14056
                     /organism="Streptococcus pneumoniae"
                     /strain="406"
                     /db_xref="taxon:1313"
     gene            1. .697
                     /gene="dexB"
     CDS             <1. .697
                     /gene="dexB"
                     /standard_name="glucosidase"
                     /function="degradation of glucans"
                     /citation=[1]
                     /citation=[2]
                     /codon_start=2
                     /transl_table=11
                     /product="alpha, 1-6-glucosidase"
                     /db_xref="PID:g728717"
                     /translation="NSLFWNNHDLPRIVSIWGNDQEYREKSAKAFAILLHLMRGTPYI
                     YQGEEIGMTNYPFETLDQVEDIESLNYPREALEKGVPIEEIMDSIRVIGRDNARTPMQ
                     WDESKNAGFSTGQPWLAVNPNYEMINVQEALANPDSIFYTYQKLVQIRKENSWLIRAD
                     FELLDTADKVFAYIRKDGDRRFLVVANLSNEEQDLTVEGKVKSVLIENTLAQEVFEKQ
                     ILVPWDAFCVELL"
     terminator      1042. .1087
                     /citation=[1]
                     /citation=[2]
     misc_feature    1658. .1689
                     /citation=[2]
                     /function="recognition sequence for the insertion sequence
                     IS1202"
     RBS             3639. .3643
                     /citation=[2]
     gene            3649. .4341
                     /gene="orf1"
     CDS             3649. .4341
                     /gene="orf1"
                     /note="The deduced amino acid sequence contains a sugar
                     transport signature (LIGFLAGGIGTSVIVLLLELLDTHVK) typical
                     of glucose transporters"
                     /citation=[1]
                     /citation=[2]
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /db_xref="PID:g695742"
                     /translation="MKEQNTLEIDVLQLFRALWKRKLVILLVAIITSSVAFAYSTFVI
                     KPEFTSTTRIYVVNRNQGEKSGLTNQDLQAGSYLVKDYREIILSQDVLEEVISDLKLD
                     LTPKGLANKIKVTVPVDTRIVSVSVNDRVPEEASRIANSLREVAAQKIISITRVSDVA
```



```
                     TLEEARPAISPSSPNIKRNTLIGFLAGGIGTSVIVLLLELLDTHVKRPEDIEDTLQMT
                     LLGVVPNLGKLK"
     RBS             4340. .4344
                     /citation=[2]
     gene            4351. .4815
                     /gene="orf2"
     CDS             4351. .4815
                     /gene="orf2"
                     /citation=[1]
                     /citation=[2]
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /db_xref="PID:g695743"
                     /translation="MPTLEIAQKKLEFIKKAEEYYNALCTNIQLSGDKLKVISVTSVN
                     PGEGKTTTSVNIARSFARAGYKTLLIDGDTRNSVMSGFFKSREKITELTEFLSGTADL
                     SHGICDTNIENLFVVQSGSVSPNPTALLQSKNFVVMVKIFSKEVSILSWWKR"
     terminator      4851. .4903
                     /citation=[1]
                     /citation=[2]
     gene            5819. .7095
                     /gene="cap3A"
     -35_signal      5819. .5824
                     /gene="cap3A"
                     /citation=[1]
                     /citation=[2]
                     /citation=[3]
                     /evidence=experimental
     -10_signal      5841. .5846
                     /gene="cap3A"
                     /citation=[1]
                     /citation=[2]
                     /citation=[3]
                     /evidence=experimental
     mRNA            5853. .>7095
                     /gene="cap3A"
                     /citation=[1]
                     /citation=[2]
                     /evidence=experimental
     RBS             5893. .5900
                     /gene="cap3A"
                     /citation=[2]
     CDS             5911. .7095
                     /gene="cap3A"
                     /standard_name="UDP-Glc dehydrogenase"
                     /function="synthesis of UDP-glucuronic acid from
                     UDP-glucose"
                     /note="The Cap3A protein has been biochemically
                     characterized, J. Bacteriol. 178:2971-2974(1996)"
                     /citation=[1]
                     /citation=[2]
```



```
                     /citation=[3]
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="uridine diphosphate glucose dehydrogenase"
                     /db_xref="PID:e280488"
                     /db_xref="PID:g1658317"
                     /translation="MKIAIAGSGYVGLSLAVLLAQHHEVKVIDVIKDKVESINNRKSP
                     IKDEAIEKYLVEKELNLEASLDPAHVYKDVEYAIIATPTNYDVDLNQFDTSSVEAAIK
                     TCMEYNDTCTIVIKSTIPEGYTKEVREKFNTDRIIFSPEFLRESKALYDNLYPSRIVV
                     GTDLDDSELTKRAWQFADLLKGGAIKEEVPILVVAFNEAEVAKLFSNTYLATRVRIFN
                     EIDTYSEVKGLNPKTIIDIVCYDPRIGSYYNNPSFGYGGYCLPKDTKQLKASFRDVPE
                     NLITAVVQSNKTRKDYIAGAILAKQPSVVGIYRLIMKSDSDNFRSSAVKGVMERLDNY
                     GKEIVIYEPTIECDTFMGYRVIKSLDEFKNISDIVVANRMNDDLRDIQEKLYTRDLFG
                     RE"
     gene            7094. .8364
                     /gene="cap3B"
     RBS             7094. .7099
                     /gene="cap3B"
                     /citation=[2]
     CDS             7111. .8364
                     /gene="cap3B"
                     /function="Synthesis of type 3 pneumococcal capsule"
                     /note="Expression and characterization of this gene has
                     been recently reported, J. Exp. Med. 184:449-455(1996)"
                     /citation=[1]
                     /citation=[2]
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="Synthase of the type 3 pneumococcal capsular
                     polysaccharide"
                     /db_xref="PID:e280489"
                     /db_xref="PID:g1658318"
                     /translation="MYTFILMLLDFFQNNDFHFFMLFFVFILIRWAVIYFHAVRYKSY
                     SCSVSDKLFSSVIIPVVDEPLNLFESVLRRISRHKPSEIIVVINGPKNERLVKLCHD
                     FNEKLENNMTPIQCYYTPVPGKRNAILPGLEHVDSQSDITVLVDSDTVWTPRTLSELL
                     KPFVCDKKIGGVTTRQKILDPERNLVTMPANLLEEIRAEGTMKAMSVTGKVGCLPGRT
                     IAFRTEILRECIHEFMNETFMGFHKEVSDDRSLTNLTLKKGYKTVMQDTSVVYTDAPT
                     SWKKFIRQQLRWAEGSQYNNLKMTPWMIRNAPLMFFIYFTDMILPMLLISFGVNIPLL
                     KILNITTIVYTASWWEIILYVLLGMIFSPGGRNFKAMSRMKWYYVFLIPVFIIVLSII
                     MCPIRLLGLMRCSDDLGWGTRNLTE"
     gene            8651. .9582
                     /gene="cap3C"
     RBS             8651. .8654
                     /gene="cap3C"
                     /citation=[2]
     CDS             8662. .9582
                     /gene="cap3C"
                     /standard_name="UDP-glucose pyrophosphorylase"
                     /function="synthesis of UDP-glucose"
```



```
                     /citation=[1]
                     /citation=[2]
                     /codon_start=1
                     /transl_table=11
...
Note: remainder of annotations omitted.

  Query Match           19.5%;  Score 17;  DB 15;  Length 14056;
  Best Local Similarity 100.0%;  Pred. No. 1.92e+01;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1617 CATTTTACTATATTTTT 1633
            |||||||||||||||||
Cp     50 CATTTTACTATATTTTT 34


RESULT   29
LOCUS       SPCPS14E    14943 bp    DNA             BCT       15-JAN-1998
DEFINITION  S.pneumoniae cps14 locus.
ACCESSION   X85787
NID         g2198539
KEYWORDS    cps14A gene; cps14B gene; cps14C gene; cps14D gene; cps14E gene;
            cps14F gene; cps14G gene; cps14H gene; cps14I gene; cps14J gene;
            cps14K gene; cps14L gene; dexB gene; glucosyl-1-phosphate
            transferase; glucosyltransferase;
            ss-1,3-N-acetylglucosaminyltransferase;
            ss-1,4-galactosyltransferase; tasA gene.
SOURCE      Streptococcus pneumoniae.
  ORGANISM  Streptococcus pneumoniae
            Eubacteria; Firmicutes; Low G+C gram-positive bacteria;
            Streptococcaceae; Streptococcus.
REFERENCE   1  (bases 1 to 14943)
  AUTHORS   Kolkman,M.A., Morrison,D.A., Van Der Zeijst,B.A. and Nuijten,P.J.
  TITLE     The capsule polysaccharide synthesis locus of streptococcus
            pneumoniae serotype 14: Identification of the glycosyl transferase
            gene cps14E
  JOURNAL   J. Bacteriol. 178 (13), 3736-3741 (1996)
  MEDLINE   96272244
REFERENCE   2  (bases 1 to 14943)
  AUTHORS   Kolkman,M.A., van der Zeijst,B.A. and Nuijten,P.J.
  TITLE     Functional analysis of glycosyltransferases encoded by the capsular
            polysaccharide biosynthesis locus of Streptococcus pneumoniae
            serotype 14
  JOURNAL   J. Biol. Chem. 272 (31), 19502-19508 (1997)
  MEDLINE   97382283
REFERENCE   3  (bases 1 to 14943)
  AUTHORS   Kolkman,M.A., Wakarchuk,W., Nuijten,P.J. and van der Zeijst,B.A.
  TITLE     Capsular polysaccharide synthesis in Streptococcus pneumoniae
            serotype 14: molecular analysis of the complete cps locus and
            identification of genes encoding glycosyltransferases required for
            the biosynthesis of the tetrasaccharide subunit
```

```
  JOURNAL   Mol. Microbiol. 26 (1), 197-208 (1997)
  MEDLINE   98043549
REFERENCE   4  (bases 1 to 14943)
  AUTHORS   Kolkman,M.A.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-MAR-1995) M.A.B. Kolkman, Institute of Infectious
            Diseases and, Immunology,Department of Bacteriology, School of
            Veterinary Medicine, University of Utrecht, P.O. Box 80.165, NL
            -3508 TD Utrecht, NETHERLANDS
  REMARK    revised by [3]   MAT
REFERENCE   5  (bases 1 to 1821)
  AUTHORS   Kolkman,M.A.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-SEP-1995) M.A.B. Kolkman, Institute of Infectious
            Diseases and, Immunology,Department of Bacteriology, School of
            Veterinary Medicine, University of Utrecht, P.O. Box 80.165, NL
            -3508 TD Utrecht, NETHERLANDS
  REMARK    revised by [4]
REFERENCE   6  (bases 1 to 14943)
  AUTHORS   Kolkman,M.A.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-APR-1996) M.A.B. Kolkman, Institute of Infectious
            Diseases and, Immunology,Department of Bacteriology, School of
            Veterinary Medicine, University of Utrecht, P.O. Box 80.165, NL
            -3508 TD Utrecht, NETHERLANDS
  REMARK    revised by [5]
REFERENCE   7  (bases 1 to 14943)
  AUTHORS   Kolkman,M.A.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUN-1997) M.A.B. Kolkman, Institute of Infectious
            Diseases and, Immunology,Department of Bacteriology, School of
            Veterinary Medicine, University of Utrecht, P.O. Box 80.165, NL
            -3508 TD Utrecht, NETHERLANDS
COMMENT     On Jun 17, 1997 this sequence version replaced gi:1279732.
FEATURES             Location/Qualifiers
     source          1..14943
                     /organism="Streptococcus pneumoniae"
                     /db_xref="taxon:1313"
     gene            1..353
                     /gene="dexB"
     CDS             1..353
                     /gene="dexB"
                     /function="degradation of glucans"
                     /codon_start=3
                     /transl_table=11
                     /product="alpha, 1-6-glucosidase"
                     /db_xref="PID:e1169882"
                     /db_xref="PID:g2198540"
                     /translation="PWLAVNPNYEMINVQEALANPDSIFTTYQKLVQIRKENSWLVRA
                     DPELLDTADKVFAYIRKDGDRRPLVVANLSNEEQDLTVEGKVKSVLIENTAAKEVLEK
                     QVLAPWDAFCVELL"
```



```
     source          1..5627
                     /note="serotype 14"
                     /organism="Streptococcus pneumoniae"
                     /isolate="NCTC 11902"
                     /db_xref="taxon:1313"
                     /sub_clone="cMK02"
                     /sub_clone="pMK100"
                     /sub_clone="pMK102"
                     /sub_clone="pMK118"
                     /sub_clone="pMK114"
                     /sub_clone="pGV100"
                     /sub_clone="pMK110"
     terminator      700..753
     gene            1054..11416
                     /gene="cps14K"
     RBS             1054..10552
                     /gene="cps14K"
     repeat_region   1196..1305
     -35_signal      1360..1365
                     /gene="cps14K"
     -10_signal      1383..1388
                     /gene="cps14K"
     RBS             1408..1410
                     /gene="cps14A"
     gene            1408..2864
                     /gene="cps14A"
     CDS             1419..2864
                     /gene="cps14A"
                     /function="putative regulatory"
                     /codon_start=1
                     /transl_table=11
                     /product="capsular polysaccharide synthesis protein"
                     /db_xref="PID:e1169883"
                     /db_xref="PID:g2198541"
                     /translation="MSRRFKKSRSQKVKRSVNIVLLTIYLLLVCFLLFLIPKYNILAF
                     RYLNLVVTALVLLVALVGLLLIYKKAEKFTIFLLVTSILVSSVSLFAVDQFVGLTNR
                     LNATSMYSEYSISVAVLADSDIENVTQLTSVTAPTGTDNENIQKLLADIKSSQNTDLT
                     VDQSSSYLAAYKSLIAGETKAIVLNSVFENIIESEYPDYASKIKKIYTEGPTKKVEAP
                     KTSKNQSFNIYVSGIDTYGPISSVSRSDVNILMTVNRDTKKILLTTTPRDAYVPIADG
                     GNNQKDKLTHAGIYGVDSSIHTLENLYGVDINYYVRLNFTSFLKMIDLLGGVDVHNDQ
                     EFSALHGKFHPPVGNVHLDSEQALGFVRERYSLADGDRDRGRNQQKVIVAILQKLTST
                     EALKNYSTIIDSLQDSIQTNMPLETMINLVNTQLESGGNYKVNSQDLKGTGRTDLPSY
                     AMPDSNLYVMEIDDSSLAVVKAAIQDVMEGR"
     RBS             2852..2857
                     /gene="cps14B"
     gene            2852..3597
                     /gene="cps14B"
     CDS             2866..3597
                     /gene="cps14B"
                     /codon_start=1
                     /transl_table=11
```



```
                     /product="capsular polysaccharide synthesis protein"
                     /db_xref="PID:e1169884"
                     /db_xref="PID:g1279733"
                     /translation="MIDIHSHIVFDVDDGPKSREESKALLIESYRQGVRTIVSTSHRR
                     KGMFETPEEKIAENFLQVREIAKEVADDLVIAYGAEIYYTLDALEKLEKKEIPTLNDS
                     RYALIEFSWHTSYFQIHTGLSWILMLYITPVIAHIERYDALENNEKHVRELIDMGCYT
                     QINSYHVSKPKFFGEKYKFMKKRARYFLERDLVHVVASDMHNLDSRPPYMQQAYDIIA
                     KKYWSQKSERTFVDNPRKIIMDQLI"
     RBS             3596..3600
                     /gene="cps14C"
     gene            3596..4298
                     /gene="cps14C"
     CDS             3606..4298
                     /gene="cps14C"
                     /function="putative chain length regulator"
                     /codon_start=1
                     /transl_table=11
                     /product="capsular polysaccharide synthesis protein"
                     /db_xref="PID:e1169885"
                     /db_xref="PID:g1279734"
                     /translation="MKEQNTLEIDVLQLSRALWKRKLVILLVAIITSSVAFAYSTFVI
                     KPEFTSTTRIYVVNRDQGEKSGLTNQDLQAGSYLVKDYREIILSQYVLEEVVSDLKLD
                     LTPKGLANKIKVTVPVDTRIVSVSVNDRVPEEASRIANSLREVAAQKIISITRVSDVT
                     TLEEARPAISPSSPNIKRNTLIGFLAGVSGTSVIVFLLEPLNTRVKRPEDIENTLQMT
                     LLGVVPNLSKLK"
     RBS             4297..4302
                     /gene="cps14D"
     gene            4297..4991
                     /gene="cps14D"
     CDS             4308..4991
                     /gene="cps14D"
                     /function="putative polysaccharide export protein"
                     /codon_start=1
                     /transl_table=11
                     /product="capsular polysaccharide synthesis protien"
                     /db_xref="PID:e1169886"
                     /db_xref="PID:g1279735"
                     /translation="MPTLEIAQKKLEFIKKAEETYNALCTNIQLSGDKLKVISVTSVN
                     PGEGKTTTSVNIARSFARAGYKTLLIDGDTRNSVISGVFKSREKITGLTEFLSGTADL
                     SYGLCDTNIENLFVVQSGPVSPNPTALLQSKNFNDMIETLRKYFDYIIVDTPPVGIVI
                     DAVIITQKCDASILVTATGEANKRDVQKAKQQLEQTGKLFLGVVLNKLDISVDKYGVY
                     ASYGNYGKK"
     RBS             4995..5000
                     /gene="cps14E"
     gene            4995..6374
                     /gene="cps14E"
     CDS             5007..6374
                     /gene="cps14E"
                     /function="capsular polysaccharide synthesis"
                     /note="homologous to Cps19fE ofS. pneumoniae serotype 19F
                     and other rfbP-like proteins"
```



```
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="glucosyl-1-phosphate transferase"
                     /db_xref="PID:e1169887"
                     /db_xref="PID:g758302"
                     /translation="MDKKGLEIPLAVLQSIIVILLVYFLSFVRETELERSSMVILYLL
                     HFFVFTFSSYGNKFFKRGYLVEFNSTIRYIFFFAIAISVLNFFIAERFSISRRGMVYF
                     LTLEGISLYLLNFLVKKYWKHVFFNPKNSKKILLLTVTENIEKVLDKLLESDELSWKL
                     VAVSVLDKSDFQHDKIPVIEKEKIIEFATHEVVDEVFVDLPGESYDIGEIISKFETMG
                     IDVTVNLNAFNKNLGRNKQIHEIVGLNVVTFSTNFYKTSHVISKRILDICGATIGLIL
                     PAIASLVLVPLIRKDGGPAIFAQTRIGKNGRHFTFYKFRSMRIDAEVIKEQLMDQNTM
                     RGGMFKMDNDPRVTKIGRFIRKTSLDELPQFWNVFIGDMSLVGTRPPTVDEYDQYTPE
...
Note: remainder of annotations omitted.

  Query Match           19.5%;  Score 17;  DB 15;  Length 14943;
  Best Local Similarity 100.0%;  Pred. No. 1.92e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1290 CATTTTACTATATTTTT 1306
            |||||||||||||||||
Cp     50 CATTTTACTATATTTTT 34


RESULT   30
LOCUS       HSPAXIII6   16621 bp    DNA             PRI       08-NOV-1995
DEFINITION  Human paired box homeotic protein (PAX3) gene, exon 8 and partial
            cds.
ACCESSION   U12259
NID         g555816
KEYWORDS    .
SEGMENT     6 of 6
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 16621)
  AUTHORS   Macina,R.A., Barr,F.G., Galili,N. and Riethman,H.C.
  TITLE     Genomic organization of the human PAX3 gene: DNA sequence analysis
            of the region disrupted in alveolar rhabdomyosarcoma
  JOURNAL   Genomics 26 (1), 1-8 (1995)
  MEDLINE   95301273
REFERENCE   2  (bases 1 to 16621)
  AUTHORS   Riethman,H.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUL-1994) Harold C. Riethman, The Wistar Institute,
            3601 Spruce St, Philadelphia, PA 19104, USA
FEATURES             Location/Qualifiers
     source          1..16621
                     /organism="Homo sapiens"
```

```
                     /db_xref="taxon:9606"
                     /clone="Hup2SE1.9"
                     /chromosome="2"
                     /map="2q35"
     gene            join(U12263:1238. .1640,U12261:1. .390,U12262:1. .338,
                     U12260:1. .405,U12258:1. .552,1. .16208)
                     /gene="PAX3"
     CDS             join(U12262:<62. .267,U12260:80. .245,U12258:77. .291,
                     15942. .16208)
                     /gene="PAX3"
                     /standard_name="paired box homeotic gene 3"
                     /codon_start=3
                     /product="paired box homeotic protein"
                     /db_xref="PID:g555819"
                     /translation="SAPQSDEGSDIDSEPDLPLKRKQRRSRTTFTAEQLEELERAFER
                     THPDIYTREELAQRAKLTEARVQVWFSNRRARWRKQAGANQLMAFNHLIPGGFPPTA
                     MPTLPTYQLSETSYQPTSIPQAVSDPSSTVHRPQPLPPSTVHQSTIPSNPDSSSAYCL
                     PSTRHGFSSSYTDSFVPPSGPSNPMNPTIGNGLSPQVMGLLTNHGGVPHQPQTDYALSP
                     LTGGLEPTTTVSASCSQRLDHMKSLDSLPTSQSYCPPTYSTTGYSMDPVTGYQYGQYG
                     QSKPWTF"
     satellite       1549. .1598
                     /gene="PAX3"
                     /rpt_family="microsatellite (TC)8(TG)10(TT(TG)2TA(TG)2TA"
     satellite       6241. .6272
                     /gene="PAX3"
                     /rpt_family="microsatellite (GT)16"
     repeat_unit     7732. .8016
                     /gene="PAX3"
                     /rpt_family="Alu J"
     repeat_unit     8998. .9284
                     /gene="PAX3"
                     /rpt_family="Alu J"
     satellite       10841. .10900
                     /gene="PAX3"
                     /rpt_family="microsatellite
                     (GT)3(AT)3GT(AT)2GT(AT)4AC(AT)3(GT)4(AT)8"
     exon            15942. .>16208
                     /gene="PAX3"
                     /number=8
BASE COUNT     5149 a   3103 c   3136 g   5233 t
ORIGIN

  Query Match           19.5%;  Score 17;  DB 25;  Length 16621;
  Best Local Similarity 100.0%;  Pred. No. 1.92e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1043 TAAAATGTGTAATTTGT 1059
            |||||||||||||||||
Qy     44 TAAAATGTGTAATTTGT 60
```



```
RESULT   31
LOCUS       SPAJ6986     17340 bp    DNA             BCT       04-SEP-1998
DEFINITION  Streptococcus pneumoniae type 33F DNA, capsular gene cluster.
ACCESSION   AJ006986
NID         g3550625
KEYWORDS    aliA gene; alpha,1-6-glucosidase; cap33fA gene; cap33fB gene;
            cap33fC gene; cap33fD gene; cap33fE gene; cap33fF gene; cap33fG
            gene; cap33fH gene; cap33fJ gene; cap33fK gene; cap33fL gene;
            cap33fM gene; cap33fN gene; cap33fO gene; capp33fI gene; dexB gene;
            glycosyl transferase; oligopeptide permease; repeating unit
            transporter; transmembrane protein; UDP-galactopyranose mutase.
SOURCE      Streptococcus pneumoniae.
  ORGANISM  Streptococcus pneumoniae
            Eubacteria; Firmicutes; Clostridium group firmicutes;
            Streptococcaceae; Streptococcus.
REFERENCE   1  (bases 1 to 17340)
  AUTHORS   Llull,D., Lopez,R., Garcia,E. and Munoz,R.
  TITLE     Nucleotide sequence of the capsular gene cluster encoding the
            polysaccharide capsule of Streptococcus pneumoniae type 33F
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 17340)
  AUTHORS   Garcia,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUN-1998) Garcia E., Centro de Investigaciones
            Biologicas, CSIC, Microbiologia Molecular, Velazquez 144, Madrid,
            Madrid, Spain, 28006
FEATURES             Location/Qualifiers
     source          1. .17340
                     /organism="Streptococcus pneumoniae"
                     /strain="SSISP 33F Type 33F"
                     /db_xref="taxon:1313"
     gene            1. .864
                     /gene="dexB"
     CDS             <1. .462
                     /gene="dexB"
                     /function="degradation of glucans"
                     /codon_start=1
                     /transl_table=11
                     /product="alpha 1-6-glucosidase"
                     /db_xref="PID:e1318802"
                     /db_xref="PID:g3550626"
                     /translation="GVPIEEIMDSIRVIGRDNARTPMQWDESKNAGFSTGQPWLAVNP
                     NYEMINVQEALANPDSIFYTYQKLVQIRKENSWLIRADFELLDTADKVFAYIRKDGDR
                     RFLVVANLSNEEQDLTVEGKVKSVLIENTLAQEVPEKQILVPWDAFCVELL"
     terminator      810. .864
                     /gene="dexB"
     gene            1474. .2978
                     /gene="cap33fA"
     -35_signal      1474. .1479
                     /gene="cap33fA"
     -10_signal      1497. .1502
```



```
                     /gene="cap33fA"
     CDS             1533. .2978
                     /gene="cap33fA"
                     /codon_start=1
                     /transl_table=11
                     /db_xref="PID:e1318803"
                     /db_xref="PID:g3550627"
                     /translation="MSRRFKKSRSQKVKRSVNIVLLTIYLLLVCFLLFLIFKYNILAF
                     RYLNLVVTALVLLVALVGLLLIIYKKAEKFTIFLLLFSILVSSVSLFAVQQFVGLTNR
                     LNATSNYSEYSISVAVLADSEIENVTQLTSVTAPTGTDNENIQKLLADIKSSQNTDLT
                     VNQSSSYLAAAKSLIAGETKAIVLNSVFENIISEYPDYASKIKKIYTKGFTKKVEAP
                     KTSKNQSFNIYVSGIDTYGPISSVSRSDVNILMTVNRDTKKILLTTPRDAIVPIADG
                     GNNQKDKLTHAGIYGVDSSIHTLENLYGVDINYYVRLNFTSPLKLIDLLGGIDVYNDQ
                     EFTAHTNGKYYPAGNVHLDSEQALGFVRERYSLADGDRDRGRNQQKVIVAILQKLTST
                     EALKNYSTIINSLQDSIQTNMPLETMINLVNAQLESGGNYKVNSQDLKGTGRTDLPSY
                     AMPDSNLYVMEIDDSSLAVVKAAIQDVMEGR"
     gene            2980. .3711
                     /gene="cap33fB"
     CDS             2980. .3711
                     /gene="cap33fB"
                     /codon_start=1
                     /transl_table=11
                     /db_xref="PID:e1318804"
                     /db_xref="PID:g3550628"
                     /translation="MIDIHSHIVFDVDDGPKSREESKALLTESYRQGVRTIVSTSHRR
                     KGMFETPEEKIAENFLQVREIAKEVASDLVIAYGAEIYYTPDVLDKLEKKRIPTLNDS
                     RYALIEPSMNTPYRDIHSALSKILMLGITPVIAHIERYDALENNEKRVRELIDRGCYT
                     QVHSSHVLKPKLFGERYKFMKKRAQYFLEQDLVHIIASDMHNLDGRPPHMAEAYDLVT
                     QKVGEAKAQELFIDNPRKIVMDQLI"
     gene            3717. .4412
                     /gene="cap33fC"
     CDS             3717. .4412
                     /gene="cap33fC"
                     /codon_start=1
                     /transl_table=11
                     /db_xref="PID:e1318805"
                     /db_xref="PID:g3550629"
                     /translation="MMKEQNTIEIDVFQLFKTLWKRKLMILIVALVTGAGAFAYSTFI
                     VKPEYTSTTRIYVVNRNQGDDKPGLTNQDLQAGSTLVKDYREIILSQDALEKVATNLKL
                     DMPAKTLASKVQVTVPADTRIVSISVKDKQPEEASRIANSLREVAAEKIVAVTRVSDV
                     TTLEEARPATTPSSPNVRRNSLFGPLGGAVVTVIAVLLIELLDTRVKRPEDVEDVLKI
                     PLLGLVPDFDKIK"
     gene            4422. .5111
                     /gene="cap33fD"
     CDS             4422. .5111
                     /gene="cap33fD"
                     /codon_start=1
                     /transl_table=11
                     /db_xref="PID:e1318806"
                     /db_xref="PID:g3550630"
                     /translation="MPTLEISQAKLDSVKKAEEYYNALCTNLQLSGDGLKVFSITSVK
```



```
                     IGEGKSTTSANIAWAFARAGYKTLLIDGDIRNSVMLGVFKARNKITGLTEFLSGTTDL
                     SQGLCDTNIENLFVIQAGSVSPNPTALLQSKNFTTMLETLRKYFDYIIVDTAPVGVVI
                     DAAIITRNCDASILVTEAGEINRRDIQKAKEQLEHTGKPFLGIVLNKFDTSVDKYGSY
                     GNYRDYGKNKK"
     gene            5126. .6493
                     /gene="cap33fE"
     CDS             5126. .6493
                     /gene="cap33fE"
                     /function="synthesis of capsular polysaccharide"
                     /codon_start=1
                     /transl_table=11
                     /product="glycosyl transferase"
                     /db_xref="PID:e1318807"
                     /db_xref="PID:g3550631"
                     /translation="MNGKIVKPSLAIIQSFLVILLTYLLSAVREAEIVSTTAIALYIL
                     HYFVPYISVYGQDFFRRGYLIELVQTLKYILFFALAISISNFFLEDRFSISRRGMIYF
                     LTLHALLVYVLNLFIKWYWKRTYPNFKGSKKILLLTATFRVEKVLDRLIESNEVVGEL
                     VAVSVLDKPDFQHDCLKVVAEGEIVNFATHEVVDEVFINLPSEKYNIGELVSQFETMG
                     IDVTVNLNAFDRSLARNKQIREMAGLNVVTFSTTFYKTSHVIAKRIIDIMGALVGLIL
                     CGLVSIVLVPLIRKDGGSAIFAQTRIGKNGRQFTFYKFRSMCVDAEAKKRELMEQNTM
                     QGGMFKVDDDPRITKIGRFIRKTSLDELPQFYNVLKGDMSLVGTRPPTVDEYEHYTPE
                     QKRRLSFKPGVTGLWQVSGRSEIKNFDEVVKLDVAYIDDWTIWKDIEILLKTVKVVLM
                     KDGAK"
     gene            6480. .7286
                     /gene="cap33fF"
     CDS             6480. .7286
                     /gene="cap33fF"
                     /note="an alternative ATG initiation codon is located at
                     position 6540-6542"
                     /codon_start=1
                     /transl_table=11
                     /db_xref="PID:e1318808"
                     /db_xref="PID:g3550632"
                     /translation="MERSRLIDVKIIVATHKEVKMPQDNSLYLPIHVGRDGKSDIGFI
                     GDNTGDNISSLNPYYCELTGLYWAWKNLDYNYLGLVHYRRYFTNKSQGYNENVNNDDL
                     ILSRANVEILLEKSDIIVPKKRKYYIETLYSHYAHTLNGEHLDLARKIIEQNSSEYLS
                     SFDKVMKQRSGYMFNMFIMKKELLDDYLPWLFSILDTMYEQMDLTDYTLFESRLPGRV
                     SELLFNVWLCQKGITPKEVPFMYMERVDLPEKGKSPLMAKFFGKKYGQSF"
     gene            7290. .8330
                     /gene="cap33fG"
     CDS             7290. .8330
                     /gene="cap33fG"
                     /note="transmembrane protein"
                     /codon_start=1
                     /transl_table=11
                     /db_xref="PID:e1318809"
                     /db_xref="PID:g3550633"
                     /translation="MKVTIIGQIKNKRTGLGKAINDFRDYCCNRATRVTEIDITNNFN
                     FLSSLFQILISDTDVYYFTPAGSVAGNIRDSLFLFFMIMKRKKIVTHFHNSAFGNVMR
                     QHPTLMIINRILYSKVDLIILLGEKSKIMFQQLRILDEKFKIIRNGVDGYLFIEKNEL
                     NKKMSDLPINIIFFSNMIREKGYEILLEVAKKMVGDEKYHFYFSGKFQDNNLKTRFIN
```

```
                     EIYSMNNVTYLDGVYGSDKKKLLQKMHYFVLPSYYKDETLPISMLEAMANGLYIIVSD
                     VGVVSEVINKETASLIEMINEETADSIIEIINQTSNKLNELDPNVSKYKQELLNENIQ
                     ASIYQQLERIAN"
     gene            8357. .9325
                     /gene="cap33fH"
     promoter        8357. .8383
                     /gene="cap33fH"
     CDS             8408. .9325
                     /gene="cap33fH"
                     /function="synthesis of capsular polysaccharide"
                     /codon_start=1
                     /transl_table=11
                     /product="glycosyl transferase"
                     /db_xref="PID:e1318810"
                     /db_xref="PID:g3550634"
                     /translation="MPTILSAKNIERVELLIVNDGSTDRTEEIARQFEREYEGIVRVI
                     EKENQGHGSAVNAGIENAVGNYFKVVDADDWVNTNNLEDLIVFLSEVDVDQVLSPYDK
                     IFVNYRGDIERKEEECNEFSQVFNEVITSAEEFTTRIKQTVGMGHSITVTETSLLQENNIR
                     LSEKMFYVDMEYIVYMLPYVKKVVLFDKSIYRYRLGTETQSISMASYIKNRDMKKQVI
                     YHLVDFYNQMRSSAVLRRITWKLILNLIRQQWIITFNLSKKEGKNSECFEFDNWLIKE
                     GRIKKIPLYFFKAVKYIRFKVKYFLGIRK"
     gene            9327. .10295
                     /gene="capp33fI"
     CDS             9327. .10295
                     /gene="capp33fI"
                     /codon_start=1
                     /transl_table=11
                     /db_xref="PID:e1318811"
                     /db_xref="PID:g3550635"
                     /translation="MRKIGKVINEYFVLRKSFTPAIARNKLFEKFWGRIGNYKIFNNI
                     ASNFYQYKHETIINFLEKDFSQFLKSYNFKEVSHKEIEQRKIFSMWIQGYESAPKLVQ
                     KTIDSQRKYAEKYGYKFVFLDENNIRETVTLPSEIVEKYENGTIDFIKYSDVVRGTLL
                     SKYGGVWLDSTIYVDSSRELNYLKKDFYTIRAKTHERVPKYIANGRWSAFCLSGEKQN
                     IVFDPLEKFHVAYFMKYDIVLDYFLIDYIIELGYRTNDLIRNYIDKVEENNQELFFLA
                     DNFSNQYDEKEWAGVLSTTALFKCSYKCPINEATGTYFDRLMKGEL"
     gene            10315. .11286
                     /gene="cap33fJ"
     CDS             10315. .11286
                     /gene="cap33fJ"
                     /function="synthesis of capsular polysaccharide"
                     /codon_start=1
                     /transl_table=11
                     /product="glycosyl transferase"
                     /db_xref="PID:e1318812"
                     /db_xref="PID:g3550636"
                     /translation="MISVIVPVYNVADYLRFALDSLLEQTYKDFEVILVNDGSTDNSG
                     EICDEYGKLYDNIHVFHKKNGGLSDARNPGLEKSRGEFITFLDSDDIPEPYALELLIT
                     IQKKYDVDIVSTKGGITYSHDIFSKKLMAEDYLTVKILTKKEPLAAVTYNDEMTVSAW
...
Note: remainder of annotations omitted.
```



```
  Query Match           19.5%;  Score 17;  DB 16;  Length 17340;
  Best Local Similarity 100.0%;  Pred. No. 1.92e+01;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1405 CATTTTACTATATTTTT 1421
          |||||||||||||||||
Cp     50 CATTTTACTATATTTTT 34


RESULT   32
LOCUS       SPZ83335    23770 bp    DNA             BCT     07-OCT-1997
DEFINITION  S.pneumoniae dexB, capl[A,B,C,D,E,F,G,H,I,J,K] genes, dTDP-rhamnose
            biosynthesis genes and aliA gene.
ACCESSION   Z83335 Z84382
NID         g1944619
KEYWORDS    aliA gene; alpha,1-6-glucosidase; caplA gene; caplB gene; caplC
            gene; caplD gene; caplE gene; caplF gene; caplG gene; caplH gene;
            caplI gene; caplJ gene; caplK gene; dexB gene;
            dTDP-4-dehydrorhamnose 3,5-epimerase; dTDP-4-dehydrorhamnose
            reductase; dTDP-glucose 4,6-dehydratase; galacturonic acid
            acetylase; galacturonosyl transferase; glucose-1-phosphate
            thymidylyltransferase; oligopeptide permease; uridine diphosphate
            glucose dehydrogenase.
SOURCE      Streptococcus pneumoniae.
  ORGANISM  Streptococcus pneumoniae
            Eubacteria; Firmicutes; Low G+C gram-positive bacteria;
            Streptococcaceae; Streptococcus.
REFERENCE   1  (bases 1 to 23770)
  AUTHORS   Munoz,R., Mollerach,M.E., Lopez,R. and Garcia,E.
  TITLE     Molecular organization of the genes required for the synthesis of
            type 1 capsular polysaccharide of Streptococcus pneumoniae:
            formation of binary encapsulated pneumococci and identification of
            cryptic dTDP-rhamnose biosynthesis genes
  JOURNAL   Mol. Microbiol. 25, 79-92 (1997)
REFERENCE   2  (bases 1 to 23770)
  AUTHORS   Garcia,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-DEC-1996) Garcia E., Centro de Investigaciones
            Biologicas. CSIC, Microbiologia Molecular, Velazquez, 144, Madrid,
            Madrid, Spain, 28006
  REMARK    Revised by [3]
REFERENCE   3  (bases 1 to 23770)
  AUTHORS   Garcia,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-APR-1997) Garcia E., Centro de Investigaciones
            Biologicas. CSIC, Microbiologia Molecular, Velazquez, 144, Madrid,
            Madrid, Spain, 28006
COMMENT     On Apr 18, 1997 this sequence version replaced gi:1870153.
FEATURES             Location/Qualifiers
     source          1. .23770
                     /organism="Streptococcus pneumoniae"
```



```
                     /strain="13868"
                     /db_xref="taxon:1313"
     gene            1. .697
                     /gene="dexB"
     CDS             <1. .697
                     /gene="dexB"
                     /standard_name="glucosidase"
                     /function="degradation of glucans"
                     /codon_start=2
                     /transl_table=11
                     /product="alpha, 1-6-glucosidase"
                     /db_xref="PID:e289125"
                     /db_xref="PID:g1870154"
                     /db_xref="SPTREMBL:P96472"
                     /translation="NSLFWNNHDLPRIVSIWGNDQEYREKSAKAFAILLHLMRGTPYI
                     YQGEEIGMTNYPFETLDQVEDIESLNYARERALEKGVPIEEIMDSIRVIGRDNARTPMQ
                     WDESKNAGFSTGQPWLAVNPNYEMINVQEALANPDSIFTTYQKLVQIRKENSSWLVRAD
                     FELLDTADAKVFAYIRKDGDRRFLVVANLSNEEQDLTVEGKVKSVLIENTLAQEVFEKQ
                     ILVPWDAFCVELL"
     terminator      1042. .1099
                     /standard_name="Terminator"
     source          complement(1166. .2629)
                     /note="Similar, although not identical, to the insertion
                     sequence element IS1167 described by Zhou et al. (Plasmid
                     33:127-138; 1995)."
                     /organism="Streptococcus pneumoniae"
                     /insertion_seq="IS1167"
                     /db_xref="taxon:1313"
     -35_signal      3536. .3541
                     /gene="caplA"
     gene            3536. .5040
                     /gene="caplA"
     -10_signal      3559. .3564
                     /gene="caplA"
     RBS             3582. .3586
                     /gene="caplA"
     CDS             3595. .5040
                     /gene="caplA"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /db_xref="PID:e289126"
                     /db_xref="PID:g1870155"
                     /db_xref="SPTREMBL:P96473"
                     /translation="MSRRFKKSRSQKVKRSVNIVLLTIYLLLICFLLFLIFKYNILAF
                     RYLNLVVTALVLLVALVGLLLIIYKKAEKFTIFLLLFSILVSSVSLFTVQQFVGLTNR
                     LNATSNYSEYSISVAVLADSKIENVTQLTSVTAPTGTDNENIQKLLADIKSSQNTDLT
                     VNQSSSYLAATKSLIAGETKAIVLNSVFENIIESEYPDYASKIKKIYTKGFTKKVEAP
                     KTSKNQSFNIYVSGIDTYGPISSVSRSDVNILMTVNRDTKKILLTTTPRDAYVPIADS
                     GNNQKDKLTHAGIYGVDSSIHTLENLYGVDINYTVRLNFTSFLKMIDLLGGVDVHNDQ
                     EFSALHGKFHFPVGNVHLDSEQALGFVRERYSLTDGDRDRGRNQQKVIVAIIQKLTST
```



```
                     EVLKNYSSILQGLQDSLQTNIPIETMIDLVNTQLESGGNYKVNSQDLKGTGRMDLPSY
                     AMPDSNLYVMEIDDSSLTVVKAAIQDVMEGR"
     RBS             5028. .5032
                     /gene="caplB"
     gene            5028. .5773
                     /gene="caplB"
     CDS             5042. .5773
                     /gene="caplB"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /db_xref="PID:e289127"
                     /db_xref="PID:g1870156"
                     /translation="MIDIHSHIVFDVDDGPKSREESKALLAESYRQGVRTIVSTSHRR
                     KGMFETPEEKIAENFLQVREIAKEVADDLVIAYGAEIYYTLDALEKLEKKEIPTLNDS
                     RYALIEPSMHTSYRQIHTGLSNILMLGITPVIAHIERYDALENNEKRVRELIDMGCYT
                     QINSYHVSKPKFPGEKYKFMEKHARYFLERDLVHVVASDMHNLDSRPPYMQQAYDIIA
                     KKYGAKKAKELFVDNPRKIIMDQLI"
     gene            5772. .6474
                     /gene="caplC"
     RBS             5772. .5776
                     /gene="caplC"
     CDS             5782. .6474
                     /gene="caplC"
                     /note="The deduced amino acid sequence contains a sugar
                     transport signature (LIGPLAGGIGTSVUVLLLELLDTHVK) typical
                     of glucose transporters"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /db_xref="PID:e289128"
                     /db_xref="PID:g1870157"
                     /db_xref="SPTREMBL:P96474"
                     /translation="MKEQNTLEIDVLQLFRALWKRKLVILLVAIITSSVAFAYSTFVI
                     KPEFTSTTRIYVVNRNQGEKPGLTNQDLQAGAYLVKDTREIVLSQDVLEEVISDLKLD
                     LMPKGLANKIKVTVPVDTRIVSVSVSDRVPEEASRIANSLREVAAQKISITRVSDVT
                     TLEEARPAISPSSPNIKRNTLIGPLAGVIGTSVIVLLLELLDTRVKRPEDIENALQMT
                     LLGVVPNLGKLK"
     gene            6473. .7149
                     /gene="caplD"
     RBS             6473. .6477
                     /gene="caplD"
     CDS             6484. .7149
                     /gene="caplD"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /db_xref="PID:e289129"
                     /db_xref="PID:g1870158"
                     /db_xref="SPTREMBL:P96475"
                     /translation="MPTLEIAQKKLEFIKKAEEYYNTLCTNIQLSGDKLKVISVTSVN
```

```
                PGEGKTTTSVNIAMSFARAGYETLLFDGDIRNSVMSGFPKSREKITGLTEPLSGTADL
                SNGLCDTNIENLFVVQSGGVSPNPTALLQSKNFNDMIETLRKYFDYIIVDTPPIGIVI
                DAAIITQKCDASILITATGEANKRDVQKAKQQLEQTGELFLGVVLNKLDISVDKYGVY
                GRK"
     gene       7275. .8426
                /gene="caplE"
     RBS        7275. .7278
                /gene="caplE"
     CDS        7290. .8426
                /gene="caplE"
                /function="synthesis of type 1 pneumococcal capsular
                polysaccharide"
                /codon_start=1
                /transl_table=11
                /product="galacturonosyl transferase"
                /db_xref="PID:e289130"
                /db_xref="PID:g1870159"
                /db_xref="SPTREMBL:P96476"
                /translation="MNLQKKKIMLITNHDDDIYCFRKEVIESLVEYGYEVLVSCPYGD
                KIELMKSIQFEYEDITIDRRGINPLKDIRLFLSYIRLLRKYRPDIDLTYTVKPNVYAS
                LAATILGIKYINNVTGLGSVLTMGSCVRSLVLLLFRISPRKSNCVFFQNEYNMKMLKT
                LGLILGDSKLIPGSGVNIEKYPVQQYPNGGNGIQGETIVFNFIGRILKEKGIDTYLAA
                AQIIKSRYPKTEFNIIGFIEPTESNYELKICDLEKKGIVYYLGQQKDAIPHITRSHAI
                IHPSVYGEGMSNVLLENASSGRVLITTDNPGCKEIVKDRETGYIFQGGNVEELVSILE
                VFLGLENEKRREMGLQGRMYIEQYFSRQIIIDAYLDKIKGALDE"
     RBS        8406. .8412
                /gene="caplF"
     gene       8406. .8970
                /gene="caplF"
     CDS        8419. .8970
                /gene="caplF"
                /function="acetylation of type 1 pneumococcal capsular
                polysaccharide"
                /codon_start=1
                /transl_table=11
                /product="galacturonic acid acetylase"
                /db_xref="PID:e289131"
                /db_xref="PID:g1870160"
                /db_xref="SPTREMBL:P96477"
                /translation="MNRIIKNIVGLLRANICLILLEVCRNEDLTFHYFLKFGKNVTVE
                QESNAKLLGKKIRVNAGGVLEVRKGAKLKISDDVFLSNNCMIACRKYIDIKSGVKCG
                PGVLIYDHDYDVSVPGGLAAKKFKTAPVNIGENVWIGANSIVLKGVSIGENSVVAAGS
                VVTKDIPADTIFIQKRLSREMKL"
     RBS        8953. .8958
                /gene="caplG"
     gene       8953. .10121
                /gene="caplG"
     CDS        8967. .10121
                /gene="caplG"
                /codon_start=1
                /transl_table=11
```



```
                /product="unknown"
                /db_xref="PID:e289132"
                /db_xref="PID:g1870161"
                /db_xref="SPTREMBL:P96478"
                /translation="MKVPIATEFRCTIYKNEIYLKPKAYFIYERYAKAFGEIVLCSRY
                EKVESYPEGYHKANFINKFIPIEGLHQVLFGQNKNKIIEGMIDSDLIVVRIPSIIGSK
...
Note: remainder of annotations omitted.

  Query Match           19.5%;  Score 17;  DB 15;  Length 23770;
  Best Local Similarity 100.0%;  Pred. No. 1.92e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   3463 CATTTTACTATATTTTT 3479
          |||||||||||||||||
Cp     50 CATTTTACTATATTTTT 34


RESULT   33
LOCUS       CE2K666     30589 bp    DNA              INV       22-SEP-1998
DEFINITION  Caenorhabditis elegans cosmid ZK666, complete sequence.
ACCESSION   Z49132
NID         g790416
KEYWORDS    HTG; Cytoskeleton assembly control protein; kinase; SLA1;
            Ubiquinol-cytochrome-C reductase.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 30589)
  AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
  TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
  JOURNAL   Nature 368 (6466), 32-38 (1994)
  MEDLINE   94150718
REFERENCE   2  (bases 1 to 30589)
  AUTHORS   Wilkinson,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-NOV-1996) Nematode Sequencing Project, Sanger Centre,
            Hinxton, Cambridge CB10 1RQ, England and Department of Genetics,
            Washington University, St. Louis, MO 63110, USA. E-mail:
```



```
            jea@sanger.ac.uk or rw@nematode.wustl.edu
  COMMENT   Coding sequences below are predicted from computer analysis, using
            predictions from Genefinder (P. Green, U. Washington), and other
            available information.
            For a graphical representation of this sequence and its analysis
            see:-
            http://weface.sanger.ac.uk/cgi-
            bin/display?db=wormace&class=Sequence &object=ZK666
            Current sequence finishing criteria for the C. elegans genome
            sequencing consortium are that all bases are either sequenced
            unambiguously on both strands, or on a single strand with both a
            dye primer and dye terminator reaction, from distinct subclones.
            Exceptions are indicated by an explicit note.
            IMPORTANT:  This sequence is NOT necessarily the entire insert of
            the specified clone.  It may be shorter because we only sequence
            overlapping sections once, or longer because we arrange for a small
            overlap between neighbouring submissions.
            IMPORTANT: This sequence is not the entire insert of clone ZK666.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we arrange for a small overlap between
            neighbouring submissions.
            The true left end of clone ZK666 is at 1 in this sequence. The true
            right end of clone ZK666 is at 7572 in
            sequence Z48716.
            The true left end of clone F59B10 is at 30489 in this sequence. The
            true right end of clone ZK673 is at 9193 in this sequence. The
            start of this sequence (1. .101) overlaps with the end of sequence
            Z48585.
            The end of this sequence (30489. .30589) overlaps with the start of
            sequence Z48716.
  FEATURES              Location/Qualifiers
     source             1. .30589
                        /organism="Caenorhabditis elegans"
                        /db_xref="taxon:6239"
                        /chromosome="II"
                        /clone="ZK666"
     gene               complement(1448. .1861)
                        /gene="ZK666.1"
     CDS                complement(1448. .1861)
                        /gene="ZK666.1"
                        /codon_start=1
                        /db_xref="PID:g790417"
                        /db_xref="SPTREMBL:Q23559"
                        /translation="MEYSYIPSFALYNNADMFQFFPRNKRLSTHQKSCLDEHARKMAV
                        FMMKRQEEYRYYENLANQEQFSDDAHFSDNGMDSGFCSGATSTGQSASTSPAPVQLIP
                        NPFFDPEAKTQSPVKLVDNPELFMESKKKSARISA"
     gene               complement(2210. .2404)
                        /gene="ZK666.2"
     CDS                complement(2210. .2404)
                        /gene="ZK666.2"
                        /codon_start=1
```



```
                        /db_xref="PID:g790418"
                        /db_xref="SPTREMBL:Q23560"
                        /translation="MYAPGYTPYVAHGSPYGAPMYGSPYGYGYASPAYVQPVGPPVYA
                        GHGGYYNDDCCGCCCGGYYY"
     gene               complement(3134. .4183)
                        /gene="ZK666.10"
     CDS                complement(join(3134. .3347,3498. .3619,3663. .3791,
                        3841. .4056,4103. .4183))
                        /gene="ZK666.10"
                        /codon_start=1
                        /db_xref="PID:e1187667"
                        /db_xref="PID:g2652945"
                        /db_xref="SPTREMBL:O18298"
                        /translation="MNLLAMYMIWFQSAGMYGYRHCLTYVQIVSFLVEIIMAWMIPLH
                        IFFPLVAAITLREGHRIIMSNQMALTIWIFLLCLVLPASITCFIYRHDAALQINENNS
                        SKYSHKKLAMILNHIPPFLTAIGTSNCQSTIEQKYEYVLQNFPQCISWVASENFEAYD
                        YHVNPWLVRTVSAGIGFMVVSTIHGACLGVHTMIVLQRLRSHMSVQIYQTHRTALISL
                        AMQMACPCVFIPPVYPYAVVSWLEAVELQGEFLLQ"
     gene               complement(4735. .6653)
                        /gene="ZK666.3"
     CDS                complement(join(4735. .4833,4892. .4984,5030. .5142,
                        5267. .5461,5509. .5708,5933. .6031,6134. .6348,6395. .6534,
                        6614. .6653))
                        /gene="ZK666.3"
                        /note="similar to Lectin C-type domain short and long
                        forms, von Willebrand factor type A domain"
                        /codon_start=1
                        /db_xref="PID:e1187666"
                        /db_xref="PID:g2652944"
                        /db_xref="SPTREMBL:Q23561"
                        /translation="MNTVFACLAIFVFGIYCSPVPQNAAAYTDPACPGNIKNLWLDVV
                        VVVDRSMQMTDAQLWQVRQTLTQVPGSDIRIGTGYSDKRSTCVGPVTYNSNATVTAEL
                        DIAKSFPDIYNIIQGSLVDVASTNASSLGEGLLAAQRVLNSGRQRTNRYNVKQVILAF
                        AADPQDEFSNLNAIELSRDLQSNGISIITVACTKDSYALPRINMCATPGYAFVDEMNT
                        SKLVKQLTGALININCFCPEDWVQYTGSDKKKLGVCIKGFDPFSGSAWGYDHAVQYCQN
                        QVSRGHLANEFSKEKHDFINLYHMNTFNRGMHLDYSIGLRYLNSTWVWEEPKGQPKLP
                        LNPDIYSTWAPGYPQRNSTGQVVGVKANGILTNWVGSATSEWLFVCQVQSSSTEHYTT
                        YVEQQ"
     gene               7773. .8207
                        /gene="ZK666.4"
     CDS                7773. .8207
                        /gene="ZK666.4"
                        /codon_start=1
                        /db_xref="PID:g790420"
                        /db_xref="SPTREMBL:Q23562"
                        /translation="MEYTQIPSFALYNNASIFQFFPRNKRLSTHQKSCLDEHARKMAV
                        FMMERQEEYRYYENLANQEQFSDDACFSDTGMDSGFCSGDTSTGQSASTSPAPAHLIP
                        NPFFDPEYVASKTTMKLVANPFYNPKLAKEIKKKSTHTSKFQ"
     gene               complement(8536. .9962)
                        /gene="ZK666.5"
     CDS                complement(join(8536. .9056,9106. .9302,9350. .9549,
```

```
                     9595. .9708,9757. .9875,9923. .9962))
                     /gene="ZK666.5"
                     /note="weak similarity to cytoskeleton assembly control
                     protein SLA1 (Swiss Prot accession number P32790)"
                     /codon_start=1
                     /db_xref="PID:g790421"
                     /db_xref="SPTREMBL:Q23563"
                     /translation="MSKNICLLFFLFHAYLSSSYTDRICGQDLSNLWLDVVAVVDNSA
                     GMTKGGLTSVAANIASIFSKNTQIGTNPTSPKTTRLALVTTNVDATTAADLNKFQSID
                     DIYSGINSALATISSSEESYLARGLSQAEKVFQAGKHGFNRAHYQKVVIVYASTYKGS
                     GDLNPVPVAQRLKTSGVTIITVAYDQNKDGDILVDLEKIATPYHNLSNENLNVIGEIQ
                     GFLLQANCFCPTSWIQYRALHDDVDSYRYGICILPVGLTAVWRAAQLSCRNRWRNSYL
                     LNEYDNRKHNFVISSVKNPSGFHFPLSYHIGLNKVNGMWQWDQPQDWTQPQLQNYEIW
                     NSGYPISDSSLTTVLNRQTTVDLEIGWQNVHQFAAASNTVCETAACDTNNYCSSGGDS
                     ISNH"
          gene       complement(10666. .12331)
                     /gene="ZK666.6"
          CDS        complement(join(10666. .11186,11234. .11436,11484. .11683,
                     11845. .11958,12005. .12147,12292. .12331))
                     /gene="ZK666.6"
                     /note="similar to von Willebrand factor type A domain"
                     /codon_start=1
                     /db_xref="PID:g790422"
                     /db_xref="SPTREMBL:Q23564"
                     /translation="MKLTFLIVLIGVYACSAQDSSTTPSPSYTDRRCGEDLGNLWLDV
                     VAVVDNSIGMTNGGLTSIAANIASVVSSGTRIGTNPSEPRTTRLGLVTYNKAAAIQAD
                     LNQYQSLDDVYDNVFRALSSVSTSEESYLANGLARAEDVLEAGKQGYNRTHYQRVVIV
                     YASAYKGSGALDPVPVAERLKTSGVTVITVAYDQDGDALLADLAKIASPPYFNTNTE
                     DNGQVIGEIQDALLQVNCFCRNSWIQYRVSYTDVHSYRYGVCILPVGFNAVWRAAKSS
                     CMNRNAYLVTEFDSNKHDYVLTAVSNYTGFSVPLTYHNGLSMVKGVWVWDQPQGMPQP
                     QLQSYAGWNPGYPISDSTLTGVLNQQSSVQLQSGWQNIKPLTTGANYVCETASCDTDN
                     YCTEAGLAEISVKH"
          gene       complement(19273. .21005)
                     /gene="ZK666.7"
          CDS        complement(join(19273. .19467,19513. .19625,19672. .20066,
                     20344. .20543,20602. .20715,20767. .20915,20960. .21005))
                     /gene="ZK666.7"
                     /note="similar to Lectin C-type domain short and long
                     forms, von Willebrand factor type A domain; cDNA EST
                     EMBL:C13164 comes from this gene; cDNA EST EMBL:C11989
                     comes from this gene; cDNA EST yk296b6.5 comes from this
                     gene; cDNA EST yk378d5.5 comes from this gene; cDNA EST
                     yk363g11.5 comes from this gene; cDNA EST yk445d8.5 comes
                     from this gene; cDNA EST EMBL:C13343 comes from this gene;
                     cDNA EST EMBL:C12024 comes from this gene; cDNA EST
                     yk178d5.3 comes from this gene; cDNA EST yk363g11.3 comes
                     from this gene; cDNA EST EMBL:C10360 comes from this gene;
                     cDNA EST yk296b6.3 comes from this gene; cDNA EST
                     yk445d8.3 comes from this gene"
                     /codon_start=1
                     /db_xref="PID:g790423"
```



```
                     /db_xref="SPTREMBL:Q23565"
                     /translation="MAMKFLLTCFLLGITFCLGSPVSQSSNCQDGYMDRVCGEDETHL
                     WLDIVAVVDNSKGMTDKGVVTVAGQIVSLFVDGQQLGIDPNQPRTTRIGIVTYNRDAT
                     VVADLNKITSIDQLADIVFGALHKASSIADSYLHAGLEAADDLLQRQSFATSRGHYKK
                     LVIVYASESYSGDTQDPLPLATRMKVDVAIATVAYGQDNVNGPLRQLSQIATPGYNFT
                     NQNGIQLVPELRATMVQVNCYCPNQWFQMRQSYSDESSLKYGVCIHPVTLGATRTAAK
                     LSCRNQWNNAYMINEYTQTKHDFTLQVIENITAFTQPYTYYIGLSYSNGNWQNEQPNG
                     QPLAQLQTWSDWCPGYPTASSTDTAVVNVQSSSSSAKTCWQNTKSSLTNMYVCEVATC
                     DSTNYCTAEKI"
          gene       complement(22233. .23476)
                     /gene="ZK666.8"
          CDS        complement(join(22233. .22457,22505. .22768,22818. .23002,
                     23050. .23476))
                     /gene="ZK666.8"
                     /note="similar to serine/threonine kinase"
...
Note: remainder of annotations omitted.

  Query Match           19.5%;  Score 17;  DB 18;  Length 30589;
  Best Local Similarity 100.0%;  Pred. No. 1.92e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  13235 AAAAATATAGTAAAATG 13251
          |||||||||||||||||
Qy     34 AAAAATATAGTAAAATG 50


RESULT   34
LOCUS       CEK09G1     34657 bp    DNA             INV       21-SEP-1998
DEFINITION  Caenorhabditis elegans cosmid K09G1, complete sequence.
ACCESSION   Z81101
NID         g1628038
KEYWORDS    HTG; Dopamine receptor.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 34657)
  AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
  TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
```



```
  JOURNAL   Nature 368 (6466), 32-38 (1994)
  MEDLINE   94150718
REFERENCE   2  (bases 1 to 34657)
  AUTHORS   Lloyd,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-NOV-1996) Nematode Sequencing Project, Sanger Centre,
            Hinxton, Cambridge CB10 1RQ, England and Department of Genetics,
            Washington University, St. Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     Coding sequences below are predicted from computer analysis, using
            predictions from Genefinder (P. Green, U. Washington), and other
            available information.
            For a graphical representation of this sequence and its analysis
            see:-
            http://webace.sanger.ac.uk/cgi-
            bin/display?db=wormace&class=Sequence &object=K09G1
            Current sequence finishing criteria for the C. elegans genome
            sequencing consortium are that all bases are either sequenced
            unambiguously on both strands, or on a single strand with both a
            dye primer and dye terminator reaction, from distinct subclones.
            Exceptions are indicated by an explicit note.
            IMPORTANT:  This sequence is NOT necessarily the entire insert of
            the specified clone.  It may be shorter because we only sequence
            overlapping sections once, or longer because we arrange for a small
            overlap between neighbouring submissions.
            IMPORTANT: This sequence is not the entire insert of clone K09G1.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we arrange for a small overlap between
            neighbouring submissions.
            The true left end of clone K09G1 is at 1 in this sequence. The true
            right end of clone K09G1 is at 4549 in
            sequence Z70756.
            The true left end of clone T06E4 is at 34554 in this sequence. The
            true right end of clone F25B3 is at 10833 in this sequence. The
            start of this sequence (1. .104) overlaps with the end of sequence
            Z70752.
            The end of this sequence (34554. .34657) overlaps with the start of
            sequence Z70756.
FEATURES    Location/Qualifiers
     source 1. .34657
            /organism="Caenorhabditis elegans"
            /db_xref="taxon:6239"
            /chromosome="V"
            /clone="K09G1"
     gene   8204. .9268
            /gene="K09G1.1"
     CDS    join(8204. .8299,8448. .8570,8738. .8920,8970. .9083,
            9152. .9268)
            /gene="K09G1.1"
            /note="predicted using Genefinder; cDNA EST EMBL:T00664
            comes from this gene; cDNA EST EMBL:C11810 comes from this
```



```
            gene; cDNA EST EMBL:C10244 comes from this gene; cDNA EST
            yk256c5.3 comes from this gene; cDNA EST yk256c5.5 comes
            from this gene"
            /codon_start=1
            /db_xref="PID:e275821"
            /db_xref="PID:g1628041"
            /translation="MQTEEGTTVELGPAKDVPHSDDITRPDSSPKKEKKKLQFHMANK
            VKIIPTNRQGQLNTGYLATLAGVLKIAEIVLSFIAFILAICADRRTTTAAWTEHISPE
            TLLIVSALLLGYVVFPHLTIKDEATREGLIVVELIFYGVNTLLYFISIWLMVHLSASW
            GTDGRGAAIMTAVICVALTVLFAIETVVKLKAWRGENEPTTVIEAGNAQA"
     gene   9809. .10239
            /gene="K09G1.2"
     CDS    join(9809. .9883,9931. .10239)
            /gene="K09G1.2"
            /note="predicted using Genefinder"
            /codon_start=1
            /db_xref="PID:e275822"
            /db_xref="PID:g1628042"
            /translation="MSEKKKKPDSLRAETPQPKRSTGIKDVIDIIWAALIFVLMIVVR
            FLHSAFEILSDIFVKVYRILIFCYRKPNHAKELFKSTLNIIKVSYDAEIWTWTDMWEI
            FKSHLVTPMTKQPPARPAPQSAKQK"
     gene   complement(16385. .16600)
            /gene="K09G1.3"
     CDS    complement(join(16385. .16463,16518. .16600))
            /gene="K09G1.3"
            /note="predicted using Genefinder"
            /codon_start=1
            /db_xref="PID:e276188"
            /db_xref="PID:g1628040"
            /translation="MTKKKKRKKSEVGNYLKENVMDEEDDYNCKPSSKYSQKSKINRP
            LDYTARLLS"
     gene   27218. .33831
            /gene="K09G1.4"
     CDS    join(27218. .27784,28264. .28386,29529. .29641,29855. .29976,
            30031. .30138,30189. .30285,30331. .30398,31000. .31404,
            31766. .31880,32327. .32515,32562. .32830,32876. .33390,
            33467. .33610,33661. .33831)
            /gene="K09G1.4"
            /note="Similarity to Fugu dopamiune receptor (TR:E112838);
            cDNA EST yk421g1.3 comes from this gene; cDNA EST
            yk421g1.5 comes from this gene"
            /codon_start=1
            /db_xref="PID:e284810"
            /db_xref="PID:g1729637"
            /db_xref="SPTREMBL:P90927"
            /translation="MYQIHYVDSHLSIGKEPQTLSETLRTFQANYIFLTLQILSVIES
            KKLQNSIPFLKCSVPFLVLKHSEKLKSITFQKVGPFVELLTAKKHLKIFVLYIIALIL
            GTPLFLPSSNRLVGLKNVIKNCPKVIQLQKMEAGETWNVSLEWPPPSLDLSTITQTPS
            TIVGSGIPLNYAGLSLIVIPLITLLGNLLVIISVLRYRALQSAINFLILGLAVADLLV
            AIIVMPYAVTYVYTNGDWYLGNLMCDIYMASDVCCSTASILLLAVISFDRYRAVSLPI
            QYSRQSQNVKRVWTLIAVIWLVSLTLASPMVFGVNVRPPDANPYECRFYNAEFSILSS
```

```
                     MISFVIPCFLVLFVYIRIIIALKKREKAAKMRREKNTIAHGLTMRPDTGEEQVDEEAA
                     GRIVAGPVVNVMMAALPSMTRRMRQFERHRRAIELAGDEEWEEDELDVMDECCGGDDA
                     GDDDDDTHADNGQGVVEASAPRTTSMLRRIINAASVGTANSTAQSVASASGMPAFFAQ
                     NISTTSPSSSSCARTTTTTSAIPKASDEREFGNSSTPRSSLESLSENVNVITNDFVSE
                     NCTTFSRRSSYADDSQPTSSQTSSGDGRSYSIKGQRRFRNLSRNYSTKHHRKVVKVNR
                     GNSRNNSRTASITNQSDDALIPAAIRTISRKSPRLFRRDKTDIKKHSNILANPITEPP
                     KEYRRVSMPIHPTNSQTETETISASRDIENLPTTTISRSTTANSAELLGSPDDFEKFP
                     ALITETVLEDVLAETREGCFMQPTVSFALTVREMEGNALNNLKGCSVESSRRVSQVDP
                     PLAIQILTRPSLPRKFLKFFFTSKYVFLLHQKMFFETLPYFADLDLQRMDSIGTTCSS
                     KTRADSLRSVDSKGSKKSNRNGIAVKLVKRAIKHEHSLKRKVSKAQRKEKRATKTLGV
                     VVGVPLVCWVPFFVINILNAVCILLNKDSCQVGYDLFFYCTWIGYMNSFMNPIITTIF
                     NTEFRRAFKSIIFGRNSTRHHFSNKQAHV"
     gene            join(complement(1507. .1661),complement(1086. .1452),
                     complement(573. .1041),complement(105. .517),
                     complement(Z70752:30029. .30182),
                     complement(Z70752:29786. .29978),
                     complement(Z70752:29650. .29740))
                     /gene="F25B3.6"
     CDS             join(complement(1507. .1661),complement(1086. .1452),
                     complement(573. .1041),complement(105. .517),
                     complement(Z70752:30029. .30182),
                     complement(Z70752:29786. .29978),
                     complement(Z70752:29650. .29740))
                     /gene="F25B3.6"
                     /note="cDNA EST EMBL:T00216 comes from this gene; cDNA EST
                     EMBL:T02104 comes from this gene; cDNA EST EMBL:D65104
                     comes from this gene; cDNA EST EMBL:D68420 comes from this
                     gene; cDNA EST EMBL:C13751 comes from this gene; cDNA EST
                     EMBL:C11518 comes from this gene; cDNA EST yk343c1.3 comes
                     from this gene; cDNA EST yk484g10.5 comes from this gene"
                     /codon_start=1
                     /db_xref="PID:e1322808"
                     /db_xref="PID:g3643326"
                     /translation="MSSSESASSDEETKRRAPATSDSDSDSDAGPKPGKPLSTDSSAS
                     DSDAEKPQAKPAKKKTLTKRKRRATGSSDDDQVDDDLPADKEDKARWKKLTELEKEQE
                     IFERKEARENAIAREEIAQQLAKKAKSSEKGVKTEKRRKMNSGGSDAGSPKRKASSD
                     SDSEMDAAFHRPSDINRKHKEKNAMDALKNKRKEIEKKNAKNEALSIDAVPGANSGSS
                     SSSSSSESSRSSSSSRESSPERVSEKDKIVKKDVDGLSELRRARLSRHKLSLMIHAPF
                     FDSTVVGCIVRLGQGQMSGSGSKYRIWKIVGVEESNKVYELEGKKTNKIIKCQNGGSE
                     RPFRMQFVSNADFEQIEFDEWLLACKRHGNLPTVDIMDKKKQDIEKAINHKYSDKEVD
                     LMIKEKSKYQTVPRNFAMTKANWSKQKELAQQRGDIREAEQIQTKIDEIERQADELEK
                     ERSKSISAIAFINHRNRSKIKDQVLSGQLKIEENSQDDPPTRKKGGMRVVSGSKSRLD
                     GTLSASSSTTNLSDGGKDKSSSLAKPTQPPPSTQIKKKTDISSLHDFDLDIDLGKLKD
                     FSTPESSGNKRPSISSSKGVSLSDYRMRRSGGGDAGSSTSAAPSSAV"
BASE COUNT    10941 a   6059 c   5703 g   11954 t
ORIGIN

  Query Match          19.5%;  Score 17;  DB 18;  Length 34657;
  Best Local Similarity 100.0%;  Pred. No. 1.92e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```



```
Db  34348 GATAGATTTTATTAAGA 34364
          |||||||||||||||||
Cp     76 GATAGATTTTATTAAGA 60


RESULT   35
LOCUS       AC005524    35800 bp    DNA             PRI       24-AUG-1998
DEFINITION  Homo sapiens chromosome 19, cosmid R30149, complete sequence.
ACCESSION   AC005524
NID         g3451031
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 35800)
  AUTHORS   Lamerdin,J.E., McCready,P.M., Skowronski,E., Adamson,A.W.,
            Burkhart-Schultz,K., Gordon,L., Kyle,A., Ramirez,M., Stilwagen,S.,
            Phan,H., Velasco,N., Do,L., Regala,W., Terry,A., Garnes,J.,
            Danganan,L., Poundstone,P., Christensen,M., Georgescu,A., Avila,J.,
            Liu,S., Attix,C., Andreise,T., Trankheim,M., Amico-Keller,G.,
            Coefield,J., Duarte,S., Lucas,S., Bruce,R., Thomas,P., Quan,G.,
            Kronmiller,B., Arellano,A., Montgomery,M., Ow,D., Nolan,M.,
            Trong,S., Kobayashi,A., Olsen,A.S. and Carrano,A.V.
  TITLE     Sequence analysis of an ~600 kb contig from human 19q12 between
            D19S766 and D19S326
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 35800)
  AUTHORS   Lamerdin,J.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-AUG-1998) Joint Genome Institute, Lawrence Livermore
            National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
COMMENT     Map and sequence oriented from q centromere to telomere. Cosmid
            R30149 is separated from cosmid F21866 to the left by <1 kb. R30149
            is currently the telomeric end of this sequencing tiling path.
            Additional map and sequence information may be obtained at:
            http://www-bio.llnl.gov/bbrp/genome/genome.html.
FEATURES                Location/Qualifiers
     source             1. .35800
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone="R30149"
                        /chromosome="19"
                        /map="19q12 between D19S766 and D19S326"
                        /cell_line="5HL2-B"
                        /clone_lib="LL19NC03 R chromosome 19-specific cosmid
                        library"
                        /note="Cosmid library constructed at LLNL from flow-sorted
                        chromosomes from human-hamster hybrid 5HL2-B, which
                        carries chromosome 19 as its only human chromosome."
     repeat_region      complement(203. .767)
```



```
                        /rpt_family="L1PA6"
     repeat_region      764. .2271
                        /rpt_family="L1PA6"
     repeat_region      complement(2407. .2471)
                        /rpt_family="LINE2"
     repeat_region      2718. .2766
                        /rpt_family="MLT1D"
     repeat_region      2754. .2889
                        /rpt_family="MLT1E"
     repeat_region      complement(2895. .3196)
                        /rpt_family="AluSp"
     repeat_region      complement(3811. .3831)
                        /rpt_family="AT_rich"
     repeat_region      complement(5701. .5983)
                        /rpt_family="MER79"
     repeat_region      complement(5984. .6952)
                        /rpt_family="LTR5"
     repeat_region      complement(6954. .7099)
                        /rpt_family="MER79"
     repeat_region      complement(7780. .7808)
                        /rpt_family="AT_rich"
     repeat_region      7826. .8230
                        /rpt_family="MSTA"
     repeat_region      8231. .9844
                        /rpt_family="MSTA-internal"
     repeat_region      9845. .10276
                        /rpt_family="MSTA"
     repeat_region      complement(10277. .10635)
                        /rpt_family="L1PB3"
     repeat_region      10652. .11474
                        /rpt_family="L1M4"
     repeat_region      11587. .11978
                        /rpt_family="LTR3"
     repeat_region      12179. .14429
                        /rpt_family="L1M4"
     repeat_region      14388. .15544
                        /rpt_family="L1PA16"
     repeat_region      15572. .17322
                        /rpt_family="L1MD1"
     repeat_region      complement(18278. .18396)
                        /rpt_family="TIGGER1"
     repeat_region      18397. .18661
                        /rpt_family="AluJo"
     repeat_region      complement(18676. .20812)
                        /rpt_family="TIGGER1"
     misc_feature       complement(20918. .21057)
                        /note="predicted exon, program: grail2exons_human_1.3,
                        frame: 0, quality: excellent, score: 86.000"
     repeat_region      complement(22972. .22999)
                        /rpt_family="AT_rich"
     misc_feature       complement(23594. .23717)
```



```
                        /note="predicted exon, program: grail2exons_human_1.3,
                        frame: 2, quality: good, score: 59.000"
     repeat_region      23755. .24139
                        /rpt_family="MSTA"
     misc_feature       complement(24666. .24852)
                        /note="predicted exon, program: grail2exons_human_1.3,
                        frame: 0, quality: good, score: 60.000"
     repeat_region      25597. .25700
                        /rpt_family="(TAAAA)n"
     repeat_region      25777. .25939
                        /rpt_family="MER5B"
     misc_feature       26102. .26540
                        /note="BLASTN similarity to AQ057587 (5. .443); match:
                        0.99, score: 1.3e-173; database searched: month.na;
                        CIT-HSP-2345K18.TR CIT-HSP Homo sapiens genomic clone
                        2345K18"
     repeat_region      complement(26935. .26971)
                        /rpt_family="AT_rich"
     repeat_region      complement(27622. .27682)
                        /rpt_family="(CA)n"
     repeat_region      complement(28090. .28343)
                        /rpt_family="LINE2"
     repeat_region      28356. .28493
                        /rpt_family="MER5B"
     repeat_region      29139. .29974
                        /rpt_family="MER6"
     repeat_region      complement(30060. .30200)
                        /rpt_family="L1MC4"
     repeat_region      complement(30232. .30402)
                        /rpt_family="LINE2"
     repeat_region      30416. .32262
                        /rpt_family="L1M3/4"
     repeat_region      32254. .34105
                        /rpt_family="L1MA3"
     repeat_region      complement(34106. .34347)
                        /rpt_family="AluY"
     repeat_region      34348. .35147
                        /rpt_family="L1MA3"
     repeat_region      35150. .35451
                        /rpt_family="AluSp"
     repeat_region      35454. .35800
                        /rpt_family="L1MA3"
BASE COUNT    12504 a   7016 c   6568 g   9712 t
ORIGIN

  Query Match          19.5%;  Score 17;  DB 27;  Length 35800;
  Best Local Similarity 100.0%;  Pred. No. 1.92e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  26379 TTAGAAAGGAAAAATAT 26395
          |||||||||||||||||
```

```
Qy      25 TTAGAAAGGAAAAATAT 41


RESULT   36
LOCUS       CELF40A3    41155 bp    DNA             INV       02-AUG-1997
DEFINITION  Caenorhabditis elegans cosmid F40A3.
ACCESSION   AF016423
NID         g2291195
KEYWORDS    .
SOURCE      Caenorhabditis elegans strain=Bristol N2.
  ORGANISM  Caenorhabditis elegans
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Nematoda;
            Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea;
            Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 41155)
  AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
  TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
  JOURNAL   Nature 368 (6466), 32-38 (1994)
  MEDLINE   94150718
REFERENCE   2  (bases 1 to 41155)
  AUTHORS   Geisel,C., Bradshaw,H. and Keppler,D.
  TITLE     The sequence of C. elegans cosmid F40A3
  JOURNAL   Unpublished (1997)
REFERENCE   3  (bases 1 to 41155)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUL-1997)
COMMENT     Submitted by:
                  Genome Sequencing Center
                  Department of Genetics, Washington University,
                  St. Louis, MO  63110, USA, and
                  Sanger Centre, Hinxton Hall
                  Cambridge CB10 IRQ, England
                  e-mail: rw@nematode.wustl.edu and jes@sanger.ac.uk

            NOTICE:  This sequence may not be the entire insert of this clone.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we provide a small overlap between
            neighboring submissions.
```



```
            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded or sequenced with an alternate
            chemistry; an attempt was made to resolve all sequencing problems,
            such as compressions and repeats; all regions were covered by
            sequence from more than one subclone

                        NEIGHBORING COSMID INFORMATION:

            The 5' cosmid is C37C3, 200 bp overlap;3' lies in a gap.  Actual
            start of this cosmid is at base position 197 of CELF40A3; actual
            end is at 41155 of CELF40A3

            NOTES:

            Coding sequences below are predicted from computer analysis, using
            the program Genefinder(P. Green and L. Hillier, ms in preparation).
FEATURES                Location/Qualifiers
     source             1. .41155
                        /organism="Caenorhabditis elegans"
                        /strain="Bristol N2"
                        /db_xref="taxon:6239"
                        /chromosome="V"
                        /clone="F40A3"
     gene               complement(156. .1605)
                        /gene="F40A3.5"
     CDS                complement(join(156. .216,268. .662,714. .787,864. .981,
                        1094. .1281,1357. .1431,1476. .1605))
                        /gene="F40A3.5"
                        /note="contains similarity to a protein kinase domain"
                        /codon_start=1
                        /evidence=not_experimental
                        /db_xref="PID:g2291201"
                        /translation="MELDHTPPPSVLNDNCSASYMTPYATVIAMSGLYLLAIFYPCKK
                        SKKMCQPMSDSIYPYQKRLTQLERELFKDVRAHSVFFAELKNYLIDEESIEVDDFQIG
                        QTADGFIPRGGVFPKTRNRFNAKVTTAVKISPPIVSKSISLLEDALRLSKLDHPNLIR
                        LLAVSQLSFTVFRPMIALEWLPGGTLADYFQFKVREKDDSERSPIQLKDMLSILYQVS
                        QALKYIHSQLDEFGQELTHGRIFTRNVLVTEPDLRRCEVKLGDFGDAPMGLEYSTPII
                        AYMPPEILCCAERIPPHRPENDVWMPGVFIWECLTLGAQPHFRKSVEEIKKSPRLPDR
                        GLSCPPTCPLDV"
     gene               6740. .7594
                        /gene="F40A3.4"
     CDS                join(6740. .6827,7331. .7394,7450. .7594)
                        /gene="F40A3.4"
                        /codon_start=1
                        /evidence=not_experimental
                        /db_xref="PID:g2291200"
                        /translation="MSRMDSYSSSIREEQKAILNGLLGILFADGIIYDNYDTTRKIHQ
                        TVDGIYRMVVALLLIVIALIIIAVPLICFFHCRRANNEERSPLHSNRKSHPIEI"
```



```
     gene               8506. .9283
                        /gene="F40A3.3"
     CDS                join(8506. .8600,8667. .8781,8828. .9283)
                        /gene="F40A3.3"
                        /note="Similar to phosphatidylethanolamine binding
                        protein; coded for by C. elegans cDNA yk106a12.5; coded
                        for by C. elegans cDNA yk106a12.3"
                        /codon_start=1
                        /db_xref="PID:g2291199"
                        /translation="MVVLVTRSLLPALFFASRAPFAAATTSARFQRGLATMAAEAFTK
                        HEVIPDVLASNPPSKVVSVKFNSGVEANLGNVLTPTQVKDTPEVKWDAEPGALYTLIK
                        TDPDAPSRKEPTYREWHHWLVVNIPGNDIAKGDTLSEYIGAGPPPKTGLHRYVYLIYK
                        QSGRIEDAEHGRLTNTSGDKRGGWKAADFVAKHKLGAPVFGNLFQAETYDDYVPILNKQ
                        LGA"
     gene               9824. .12167
                        /gene="F40A3.2"
     CDS                join(9824. .9917,10735. .11058,11837. .12034,12091. .12167)
                        /gene="F40A3.2"
                        /note="contains similarity to EGF-like domains"
                        /codon_start=1
                        /evidence=not_experimental
                        /db_xref="PID:g2291197"
                        /translation="MLRRILLSAILIVLVACVLRKKKPCINGTPEGDRCYCIEGWTGT
                        LCHRKMTCSGPERLQNGSCIECELGWAGTDCDIIDCHGNGMPNYDLTECTCTVPYSGK
                        YCEIADTKDIYKVNNFTSSIGPIGILTIIPLALITMCCEHPAKKRQLRRVEQHLNGV
                        YITTGRKQVDRKMLVKGLLEASDDEKSVKSVKIAADSGKEDKDKEEEEVIKMRNSATTS
                        EKKALLADPADM"
     gene               complement(23146. .24752)
                        /gene="F40A3.6"
     CDS                complement(join(23146. .23311,23512. .23597,23644. .23841,
                        23885. .24052,24594. .24752))
                        /gene="F40A3.6"
                        /codon_start=1
                        /evidence=not_experimental
                        /db_xref="PID:g2291202"
                        /translation="MSEYLGWHNRKPAAGSRGNSSFDRDIVPGGANTKSYESYVKNIF
                        HPTNRLRAKDKDQLVRDRGLSATSMLRAEREPGRSALDDKYDREHFPEKMPLESLYAE
                        LGLKNPAASNYRSFQVYEPKSSSETYFAASYEAERERRKKDLERELASYDKEDYEDFTG
                        EPKKNLYSSISSYRPPLSSYKSREDEETEKNNRKTLIDYKDFDSSPLSPYLYQSHPYV
                        RSQDSRIIGSNFVQLTSRPKDKFISKIEETLAMVRDMPRY"
     gene               24968. .26143
                        /gene="F40A3.1"
     CDS                join(24968. .25107,25199. .25454,25498. .25634,26041. .26143)
                        /gene="F40A3.1"
                        /codon_start=1
                        /evidence=not_experimental
                        /db_xref="PID:g2291198"
                        /translation="MSEIVRLPEETQETEMPLHHANRAFGARKQDAGWMAKNEGKTFR
                        VVIKNIEIFQINDWELIGSRTYSWKGNEVLRCKHWDKVAIRWHLSTCTEHGVCDGGTN
                        VDYATRAQNWTEMPVWRNAFTYVQIPGQLQKTKFLRLGLTGYDSRKTQLEAFSPLKID
                        VDWRWQVTRDGCNNFVVRMSAMSMINEKMRPKLERFCRLFDKDHFVDCTNL"
```



```
     gene               complement(26520. .30751)
                        /gene="F40A3.7"
     CDS                complement(join(26520. .26710,27914. .28026,28197. .28358,
                        28415. .28609,29618. .29761,29814. .29935,30397. .30616,
                        30669. .30751))
                        /gene="F40A3.7"
                        /note="conatains similarity to seven-transmembrane
                        receptors"
                        /codon_start=1
                        /evidence=not_experimental
                        /db_xref="PID:g2291196"
                        /translation="MSANFSHLCLTDEQQRMSGTNLEYNLQRYFYPALAIFGILGNVL
                        NLTVLLNRSMRSRANSFLAVLAPADIIPLFLLFPNILANYSFFTFNWTFRWFYLHSKV
                        HLISLANWCSSVAHWCVIAVCGDRLLGIQNPLYARATWRWWKLPLVTTIIVPTCGLLT
                        CYGHFEYFCLVRSYCRETQLYSRCLPVNAEKWFGHRPNPPSQRYQSFIAICKLAHIFL
                        MIILPIILLLFVLNLTLLWALRKRQKHLSIGKDFNADRRQNDVHMQRTEHRVTLTVTPI
                        VTMPTLTNGPSALVHLVMYATHEELYDLTMISSTLVICGKASNPIFLCLGSKHFRLRL
                        LKLTRKKINRKIDSITGSLVHTTKLSTASSRRTSMPIQEKKRSSSCAPLVAKNRALST
                        VETIYQPLLNGNRKGDF"
BASE COUNT    13599 a    7323 c    6798 g   13435 t
ORIGIN

  Query Match           19.5%;  Score 17;  DB 18;  Length 41155;
  Best Local Similarity 100.0%;  Pred. No. 1.92e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  16892 AAAATGTGTAATTTGTC 16908
          |||||||||||||||||
Qy      45 AAAATGTGTAATTTGTC 61


RESULT   37
LOCUS       HSAC002114  45852 bp    DNA             PRI       13-MAY-1997
DEFINITION  Human Cosmid gl940a085 from 7q31.3, complete sequence.
ACCESSION   AC002114
NID         g2085788
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 45852)
  AUTHORS   Iadonato,S.P., Yu,J., Wong,G.K.-S., Magness,C.L., Green,E.D.,
            Green,P. and Olson,M.V.
  TITLE     Large-scale MCD Mapping and Sequencing of Human Chromosome 7
  JOURNAL   Unpublished (1996)
  REMARK    University of Washington Human Genome Center
              Box 352145 Seattle, WA 98195
              Contact: Shawn Iadonato (iadonato@u.washington.edu)
REFERENCE   2  (bases 1 to 45852)
```

```
  AUTHORS   Iadonato,S.P., Yu,J., Wong,G.K.-S., Magness,C.L., Green,E.D.,
            Green,P. and Olson,M.V.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-1997) Human Genome Center, University of
            Washington, Box 352145, Seattle, WA 98195, USA
  REMARK    University of Washington Human Genome Center
              Box 352145 Seattle, WA 98195
              Contact: Shawn Iadonato (iadonato@u.washington.edu)
COMMENT     Overlapping Sequences:
               5': UWGC:g1940a370 (Genbank:AC002084)
               3': UWGC:g1940a142
            ----------
            Sequence Quality Assessment:
            This entry has been annotated with sequence quality
            estimates computed by the Phrap assembly program.
            All manually edited bases have been reduced to quality zero.
            Quality levels above 40 are expected to have less than
            1 error in 10,000 bp.

            Pct of Consensus above quality 40: 97.1%


            Quality  Cumulative Percentage of Consensus
            -------  ----------------------------------
                90   xxx (13.5%)
                80+  xxxxxxxxxx (47.6%)
                70+  xxxxxxxxxxxxxxx (79.9%)
                60+  xxxxxxxxxxxxxxxxx (89.6%)
                50+  xxxxxxxxxxxxxxxxxx (93.9%)
                40+  xxxxxxxxxxxxxxxxxx (97.1%)
                30+  xxxxxxxxxxxxxxxxxxx (99.5%)
                20+  xxxxxxxxxxxxxxxxxxx (99.9%)
                10+  xxxxxxxxxxxxxxxxxxx (100.0%)
                00+  xxxxxxxxxxxxxxxxxxx (100.0%)

            Base-by-base quality values are not generally visible from the
            Genbank flat file format but are available as part
            of this entry's ASN.1 file.
            ----------
            Sequence Verification:
             This sequence has been verified by Multiple Complete Digest
             Mapping. Comparison of the experimentally derived map digest
             fragments with sequence-predicted fragments is given below.
             Small fragments below a variable cutoff (approximately 400-600bp)
             are not mapped and hence do not appear in the table. There are no
             significant remaining descrepancies between the experimental and
             predicted values. Uniquely ordered fragment groups are separated
             by dashed lines.
                    EcoRI                HindIII              NsiI
               Map      Seq       Map       Seq       Map       Seq
             -------- --------  --------  --------  --------  --------
```



```
             531.72   525.00   1686.38  1667.00   3044.76  2976.00
             -------- -------- -------- --------  -------- --------
             3699.80  3616.00   687.33   669.00   6052.19  6004.00
             -------- -------- -------- --------  -------- --------
            11242.50 11244.00   851.69   836.00    772.21   768.00
             -------- -------- -------- --------  -------- --------
             2121.39  2154.00  2481.36  2474.00  16310.35 16655.00
             -------- -------- -------- --------  -------- --------
             4660.00  4695.00  3836.96  3775.00   1876.76  1900.00
             -------- -------- -------- --------  -------- --------
             6292.09  6371.00  3324.07  3357.00   5094.92  5080.00
             -------- -------- -------- --------  -------- --------
             4661.65  4654.00   973.00   956.00    940.29   938.00
             -------- -------- -------- --------  -------- --------
             1373.13  1366.00  5891.05  5977.00   3031.63  3088.00
             -------- -------- -------- --------  -------- --------
             3641.86  3691.00  3316.61  3308.00   2049.94  2062.00
             -------- -------- -------- --------  -------- --------
             2805.90  2806.00  1108.19  1106.00   3179.68  3147.00
             -------- -------- -------- --------  -------- --------
                               3704.75  3738.00    789.05   786.00
                               -------- --------  -------- --------
                               1428.27  1415.00
                               -------- --------
                               2248.23  2235.00
                               -------- --------
                                773.41   762.00
                               -------- --------
                               1592.35  1579.00
                               -------- --------
                               2671.00  2727.00
                               -------- --------

FEATURES             Location/Qualifiers
     source          1. .45852
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7q31.3"
                     /clone="NHGRI:yWSS1940"
                     /sub_clone="UWGC:g1940a085"
                     /cell_line="GM10791"
                     /clone_lib="E. Green Chromosome 7 YAC Resource"
     repeat_region   1882. .2326
                     /rpt_family="MLT2B2"
     repeat_region   5116. .5567
                     /rpt_family="MLT2B2"
     repeat_region   5837. .6049
                     /rpt_family="Alu"
     repeat_region   8123. .8206
                     /rpt_family="THE1"
     repeat_region   9141. .9419
```



```
                     /rpt_family="Alu"
     repeat_region   15918. .18328
                     /rpt_family="L1"
     repeat_region   complement(19800. .19965)
                     /rpt_family="MIR"
     repeat_region   complement(21620. .21673)
                     /rpt_family="L1"
     repeat_region   21678. .22085
                     /rpt_family="L1"
     repeat_region   27915. .30262
                     /rpt_family="L1"
     repeat_region   32491. .32785
                     /rpt_family="Alu"
     repeat_region   complement(35081. .35198)
                     /rpt_family="MIR"
     repeat_region   35506. .35808
                     /rpt_family="Alu"
BASE COUNT    14709 a   9221 c   9049 g  12873 t
ORIGIN

  Query Match           19.5%;  Score 17;  DB 26;  Length 45852;
  Best Local Similarity 100.0%;  Pred. No. 1.92e+01;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   9918 TTTTAGAAAGGAAAAAT 9934
            |||||||||||||||||
Qy     23 TTTTAGAAAGGAAAAAT 39


RESULT   38
LOCUS       AC004264    47188 bp    DNA             PRI       06-MAR-1998
DEFINITION  Homo sapiens PAC clone DJ102K02 from 22q12.1-qter, complete
            sequence.
ACCESSION   AC004264
NID         g2935613
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 47188)
  AUTHORS   Bentley,D. and Scheet,P.
  TITLE     The sequence of Homo sapiens PAC clone DJ102K02
  JOURNAL   Unpublished (1998)
REFERENCE   2  (bases 1 to 47188)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAR-1998) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     SUBMITTED BY:  WUGSC
            Genome Sequencing Center
```



```
            Department of Genetics
            Washington University
            St. Louis MO 63108, USA
            http://genome.wustl.edu/gsc
            mailto:sapiens@watson.wustl.edu

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded or sequenced with an alternate
            chemistry; an attempt was made to resolve all sequencing problems,
            such as compressions and repeats; all regions were covered by
            sequence from more than one subclone; and the assembly was
            confirmed by restriction digest.

            MAPPING INFORMATION:
            This sequence was generated from part of bacterial clone contigs of
            human chromosome 22, constructed by the Sanger Centre chromosome 22
            mapping group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr22/

            SOURCE INFORMATION:
            This clone was derived from human PAC library RPCI-1, prepared by
            Pieter de Jong and coworkers at Roswell Park Cancer Institute,
            using the method described by Ioannou et al., Nature Genetics
            6:84-9 (1994).  The library is from one male donor.  For further
            details, see http://bacpac.med.buffalo.edu/
            The clone is available from Genome Systems, Inc.
            (http://www.genomesystems.com).
            VECTOR:  pCYPAC2
            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is DJ43804, 200 bp overlap;  The
            clone sequenced to the right is DJ130H16.  Actual start of this
            clone is at base position 197 of DJ102K02.
FEATURES             Location/Qualifiers
     source          1. .47188
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="22"
                     /clone="DJ102K02"
                     /clone_lib="RPCI-1"
                     /map="22q12.1-qter"
     repeat_region   20. .300
                     /rpt_family="MaLR"
     repeat_region   1255. .1618
                     /rpt_family="MaLR"
     repeat_region   2346. .2455
                     /rpt_family="MIR"
```

```
     repeat_region   3680. .3745
                     /rpt_family="MER1_type"
     repeat_region   4474. .4541
                     /rpt_family="MIR"
     repeat_region   4623. .4918
                     /rpt_family="Alu"
     gene            complement(8354. .11398)
                     /gene="LIF"
     CDS             complement(join(8354. .8764,9458. .9636,11380. .11398))
                     /gene="LIF"
                     /note="match to protein P15018 (PID:g126279);
                     H_DJ102K02.2"
                     /codon_start=1
                     /product="leukemia inhibitory factor"
                     /db_xref="PID:g2935615"
                     /translation="MKVLAAGVVPLLLVLHWKHGAGSPLPITPVNATCAIRHPCHNNL
                     MNQIRSQLAQLNGSANALFILYYTAQGEPFPNNLDKLCGPNVTDFPPFHANGTEKAKL
                     VELYRIVVYLGTSLGNITRDQKILNPSALSLHSKLNATADILRGLLSNVLCRLCSKYH
                     VGHVDVTYGPDTSGKDVFQKKKLGCQLLGKYKQIIAVLAQAF"
     repeat_region   14379. .14475
                     /rpt_family="MIR"
     repeat_region   14553. .14592
                     /rpt_family="L2"
     repeat_region   14832. .14995
                     /rpt_family="MIR"
     repeat_region   16003. .16135
                     /rpt_family="MIR"
     repeat_region   17150. .17205
                     /rpt_family="MIR"
     repeat_region   18949. .19062
                     /rpt_family="MIR"
     repeat_region   19594. .19729
                     /rpt_family="MIR"
     repeat_region   19800. .19883
                     /rpt_family="Alu"
     repeat_region   19984. .20031
                     /rpt_family="MIR"
     repeat_region   20209. .20483
                     /rpt_family="Alu"
     repeat_region   21095. .21187
                     /rpt_family="MIR"
     repeat_region   21377. .21502
                     /rpt_family="MIR"
     repeat_region   23111. .23146
                     /rpt_family="MIR"
     repeat_region   23254. .23318
                     /rpt_family="MIR"
     repeat_region   23925. .24224
                     /rpt_family="Alu"
     repeat_region   25333. .25705
                     /rpt_family="L2"
```



```
     repeat_region   25805. .25879
                     /rpt_family="MIR"
     repeat_region   26290. .26340
                     /rpt_family="MIR"
     repeat_region   26638. .26768
                     /rpt_family="MIR"
     repeat_region   26872. .26920
                     /rpt_family="L2"
     misc_feature    27535. .27912
                     /note="match to EST AA534303 (NID:g2278556) nf73e07.s1"
     repeat_region   27939. .28033
                     /rpt_family="MIR"
     gene            complement(28586. .31502)
                     /gene="OSM"
     CDS             complement(join(28586. .29167,29705. .29847,31469. .31502))
                     /gene="OSM"
                     /note="match to protein P13725 (PID:g129168) and EST
                     T29765 (NID:g611863); H_DJ102K02.1"
                     /codon_start=1
                     /product="oncostatin M precursor"
                     /db_xref="PID:g2935614"
                     /translation="MGVLLTQRTLLSLVLALLFPSMASMAAIGSCSKEYRVLLGQLQK
                     QTDLMQDTSRLLDPYIRIQGLDVPKLREHCRERPGAFPSEETLRGLGRRGFLQTLNAT
                     LGCVLHRLADLEQRLPKAQDLERSGLNIEDLEKLQMARPNILGLRNNIYCMAQLLDNS
                     DTAEPTKAGRGASQPPTPFPASDAFQRKLEGCRFLHGYHRFMHSVGRVFSKWGESPNR
                     SRRHSPHQALRKGVRRTRPSRKGKRLMTRGQLPR"
     misc_feature    complement(28633. .28926)
                     /gene="OSM"
                     /note="match to EST T29765 (NID:g611863)"
     repeat_region   29377. .29451
                     /rpt_family="L2"
     repeat_region   32272. .32385
                     /rpt_family="MIR"
     repeat_region   33001. .33081
                     /rpt_family="L2"
     repeat_region   33324. .33630
                     /rpt_family="Alu"
     repeat_region   33947. .34229
                     /rpt_family="Alu"
     repeat_region   34353. .34548
                     /rpt_family="MaLR"
     repeat_region   34552. .34848
                     /rpt_family="Alu"
     repeat_region   34853. .34913
                     /rpt_family="MaLR"
     repeat_region   34918. .35089
                     /rpt_family="Alu"
     repeat_region   35117. .35419
                     /rpt_family="Alu"
     repeat_region   35420. .35524
                     /rpt_family="MaLR"
```



```
     repeat_region   36078. .36123
                     /rpt_family="MIR"
     repeat_region   36140. .36464
                     /rpt_family="L1"
     repeat_region   36468. .36768
                     /rpt_family="Alu"
     repeat_region   36769. .36831
                     /rpt_family="L1"
     repeat_region   36832. .37133
                     /rpt_family="Alu"
     repeat_region   37134. .37429
                     /rpt_family="Alu"
     repeat_region   37462. .37956
                     /rpt_family="L1"
     repeat_region   38728. .38805
                     /rpt_family="MIR"
     repeat_region   39305. .39669
                     /rpt_family="L2"
     repeat_region   41402. .41447
                     /rpt_family="L1"
     repeat_region   41452. .41751
                     /rpt_family="Alu"
     repeat_region   41752. .42086
                     /rpt_family="L1"
     repeat_region   42121. .42402
                     /rpt_family="Alu"
...
Note: remainder of annotations omitted.

  Query Match           19.5%;  Score 17;  DB 26;  Length 47188;
  Best Local Similarity 100.0%;  Pred. No. 1.92e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  21403 CTATATTTTTCCTTTCT 21419
          |||||||||||||||||
Cp     43 CTATATTTTTCCTTTCT 27


RESULT   39
LOCUS       EPFCPCG     70028 bp    DNA   circular  PLN       05-MAY-1995
DEFINITION  Epifagus virginiana chloroplast complete genome.
ACCESSION   M81884
NID         g336917
KEYWORDS    complete genome; maturase; protease; pseudogene; ribosomal protein;
            transfer RNA-Arg; transfer RNA-Asn; transfer RNA-Asp; transfer
            RNA-Gln; transfer RNA-Glu; transfer RNA-His; transfer RNA-Ile;
            transfer RNA-Leu; transfer RNA-Met; transfer RNA-Phe; transfer
            RNA-Pro; transfer RNA-Ser; transfer RNA-Trp; transfer RNA-Tyr;
            transfer RNA-fMet; translation initiation factor 1.
SOURCE      Chloroplast Epifagus virginiana DNA.
  ORGANISM  Chloroplast Epifagus virginiana
```



```
            Eukaryotae; mitochondrial eukaryotes; Viridiplantae;
            Charophyta/Embryophyta group; Embryophyta; Magnoliophyta;
            Magnoliopsida; Gentiananae; Scrophulariales; Orobanchaceae;
            Epifagus.
REFERENCE   1  (bases 1 to 70028)
  AUTHORS   Wolfe,K.H., Morden,C.W., Ems,S.C. and Palmer,J.D.
  TITLE     Rapid evolution of the plastid translational apparatus in a
            nonphotosynthetic plant: loss or accelerated sequence evolution of
            tRNA and ribosomal protein genes
  JOURNAL   J. Mol. Evol. 35 (4), 304-317 (1992)
  MEDLINE   93021155
REFERENCE   2  (sites)
  AUTHORS   Wolfe,K.H., Morden,C.W. and Palmer,J.D.
  TITLE     Function and evolution of a minimal plastid genome from a
            nonphotosynthetic parasitic plant
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 89 (22), 10648-10652 (1992)
  MEDLINE   93066301
FEATURES              Location/Qualifiers
     source           1. .70028
                      /organism="Epifagus virginiana"
                      /chloroplast
                      /db_xref="taxon:4177"
     misc_feature     1. .19799
                      /note="large single copy region (LSC)"
     CDS              complement(234. .569)
                      /gene="pseudo-psbA"
                      /pseudo
     gene             complement(234. .569)
                      /gene="pseudo-psbA"
     gene             complement(588. .1907)
                      /gene="matK"
     CDS              complement(588. .1907)
                      /gene="matK"
                      /codon_start=1
                      /db_xref="PID:g336924"
                      /translation="MYKQNLFIISSNDSKYNFLLRNKKLYYEIILEGLTFILEFFFFI
                      RLISYLEGKKNKIEKLNNLRSILSIFPFLEDNFSNLKFILDILIPRKVHAEILVQTLR
                      YWIKDATYLHLLRFWLNNNWNSIINQNKDGYFYPQKNKILLLLYNSYVCKYESIFVFL
                      RNQSYHFRSTPPIMLLERIYFYFKIERLVNPPLKVKDFKANLWLIKEPCLHYFRYRRQ
                      YILASKGTSLFLNKWKCYVITFWQWYFSLWFFSRSIYIKKLLNNSFEILTYHSSLNIN
                      PSFVLSQILETPFISNNTIKKVNILVPIILSISDKEKFFNVLGHPTSNLFRVDLSDYN
                      IIGRLECIYRNLTHYLSGSSKKNSLYLIKYILLLSCVRTLARKHKSTVRTFAFQKKRL
                      CLELLEGFFISEEDILFLKFQKVYSSLGGVYISQFWYLNIYSINYLA"
     tRNA             complement(2259. .2330)
                      /gene="Gln-tRNA"
                      /note="codon recognized: GGG; anticodon ttg"
                      /product="tRNA-Gln"
                      /anticodon=(pos:2291. .2293,aa:Gln)
     gene             complement(2259. .2330)
                      /gene="Gln-tRNA"
     tRNA             complement(2604. .2691)
```

```
                     /gene="Ser-tRNA"
                     /note="codon recognized: AAC; anticodon gct"
                     /product="tRNA-Ser"
                     /anticodon=(pos:2638. .2640,aa:Ser)
     gene            complement(2604. .5149)
                     /gene="Ser-tRNA"
     tRNA            2715. .2782
                     /gene="pseudo-tRNA-Arg"
                     /note="anticodon tct"
                     /pseudo
     gene            2715. .2782
                     /gene="pseudo-tRNA-Arg"
     CDS             complement(2799. .3061)
                     /gene="pseudo-atpA"
                     /pseudo
     gene            complement(2799. .3061)
                     /gene="pseudo-atpA"
     gene            complement(3129. .3842)
                     /gene="rps2"
     CDS             complement(3129. .3842)
                     /gene="rps2"
                     /codon_start=1
                     /product="ribosomal protein S2"
                     /db_xref="PID:g336925"
                     /translation="MTRRYWNIDLEEMIGAGVSFGHGTRKWNPKMAPYISVKHKGIHF
                     TWLTKTARFLSEACDLVFYAASRGKQFLIVDTKNKAADSAAWAAIKARCHCVNKKWPG
                     GMLTNWSTTETRLHKLRDLIMEQKAGRLNRLKKKDEAAVKRQLARLQTYLGGIKYMTR
                     LPDIVIIVDQHEEYKALHECITIGIPTIGLIDTNCDPDLADISIPANDDAISSIRLIL
                     NKLVFAICEGRSGYIIKNI"
     tRNA            3965. .4033
                     /gene="pseudo-Cys-tRNA"
                     /note="anticodon gca"
     gene            3965. .4033
                     /gene="pseudo-Cys-tRNA"
     tRNA            complement(4438. .4511)
                     /gne="Asp-tRNA"
                     /note="codon recognized: CTT; anticodon gtc"
                     /product="tRNA-Asp"
                     /anticodon=(pos:4472. .4474,aa:Asp)
     gene            complement(4438. .4511)
                     /gene="Asp-tRNA"
     tRNA            complement(4661. .4745)
                     /gene="Tyr-tRNA"
                     /note="codon recognized: TTA; anticodon gta"
                     /product="tRNA-Tyr"
                     /anticodon=(pos:4695. .4697,aa:Tyr)
     gene            complement(4661. .4745)
                     /gene="Tyr-tRNA"
     tRNA            complement(4826. .4898)
                     /gene="Glu-tRNA"
                     /note="codon recognized: TTG; anticodon ttc"
```



```
                     /product="tRNA-Glu"
                     /anticodon=(pos:4860. .4862,aa:Glu)
     gene            complement(4826. .4898)
                     /gene="Glu-tRNA"
     tRNA            complement(5057. .5149)
                     /gene="Ser-tRNA"
                     /note="codon recognized: TTA; anticocon tga"
                     /product="tRNA-Ser"
                     /anticodon=(pos:5091. .5093,aa:Ser)
     tRNA            complement(5334. .5407)
                     /gene="fMet-tRNA"
                     /note="codon recognized: GTT; anticodon cat"
                     /product="tRNA-Met"
                     /anticodon=(pos:5368. .5370,aa:Met)
     gene            complement(5334. .5407)
                     /gene="fMet-tRNA"
     gene            complement(5541. .5843)
                     /gene="rps14"
     CDS             complement(5541. .5843)
                     /gene="rps14"
                     /codon_start=1
                     /product="ribosomal protein S14"
                     /db_xref="PID:g336926"
                     /translation="MARKSLIQREKGRLKLENKYHFIRRSSKNEISKVPSLSDKWEIY
                     GKLESPPRNSAPTRLRRRCFTTGRPRANYRDFGLCGHILREMVNACLLPGATRSSW"
     tRNA            6041. .6129
                     /gene="pseudo-Ser-tRNA"
                     /note="anticodon gga"
     gene            6041. .6129
                     /gene="pseudo-Ser-tRNA"
     gene            complement(6312. .6920)
                     /gene="rps4"
     CDS             complement(6312. .6920)
                     /gene="rps4"
                     /codon_start=1
                     /product="ribosomal protein S4"
                     /db_xref="PID:g336927"
                     /translation="MSRYRGPSLKKIRRLGALPGLTNKRSKAENDFIKKLRSDKKSQY
                     RIRLEEKQKLRFNYGLRERQLRKYFSIAIKTRGSTGKVLMQLLEMRLDNIIFRLGMAS
                     TIPAARQLVNHRHVLVNGRIVDIPSYRCKSRDIIMARDEQQSNTFINNCINYSTHNRM
                     EAPNHLTLLHPFKGLVNQIIDSKWVGFKINELLVVEYYFRKT"
     repeat_region   6998. .7043
                     /rpt_type=tandem
                     /rpt_unit=6998. .7020
     repeat_unit     6998. .7020
     tRNA            7285. .7357
                     /gene="Phe-tRNA"
                     /note="codon recognized: TTC; anticodon gaa"
                     /product="tRNA-Phe"
                     /anticodon=(pos:7318. .7320,aa:Phe)
     gene            7285. .7357
```



```
                     /gene="Phe-tRNA"
     tRNA            7392. .7464
                     /gene="Met-tRNA"
                     /note="codon recognized: ATG; anticodon cat"
                     /product="tRNA-Met"
                     /anticodon=(pos:7425. .7427,aa:Met)
     gene            7392. .7464
                     /gene="Met-tRNA"
     CDS             complement(7625. .7774)
                     /gene="pseudo-atpB"
                     /pseudo
     gene            complement(7625. .7774)
                     /gene="pseudo-atpB"
     repeat_region   7683. .7754
                     /rpt_type=tandem
                     /rpt_unit=7683. .7718
     repeat_unit     7683. .7718
                     /gene="pseudo-atpB"
     CDS             7768. .8202
                     /gene="pseudo-rbcL"
                     /pseudo
     gene            7768. .8202
                     /gene="pseudo-rbcL"
     gene            8291. .9772
                     /gene="accD"
     CDS             8291. .9772
                     /gene="accD"
                     /note="ORF 493"
                     /codon_start=1
                     /product="acetyl-CoA carboxylase"
                     /db_xref="PID:g336928"
                     /translation="MERWRFSLMVPNNKLKHEYGIKKSLDSVSSIENTIENKDSNSKP
                     NIKGRAQNIRSCGGRDNYSYNNIIYLFGVEYNRNLIPNDTFLPRDNKGDSFYISFNIE
                     NHIFEIDINFSCLSELEIERYFYSYHGFDYFNFNNGSAIEDYLYNHYMYDTPSNNHIT
                     SCIESYLQSQICVNPSIISDSSDSYIFGCFDKRRTHNESRRSGIRNRAQSSYLTIRES
...
Note: remainder of annotations omitted.

  Query Match           19.5%;  Score 17;  DB 23;  Length 70028;
  Best Local Similarity  94.4%;  Pred. No. 1.92e+01;
  Matches    17;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db  31718 TAATAAAATCTATCTTTA 31735
          ||||||||||||||| ||
Qy     63 TAATAAAATCTATCTNTA 80


RESULT   40
LOCUS       EPFCPCG     70028 bp    DNA   circular  PLN       05-MAY-1995
DEFINITION  Epifagus virginiana chloroplast complete genome.
ACCESSION   M81884
```



```
NID         g336917
KEYWORDS    complete genome; maturase; protease; pseudogene; ribosomal protein;
            transfer RNA-Arg; transfer RNA-Asn; transfer RNA-Gln; transfer
            RNA-Glu; transfer RNA-His; transfer RNA-Ile; transfer RNA-Leu;
            transfer RNA-Met; transfer RNA-Phe; transfer RNA-Pro; transfer
            RNA-Ser; transfer RNA-Trp; transfer RNA-Tyr; transfer RNA-fMet;
            translation initiation factor 1.
SOURCE      Chloroplast Epifagus virginiana DNA.
  ORGANISM  Chloroplast Epifagus virginiana
            Eukaryotae; mitochondrial eukaryotes; Viridiplantae;
            Charophyta/Embryophyta group; Embryophyta; Magnoliophyta;
            Magnoliopsida; Gentiananae; Scrophulariales; Orobanchaceae;
            Epifagus.
REFERENCE   1  (bases 1 to 70028)
  AUTHORS   Wolfe,K.H., Morden,C.W., Ems,S.C. and Palmer,J.D.
  TITLE     Rapid evolution of the plastid translational apparatus in a
            nonphotosynthetic plant: loss or accelerated sequence evolution of
            tRNA and ribosomal protein genes
  JOURNAL   J. Mol. Evol. 35 (4), 304-317 (1992)
  MEDLINE   93021155
REFERENCE   2  (sites)
  AUTHORS   Wolfe,K.H., Morden,C.W. and Palmer,J.D.
  TITLE     Function and evolution of a minimal plastid genome from a
            nonphotosynthetic parasitic plant
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 89 (22), 10648-10652 (1992)
  MEDLINE   93066301
FEATURES             Location/Qualifiers
     source     .    1. .70028
                     /organism="Epifagus virginiana"
                     /chloroplast
                     /db_xref="taxon:4177"
     misc_feature    1. .19799
                     /note="large single copy region (LSC)"
     CDS             complement(234. .569)
                     /gene="pseudo-psbA"
                     /pseudo
     gene            complement(234. .569)
                     /gene="pseudo-psbA"
     gene            complement(588. .1907)
                     /gene="matK"
     CDS             complement(588. .1907)
                     /gene="matK"
                     /codon_start=1
                     /db_xref="PID:g336924"
                     /translation="MYKQNLFIISSNDSKYNFLLRNKKLYYEIILEGLTFILEFFFFI
                     NLISYLEGKKNKIEKLNNLRSILSIFPPLEDNFSNLKFILDILIPRKVHAEILVQTLR
                     YWIKDATYLHLLRFWLNNNWNSIINQNKDGYFYPQKNKILLLLYNSYVCKYESIFVFL
                     RNQSYHFRSTPFIMLLERIYFYFKIERLVNPFLKVKDFKANLWLIKEPCLHYFRYRRQ
                     YILASKGTSLFLNKWKCYVITFWQWYFSLWFFSRSIYIKKLLNNSFEILTYHSSLNIN
                     PSFVLSQILETPFISNNTIKKVNILVPIILSISDKEKFPNVLGHPTSNLFRVDLSDYN
                     IIGRLECIYRNLYHYLSGSSKKNSLYLIKYILLLSCVRTLARKHKSTVRTFAFQKKRL
```

```
                     CLELLEGFFISEEDILPLKPQKVYSSLGGVYISQFWYLNIYSINYLA"
     tRNA            complement(2259. .2330)
                     /gene="Gln-tRNA"
                     /note="codon recognized: GGG; anticodon ttg"
                     /product="tRNA-Gln"
                     /anticodon=(pos:2291. .2293,aa:Gln)
     gene            complement(2259. .2330)
                     /gene="Gln-tRNA"
     tRNA            complement(2604. .2691)
                     /gene="Ser-tRNA"
                     /note="codon recognized: AAC; anticodon gct"
                     /product="tRNA-Ser"
                     /anticodon=(pos:2638. .2640,aa:Ser)
     gene            complement(2604. .5149)
                     /gene="Ser-tRNA"
     tRNA            2715. .2782
                     /gene="pseudo-tRNA-Arg"
                     /note="anticodon tct"
                     /pseudo
     gene            2715. .2782
                     /gene="pseudo-tRNA-Arg"
     CDS             complement(2799. .3061)
                     /gene="pseudo-atpA"
                     /pseudo
     gene            complement(2799. .3061)
                     /gene="pseudo-atpA"
     gene            complement(3129. .3842)
                     /gene="rps2"
     CDS             complement(3129. .3842)
                     /gene="rps2"
                     /codon_start=1
                     /product="ribosomal protein S2"
                     /db_xref="PID:g3336925"
                     /translation="MTRRTWWIDLEEMIGAGVSFGHGTRKWNPKMAPYISVKHKGIHF
                     IWLTKTARFLSEACDLVFTAASRGKQFLIVDTKNKAADSAAWAAIKARCHCVNKKWPG
                     GMLTNWSTTETRLHKLRDLIMEQKAGRLNRLKKKDEAAVKRQLARLQTYLGGIKYMTR
                     LPDIVIIVDQREEYKALHECITIGIPTIGLIDTNCDPDLADISIPANDDAISSIRLIL
                     NKLVFAICEGRSGYIIKNI"
     tRNA            3965. .4033
                     /gene="pseudo-Cys-tRNA"
                     /note="anticodon gca"
     gene            3965. .4033
                     /gene="pseudo-Cys-tRNA"
     tRNA            complement(4438. .4511)
                     /gene="Asp-tRNA"
                     /note="codon recognized: CTT; anticodon gtc"
                     /product="tRNA-Asp"
                     /anticodon=(pos:4472. .4474,aa:Asp)
     gene            complement(4438. .4511)
                     /gene="Asp-tRNA"
     tRNA            complement(4661. .4745)
```



```
                     /gene="Tyr-tRNA"
                     /note="codon recognized: TTA; anticodon gta"
                     /product="tRNA-Tyr"
                     /anticodon=(pos:4695. .4697,aa:Tyr)
     gene            complement(4661. .4745)
                     /gene="Tyr-tRNA"
     tRNA            complement(4826. .4898)
                     /gene="Glu-tRNA"
                     /note="codon recognized: TTG; anticodon ttc"
                     /product="tRNA-Glu"
                     /anticodon=(pos:4860. .4862,aa:Glu)
     gene            complement(4826. .4898)
                     /gene="Glu-tRNA"
     tRNA            complement(5057. .5149)
                     /gene="Ser-tRNA"
                     /note="codon recognized: TTA; anticocon tga"
                     /product="tRNA-Ser"
                     /anticodon=(pos:5091. .5093,aa:Ser)
     tRNA            complement(5334. .5407)
                     /gene="fMet-tRNA"
                     /note="codon recognized: GTT; anticodon cat"
                     /product="tRNA-Met"
                     /anticodon=(pos:5368. .5370,aa:Met)
     gene            complement(5334. .5407)
                     /gene="fMet-tRNA"
     gene            complement(5541. .5843)
                     /gene="rps14"
     CDS             complement(5541. .5843)
                     /gene="rps14"
                     /codon_start=1
                     /product="ribosomal protein S14"
                     /db_xref="PID:g3336926"
                     /translation="MARKSLIQREKGRLKLENKYHFIRRSSKNEISKVPSLSDKWEIY
                     GKLESPPRNSAPTRLRRRCFYTGRPRANYRDFGLCGHILREMVNACLLPGATRSSW"
     tRNA            6041. .6129
                     /gene="pseudo-Ser-tRNA"
                     /note="anticodon gga"
     gene            6041. .6129
                     /gene="pseudo-Ser-tRNA"
     gene            complement(6312. .6920)
                     /gene="rps4"
     CDS             complement(6312. .6920)
                     /gene="rps4"
                     /codon_start=1
                     /product="ribosomal protein S4"
                     /db_xref="PID:g3336927"
                     /translation="MSRIRGPSLKKIRRLGALPGLTNKRSKAENDFIKKLRSDKKSQY
                     RIRLEEKQKLRFNYGLRERQLRKYFSIAIKTRGSTGKVLMQLLEMRLDNIIFRLGMAS
                     TIPAARQLVNHRHVLVNGRIVDIPSYRCKSRDIIMARDEQQSNTFINNCINYSTHNRM
                     EAPNHLTLLHPFKGLVNQIIDSKWVGFKINELLVVEYYFRKT"
     repeat_region   6998. .7043
```



```
                     /rpt_type=tandem
                     /rpt_unit=6998. .7020
     repeat_unit     6998. .7020
     tRNA            7285. .7357
                     /gene="Phe-tRNA"
                     /note="codon recognized: TTC; anticodon gaa"
                     /product="tRNA-Phe"
                     /anticodon=(pos:7318. .7320,aa:Phe)
     gene            7285. .7357
                     /gene="Phe-tRNA"
     tRNA            7392. .7464
                     /gene="Met-tRNA"
                     /note="codon recognized: ATG; anticodon cat"
                     /product="tRNA-Met"
                     /anticodon=(pos:7425. .7427,aa:Met)
     gene            7392. .7464
                     /gene="Met-tRNA"
     CDS             complement(7625. .7774)
                     /gene="pseudo-atpB"
                     /pseudo
     gene            complement(7625. .7774)
                     /gene="pseudo-atpB"
     repeat_region   7683. .7754
                     /rpt_type=tandem
                     /rpt_unit=7683. .7718
     repeat_unit     7683. .7718
                     /gene="pseudo-atpB"
     CDS             7768. .8202
                     /gene="pseudo-rbcL"
                     /pseudo
     gene            7768. .8202
                     /gene="pseudo-rbcL"
     gene            8291. .9772
                     /gene="accD"
     CDS             8291. .9772
                     /gene="accD"
                     /note="ORF 493"
                     /codon_start=1
                     /product="acetyl-CoA carboxylase"
                     /db_xref="PID:g3336928"
                     /translation="MERWRPSLMVFNNKLKHEYGIKKSLDSVSSIENTIENKDSNSKP
                     NIKGRAQNIRSCGGRDNYSYNNIIYLPGVEYNRNNLIPNDTPLFRDNKGDSFYISFNIE
                     NHIFEIDINPSCLSELEIERYFTSYHGFDYFNFNNGSAIEDYLYNHYMYDTPSNNHIT
                     SCIESYLQSQICVNPSIISDSSDSYIPGCFDKRRTHNESRRSGIRNRAQSSYLTIRES
...
Note: remainder of annotations omitted.

  Query Match          19.5%;  Score 17;  DB 23;  Length 70028;
  Best Local Similarity  94.4%;  Pred. No. 1.92e+01;
  Matches   17;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;
```



```
Db  58093 TAAAGATAGATTTTATTA 58110
          || ||||||||||||||||
Cp     80 TANAGATAGATTTTATTA 63


RESULT   41
LOCUS       TM017A05    78344 bp    DNA             PLN       09-OCT-1997
DEFINITION  Arabidopsis thaliana BAC TM017A05.
ACCESSION   AP024504
NID         g2435510
KEYWORDS    .
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryotae; Viridiplantae; Charophyta/Embryophyta group;
            Embryophyta; vascular plants; seed plants; Magnoliophyta;
            Magnoliopsida; Capparales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 78344)
  AUTHORS   Zhu,H., Graves,T. and Ozersky,P.
  TITLE     The sequence of A. thaliana TM017A05
  JOURNAL   Unpublished (1997)
REFERENCE   2  (bases 1 to 78344)
  AUTHORS   Washington University Genome Sequencing Center.
  TITLE     The A. thaliana Genome Sequencing Project
  JOURNAL   Unpublished (1997)
REFERENCE   3  (bases 1 to 78344)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-SEP-1997) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   4  (bases 1 to 78344)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-OCT-1997) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     Submitted by:
                    Genome Sequencing Center
                    Department of Genetics, Washington University,
                    St. Louis, MO 63108, USA
                    e-mail: rwilson@watson.wustl.edu


            MAPPING: Clones were assigned to the YAC map by hybridization by
            M. Lohdi, Cold Spring Harbor Laboratories, and fingerprinted
            by M. Marra, WashU, to pick the best candidates for sequencing.


            NOTICE:  This sequence may not be the entire insert of this clone.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we provide a small overlap between
            neighboring submissions.
```

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:

The actual start of this clone is at base position 1 of TM017A05; the actual end is at base position 78344 of TM017A05. The orientation of this clone is unknown.

NOTES:

Coding sequences below are predicted from computer analysis, using the program Genefinder(P. Green and L. Hillier, ms in preparation).

```
FEATURES             Location/Qualifiers
     source          1. .78344
                     /organism="Arabidopsis thaliana"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="IV"
                     /clone="TM017A05"
                     /map="4"
     gene            164. .3072
                     /gene="A_TM017A05.5"
     CDS             join(164. .304,510. .636,734. .823,908. .1102,1203. .1272,
                     1367. .1409,1722. .2765,2836. .3072)
                     /gene="A_TM017A05.5"
                     /note="similar to prokaryotic sensory transduction
                     protien"
                     /codon_start=1
                     /evidence=not_experimental
                     /db_xref="PID:g2435516"
                     /translation="MAAVIGLLDILISDDCLSNEQYATVTQIRKCSTALLRLLNNILD
                     LSKVESGKLVLEEAEPDLGRELEGLVDMFSVQCINHNVETVLDLSDDMPALVRGDSAR
                     LVQIFANLISNSIKFTTTGHIILRGWCENINSLHDEMSVSVDRRKPWAPMKTKQVQHR
                     NHLQKSCKNANKMVLWPEVDDTGCGIDPSKWDSVPESPEQADPSTTRTHGGTGLGLCI
                     VRNLVMLSMYGSTARMITSKWLRKHGIATVEASDWNELTQIIRDLLETGSRDNSPDSQ
                     HNISDPLRAELSNIVEIKNPVFVIVVDIGVLDLTTNIWKEQLNYLDRPSNKAKFAWLL
                     KHDTSNTVKTELRRKGHVMMVNKPLYKAKMIQILEAVIKNNRKRGLCNDLRNRGNGSDE
                     SHDCLEIDPTQPDTCSSDDSSETSGEKQVDKSVKPSTLHSPVLKNYLIDATTSNDDST
                     SASMTQKNPEEDWKDRLYSGIALDGKNQRSLEGIRILLAEDTPVLQRVATIMLEKMG
                     ATVTAVWDGQQAVDSLNYKSINAQAPTEEHKSFEEETANKVTTRETSLRNSSPYDLIL
                     MDCQMPKMDGYEATKAIRRAEIGTELHIPIVALTAHAMSSDEAKCLEVGMDAYLTKPI
                     DRKLMVSTILSLTKPSAPQTSLSA"
     tRNA            complement(3420. .3493)
                     /note="codon recognized: ATT; Ile"
                     /product="tRNA-Ile"
```



```
     gene            complement(4263. .5892)
                     /gene="ARAC1"
     CDS             complement(join(4263. .4364,4455. .4520,4612. .4676,
                     5151. .5214,5314. .5423,5513. .5600,5794. .5892))
                     /gene="ARAC1"
                     /note="A_TM017A05.6"
                     /codon_start=1
                     /product="Arabidopsis thaliana GTP-binding protein (ARAC1)
                     (NID:g1292907)"
                     /db_xref="PID:g2435520"
                     /translation="MSASRFIKCVTVGDGAVGKTCLLISYTSNTFPTDYVPTVFDNFS
                     ANVVVNGATVNLGLWDTAGQEDYNRLRPLSYRGADVFILAFSLISKASYENVSKKWIP
                     ELKHYAPGVPIVLVGTKLDLRDDKQFFIDHPGAVPITTAQGEELKKLIGAPAYIECSS
                     KTQENVKGVFDAAIRVVLQPPKQKKKKSKAQKACSIL"
     gene            complement(6861. .12520)
                     /gene="A_TM017A05.7"
     CDS             complement(join(6861. .7062,7142. .7255,7328. .7461,
                     7521. .7628,7732. .7812,7887. .7955,8201. .8290,8480. .8569,
                     8822. .8911,9028. .9093,9179. .9418,9462. .9516,9561. .9634,
                     9842. .9958,10314. .10448,10593. .10715,10806. .10919,
                     10990. .11046,11243. .11330,11519. .11704,11941. .12064,
                     12170. .12266,12428. .12520))
                     /gene="A_TM017A05.7"
                     /note="similar to mouse MEM3 (GB:U47024 and S. cerevisiae
                     vacuolar sorting protein 35 (SW:P34110)"
                     /codon_start=1
                     /evidence=not_experimental
                     /db_xref="PID:g2435519"
                     /translation="MIADGSEDEEKWLAAGAAAFKQNAFYMQRAIDSNNLKDALKYSA
                     QMLSELRTSKLSPHKYYDLYMRAFDELRKLEIFFMEETRRGCSVIELYELVQHAGNIL
                     PRLYLLCTAGSVYIKTKEAPAKEILKDLVEMCRGIQHPLRGLFLRSYLAQISRDKLPD
                     IGSEYEGDADTVIDAVEFVLLNFTEMNNKLWVRMQHQGPAREKERRREKERGELRDLRTF
                     SLTYQVGKNLHVLSQLEGVDLDMYRDTVLPRVLEQIVNCRDEIAQYYLIDCIIQVFPD
                     EYHLQTLDVLLGACPQLQASVDIMTVLSRLMERLSNYAALNAEVLPYFLQVEAFSKLN
                     NAIGKVIEAQEDMPILSAVTLYSSLLKFTLHVHPDRLDYADQVLISGSSLFIILVADI
                     GSGIYVLFCYIKLPTTYLIRIPCWNRVHWGSCVKQLSGKGKIDDTRATKELVSLLSAP
                     LEKTNDVVTALKLTNTPLDVETYLDTETKRIMATVIVRSIMKNNTLITTAEKVEALFEL
                     IKGIINDLDEPQGLEVDEDDPQEEQNSVALLIMDLPNDDPEMDPIVNVLKKHFLTGG
                     PKRLKFTIPPLVVSTLKLIRRLPVEGDNPPGKEASVTATKIPQFLNQIIEALPNVPSP
                     DLAFRLYLQCAEAADKCDEEPIAYEFFTQAYILYEEEISDSKAQVTALQLIIGTLQRM
                     QVPGVENRDTLTHKATGADKGKLILLQYAAKLLKKPDQCRAVYACSHLFWLEDRETIQ
                     DGERVLLCLKRALKIANSAQQVANTARGSTGSVTLPIEILNKYLYFYEKGVPQITVES
                     VESLIKLIKNEESMPSDPSAESFFATTLEFMEFQKQKEGAIGERYQAIKV"
     gene            19253. .20158
                     /gene="A_TM017A05.4"
     CDS             join(19253. .19663,19739. .20158)
                     /gene="A_TM017A05.4"
                     /codon_start=1
                     /evidence=not_experimental
                     /db_xref="PID:g2435512"
                     /translation="MANSWDQLGEIASVAQLTGIDALKLIGMIVNAANTARMHKKNCR
```



```
                     QFAHHLKLIRNLLEQIKNSEMNQRSEILEPLQGLDDALRRSYILVKSCQEKSYLYLLA
                     MGWNIVNQFEKAQNEIDLFLKIVPLINMADNARIRERLEAIERDQREYTLDEEDRKVQ
                     DVILKQESTREAATSVLKKTLSRSYPNMGFCEALKTEEEKLQLELQRSRARYDADQCE
                     VIQRLIDVTQTAATVEPNLEKVLTKKEELTSSKKRDDLLYDTDSSSIRADSRLIPPTPI
                     "
     gene            complement(29936. .32522)
                     /gene="A_TM017A05.8"
     CDS             complement(join(29936. .30132,30235. .30306,30405. .30455,
                     30542. .30545,30546. .30645,30698. .30801,30877. .31012,
                     31119. .31343,31487. .31722,32125. .32260,32315. .32362,
                     32413. .32522))
                     /gene="A_TM017A05.8"
                     /note="contains similarity to peptidase family A1"
                     /codon_start=1
                     /evidence=not_experimental
                     /db_xref="PID:g2435517"
                     /translation="MAHRDRLIRGRRLANEDQSLVTFSDGNETVRVDALGPFKVNVFM
                     ETCELFMRDLHYANVTVGTPSDWFMVALDTGSDLFWLPCDCTNCVRELKAPGGSSLDL
                     NIYSPNASSTSTKVPCNSTLCTRGDRCASPESDCPYQIRYLSNGTSSTGVLVEDVLHL
                     VSNDKSSKAIPARVTFGCGQVQTGVFHDGAAPNGLFGLGLEDISVPSVLAKEGIAANS
                     FSMCFGNDGAGRISFGDGGSVDQRETPLNIRQPHPTYNITVTKISVGGNTGDLEFDAV
                     FDSGTSFTYLTDAAYTLISESFNSLALDKRYQTTDSELPFEYCYALRLPLYSGHHNPN
                     KDSFQYPAVNLTMKGGGSYPVYHPLVVIPMKVNVYCLAIMKIEDISIIGQNFMTGYRV
                     VFDREKLILGWKESDCYTGETSARTLPSNRSSSSARPPASSPDPEATNIPSQRPNTST
                     TSAAYSLSISLSLFPFSILAIL"
     gene            35321. .39123
                     /gene="A_TM017A05.3"
     CDS             join(35321. .35374,35634. .35867,36593. .36894,38340. .39123)
                     /gene="A_TM017A05.3"
                     /codon_start=1
                     /evidence=not_experimental
                     /db_xref="PID:g2435515"
                     /translation="MNNHPIYDHPLFKNYTIQMKPSSYPKGKNNESSDKEKQSVVTQL
                     WTVNGKCPKNSIPIRRTRRKEILRTETMQRYDKKNPNIINHPKASTSNSIHEDPNSGN
                     WWLKFGDEFVGYWPSILFNHLKDGATEIQWGGEIINFKDGALHTTTRMGSGHFAESGY
                     QKASYFKDVEIIDERDIHSSPKEGYSYMTQESCYNIRIRPHSKSRILPLMASRPSVRH
                     PSHNHPLRGHKAQVEDEIICSGCDLDLLGAYFKCTKSECDYFLHKSCFDLPREIRHKS
                     HPDHPLILLYSPQNNNSTYTCDACGEYGSGFTYNCSICQYDVHVGCVSVPETMKHDEH
                     VHPLALIYKAPCPKDHIFTCDVCDETMPHNLWLYYCQKCDYGAHLHSCVAEEEEKSKK
                     GGRGRGGEGGSGGNGVNRGRSSANSELAAMLKAQREMEQMQIALHLEMQRAKIDKKSR
                     KNMLKMI"
     gene            complement(40666. .42063)
                     /gene="A_TM017A05.9"
     CDS             complement(join(40666. .40716,40773. .40854,40912. .41015,
                     41098. .41265,41368. .41451,41540. .41718,41853. .41891,
                     41892. .42063))
                     /gene="A_TM017A05.9"
                     /note="contains similarity to C3HC4-type zinc fingers"
                     /codon_start=1
                     /evidence=not_experimental
                     /db_xref="PID:g2435518"
```



```
                     /translation="MASSSSSSYRFQSGSYPLSSSPSLGNFVERIKDACHFLVSAVLG
                     TIISAILTFPFALGLFSFLVTLLEIVLGTLLGALTGALIGQETESGFIRGAAIGAISG
...
Note: remainder of annotations omitted.

  Query Match           19.5%;  Score 17;  DB 23;  Length 78344;
  Best Local Similarity  94.4%;  Pred. No. 1.92e+01;
  Matches    17;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db  16901 ATGTAGAGATAGATTTTA 16918
          ||||| ||||||||||||
Cp      83 ATGTANAGATAGATTTTA 66


RESULT   42
LOCUS       HS454G6     79376 bp    DNA             PRI       12-FEB-1998
DEFINITION  Human DNA sequence from PAC 454G6 on chromosome 1q24. Contains
            trabecular meshwork inducible glucocorticoid response protein,
            TIGR, myocilin, ESTs and STS.
ACCESSION   Z98750
NID         g2887277
KEYWORDS    1q24; myocilin; TIGR.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 79376)
  AUTHORS   Deadman,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-OCT-1997) sanger.ac.uk/HGP/Chr1/) Sanger Centre,
            Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquires:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Feb 14, 1998 this sequence version replaced gi:2465060.
            IMPORTANT: This sequence is not the entire insert of clone 454G6.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we arrange for a small overlap between
            neighbouring submissions.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variations annotated may not be found in the sequence submission
            corresponding to the overlapping clone as we submit sequences with
            only a small overlap as described above.
            This sequence was generated from part of bacterial clone contigs of
            human chromosome 1, constructed by the Sanger Centre chromosome 1
            mapping group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1/
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
```

```
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            The true left end of clone 454G6 is at 1 in this sequence. The true
            left end of clone 56089 is at 79273.
            454G6 is from the library RPCI3 constructed at the Roswell Park
            Cancer Institute by the group of Pieter de Jong.
            For further details see http://bacpac.med.buffalo.edu/.
FEATURES              Location/Qualifiers
     source           1. .79376
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="1"
                      /map="1q24"
                      /clone="454G6"
                      /clone_lib="RPCI3"
     repeat_region    435. .472
                      /note="19 copies of 2 mer 82 % conserved"
     prim_transcript  1914. .>3968
                      /note="match: multiple ESTs; match: R56676 AA043968 W63639
                      F12081 AA046699; match: F02925 AA131540 W00634 R36066
                      AA313383; match: AA163561 F02925 AA131540 W00634 R36066;
                      match: AA313383 AA163561 N89173 AA174814 AA057059; match:
                      AA329084 W47082 AA043955 AA341783 AA353681; match:
                      AA046487 AA369741 H08313 AA186895 H32730; match: H08333
                      H08236 N42052 D61944 R27102 N32353; match: N30491 AA307150
                      AA192"
     repeat_region    3703. .3746
                      /note="22 copies of 2 mer 89 % conserved"
     repeat_region    4051. .4183
                      /note="AluSq repeat: matches 1. .133 of consensus;
                      incomplete repeat"
     repeat_region    4200. .4502
                      /note="AluSq repeat: matches 2. .301 of consensus"
     repeat_region    4659. .4851
                      /note="AluSx repeat: matches 2. .194 of consensus;
                      incomplete repeat"
     repeat_region    5216. .5345
                      /note="AluJo repeat: matches 132. .1 of consensus;
                      incomplete repeat"
     repeat_region    7759. .7907
                      /note="MIR repeat: matches 174. .1 of consensus"
     repeat_region    7933. .9328
                      /note="TIGGER1 repeat: matches 1. .1472 of consensus"
     repeat_region    9332. .9626
                      /note="AluSg repeat: matches 1. .289 of consensus"
     repeat_region    9639. .10335
                      /note="TIGGER1 repeat: matches 1469. .2174 of consensus"
     repeat_region    10343. .10642
                      /note="AluSp repeat: matches 1. .300 of consensus"
     repeat_region    10643. .10856
                      /note="TIGGER1 repeat: matches 2175. .2417 of consensus"
```



```
     misc_feature     <10903. .>16855
                      /note="endogenous retroviral sequence"
     repeat_region    10949. .11384
                      /note="LTR2 repeat: matches 31. .449 of consensus"
     repeat_region    15835. .15870
                      /note="18 copies of 2 mer 83 % conserved"
     repeat_region    16856. .17286
                      /note="LTR2 repeat: matches 31. .449 of consensus"
     repeat_region    17287. .17574
                      /note="AluSg repeat: matches 15. .300 of consensus"
     repeat_region    18294. .18650
                      /note="THE1B repeat: matches 358. .1 of consensus"
     repeat_region    18877. .19180
                      /note="AluY repeat: matches 301. .2 of consensus"
     repeat_region    19767. .20013
                      /note="MLT2_internal repeat: matches 5002. .4750 of
                      consensus"
     repeat_region    20051. .20118
                      /note="MLT1F repeat: matches 539. .471 of consensus"
     repeat_region    20130. .20261
                      /note="AluJb repeat: matches 131. .1 of consensus;
                      incomplete repeat"
     repeat_region    20264. .20722
                      /note="MLT1F repeat: matches 482. .1 of consensus"
     repeat_region    20858. .21223
                      /note="MSTD repeat: matches 394. .1 of consensus"
     repeat_region    21216. .21302
                      /note="MLT2_internal repeat: matches 4520. .4433 of
                      consensus"
     repeat_region    21403. .21703
                      /note="MLT2_internal repeat: matches 3887. .3580 of
                      consensus"
     repeat_region    21978. .22357
                      /note="MLT2_internal repeat: matches 3218. .2839 of
                      consensus"
     repeat_region    22363. .22524
                      /note="MLT2_internal repeat: matches 2495. .2317 of
                      consensus"
     repeat_region    22531. .22839
                      /note="AluSx repeat: matches 302. .1 of consensus"
     repeat_region    23007. .23309
                      /note="AluSp repeat: matches 3. .301 of consensus"
     misc_feature     23286. .>23680
                      /note="STS G07544"
     misc_feature     complement(23652. .24072)
                      /note="STS G07436"
     repeat_region    24008. .24290
                      /note="AluJo repeat: matches 298. .6 of consensus"
     repeat_region    25895. .26364
                      /note="LTR2 repeat: matches 2. .449 of consensus"
     repeat_region    26397. .26697
```



```
                      /note="AluSq repeat: matches 1. .302 of consensus"
     repeat_region    26719. .27021
                      /note="AluSx repeat: matches 1. .302 of consensus"
     repeat_region    29037. .29334
                      /note="AluSx repeat: matches 300. .3 of consensus"
     repeat_region    30028. .30310
                      /note="AluSx repeat: matches 1. .301 of consensus"
     repeat_region    30402. .30539
                      /note="MLT2B repeat: matches 264. .404 of consensus"
     repeat_region    30402. .30615
                      /note="MLT2A repeat: matches 264. .453 of consensus"
     prim_transcript  <30801. .>31136
                      /note="match: 5' EST AA256902 clone 682136"
     repeat_region    30980. .31222
                      /note="MER4B repeat: matches 199. .431 of consensus"
     repeat_region    31232. .31528
                      /note="AluSc repeat: matches 299. .1 of consensus"
     repeat_region    31530. .31654
                      /note="MER4B repeat: matches 417. .540 of consensus"
     repeat_region    31836. .32135
                      /note="AluY repeat: matches 300. .1 of consensus"
     repeat_region    32200. .32301
                      /note="AluJo repeat: matches 186. .287 of consensus;
                      incomplete repeat"
     repeat_region    32365. .32493
                      /note="MIR repeat: matches 206. .67 of consensus"
     repeat_region    33283. .33399
                      /note="MIR repeat: matches 77. .196 of consensus"
     repeat_region    33408. .33667
                      /note="AluY repeat: matches 1. .291 of consensus"
     repeat_region    33956. .34043
                      /note="MIR repeat: matches 80. .167 of consensus"
     repeat_region    34725. .34904
                      /note="AluJo repeat: matches 299. .127 of consensus;
                      incomplete repeat"
     repeat_region    34907. .35207
                      /note="AluY repeat: matches 300. .1 of consensus"
     repeat_region    35212. .35344
                      /note="AluJb repeat: matches 133. .1 of consensus;
                      incomplete repeat"
     repeat_region    36545. .36842
                      /note="AluJo repeat: matches 299. .1 of consensus"
     repeat_region    38190. .38379
                      /note="MER3 repeat: matches 209. .13 of consensus"
     repeat_region    38382. .38682
                      /note="AluSx repeat: matches 1. .302 of consensus"
     repeat_region    38836. .38968
                      /note="MIR2 repeat: matches 145. .2 of consensus"
     repeat_region    38983. .39279
                      /note="AluJb repeat: matches 3. .296 of consensus"
     repeat_region    39418. .39773
```



```
                      /note="THE1A repeat: matches 354. .2 of consensus"
     repeat_region    39775. .40120
                      /note="THE1B-INTERNAL repeat: matches 1580. .1234 of
                      consensus"
     repeat_region    40757. .41023
                      /note="AluJo repeat: matches 37. .301 of consensus;
                      incomplete repeat"
     repeat_region    41307. .41589
                      /note="AluSq repeat: matches 21. .303 of consensus;
                      incomplete repeat"
     repeat_region    41969. .42024
...
Note: remainder of annotations omitted.

  Query Match           19.5%;  Score 17;  DB 26;  Length 79376;
  Best Local Similarity 100.0%;  Pred. No. 1.92e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  64778 AAAGGAAAAATATAGTA 64794
          |||||||||||||||||
Qy     29 AAAGGAAAAATATAGTA 45


RESULT   43
LOCUS       ATF24J7     86554 bp    DNA             PLN       05-FEB-1998
DEFINITION  Arabidopsis thaliana DNA chromosome 4, BAC clone F24J7 (ESSAII
            project).
ACCESSION   AL021768
NID         g2853071
KEYWORDS    .
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryotae; Viridiplantae; Charophyta/Embryophyta group;
            Embryophyta; Tracheophyta; seed plants; Magnoliophyta;
            eudicotyledons; Rosidae; Capparales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 86554)
  AUTHORS   Bevan,M., Vitale,D., Liguori,R., Argiriou,A., De Simone,V.,
            Bancroft,I., Mewes,H.W., Mayer,K. and Schueller,C.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 86554)
  AUTHORS   EU Arabidopsis sequencing,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-FEB-1998) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, Project
            Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
            Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
            E-mail: michael.bevan@bbsrc.ac.uk
FEATURES              Location/Qualifiers
     source           1. .86554
                      /organism="Arabidopsis thaliana"
                      /variety="Columbia"
```

```
                     /db_xref="taxon:3702"
                     /chromosome="4"
     gene            177. .1542
                     /gene="F24J7.10"
     exon            177. .991
                     /gene="F24J7.10"
                     /number=1
     CDS             join(177. .991,1065. .1542)
                     /gene="F24J7.10"
                     /note="Protein sequence is in conflict with the conceptual
                     translation; similarity to gene T09D09.11, Arabidopsis
                     thaliana, PATCHX:G2347195; contains EST gb:F15379, F15378"
                     /codon_start=1
                     /product="putative protein"
                     /db_xref="PID:e1250336"
                     /db_xref="PID:g2853072"
                     /translation="SFNGVSAALYTLAYNAINPTSPELYLLLNALIPLIVSFTAIIPI
                     LRQPPPEPLPPDGVRRDSLMPLLLNILAALNGVYLLLFGSNSSDLTSARLLPGGAILL
                     LVFPLCIPGLVIARNWYNRTIHTSFRLEGSGFILVDPDELELHKGMLAHEANREGYQL
                     LSDDVVQNPVKSVAVEEEDSDESCCKKLITRDQLEGLGIEHSLSLLLLTRSDPWLYYIT
                     YFCGGTIGLVYSNNLGQIAQSLGQSSNTTTLVTLYSAPSFPGRLLSATPDYIRAKVYF
                     ARTGWLAIALLPTPFALFLLASSGTASALQAGTALMGLSSGFIFAAAVSITSELFGPN
                     SVGVNHNILITNIPIGSLIYGFLAALVYDSHGFTGTKSMTSESVVCMGRDCYYLTFVW
                     WGCLSLLGLGSSLVLFIRTRRAYQRFEQARISSNINDS"
     intron          992. .1064
                     /gene="F24J7.10"
                     /number=1
     exon            1065. .1542
                     /gene="F24J7.10"
                     /number=2
     gene            1071. .4227
                     /gene="F24J7.20"
     CDS             complement(join(1071. .1274,1940. .2548,2645. .2766,
                     2817. .4227))
                     /gene="F24J7.20"
                     /note="similarity to gene F21M12.4, Arabidopsis thaliana,
                     PATCHX:G2347201; contains EST gb:F15379"
                     /codon_start=1
                     /product="putative protein"
                     /db_xref="PID:e1250337"
                     /db_xref="PID:g2853073"
                     /translation="MASQTKLKKPNSHPSLCTFLFPTVLFTIPALFLLRTSTCSSSTA
                     AVSSSSSDTNQPPWSGDLQTAQFAWNRLDFSLTNPPPKTLKLAVFSRKWPTGPNPGGM
                     ERHAFTLFTALARRGHRVVFTSPDLQSPETNKIPPVSDQIIYPIIHSHGDAEPGKWR
                     YNKAWELYQEENKKEPFDAVHSESVALPHWIAREVPWLAVSWNGIALESLQSSIYQDL
                     IRKPDEPRSQGFNASLYGAVLPKILDEIRPFHNYAHHIAISDSCGEMLRDVYQIPEKR
                     VHVILNGVDENGFTSDKKLRTLFRSKLGLPENSSAIVLGAAGRLVKDKGHPLLFEAFA
                     KIIQTYSNVYLVVAGSGPWEQRYKELGEKVSILGSLNPNELKGFTNGIDLFVNPTLRP
                     QGLDLTLMEAMLSGKPVMASRYASIKRTIVVNDEFGFMFAPNVEALTAVMEVAVAEGA
                     ERLAERGRKCKEYAAEMFTAKILYLPLISKTQHLYFPHYWKSESTTHSNVIKYICDPY
                     ISKQSGDPDVLMVSVFPHHVINPTIESLRTLQLGSSSTFKIPTPWEKDNVLYLVSHSS
```



```
                     CDGLVCLYNHDKSGYVVNPTTRWYRPLPLCDYQKLMIDLRDGYYKLKRKFFMLGFGKD
                     KFRGTYKPVWLYNSAEIGLKNATTCEVFDPSTNAWRYVTLAAPYRIAAYPAPVYLDGS
                     LYWFTECKETKVLPFNLHTETFQVVCKAPFANILWLTPTLPGPKSSDVMETAAAKINP
                     ELKPISAVPACNADAVPDDASKKSAKGVGNSAIANHPVLAK"
     exon            complement(1071. .1274)
                     /gene="F24J7.10"
                     /number=1
     intron          complement(1275. .1939)
                     /gene="F24J7.20"
                     /number=1
     exon            complement(1940. .2548)
                     /gene="F24J7.20"
                     /number=2
     intron          complement(2549. .2644)
                     /gene="F24J7.20"
                     /number=2
     exon            complement(2645. .2766)
                     /gene="F24J7.20"
                     /number=3
     intron          complement(2767. .2816)
                     /gene="F24J7.20"
                     /number=3
     exon            complement(2817. .4227)
                     /gene="F24J7.20"
                     /number=4
     CDS             complement(join(5317. .6063,6115. .6524,6567. .6663))
                     /gene="F24J7.30"
                     /note="similarity to TMV resistance protein N, Nicotiana
                     glutinosa, PIR2:A54810"
                     /codon_start=1
                     /product="putative protein"
                     /db_xref="PID:e1250338"
                     /db_xref="PID:g2853074"
                     /translation="MDDTSIKQTPRTMCMSNLKLFSPGGSKVQDFRDMYLTDCNLYKF
                     PDNFSCLSSLQSLCLSRNSIENLPGSIKKLHHLKSLYLKNCKNLISLPVLPSNQYLDV
                     HGCISLETVSKPMTLLVIAEKTHSTFVFTDCYKLNRDAQEKIVAHTQLKSQILANRSF
                     QLGHKVQSLELVLEPLSAVSFPGNDLPLWFRHQRIGSSMETNLPSHWCDDKFIGLSLC
                     TVVSFKDYEDRTSRFSVICKCKFRNEDGDYISPTCNLGGWKEQCGSSSHEESRRLSSD
                     HVFISYSNCYHAKKNDDLNRCCNFTTASFKFFVTDGRAKRKLDCCEVVKCGMSLLYAPD
                     ENDCRLHGLHENNLEKEISGKDTELNETALDKVVVSKRGRLCVQEEELLNRKRIKEDV
                     SFGNYDCSVTSILSYMEPAKPHLWK"
     exon            complement(5317. .6063)
                     /gene="F24J7.30"
                     /number=1
     gene            5317. .6663
                     /gene="F24J7.30"
     intron          complement(6064. .6114)
                     /gene="F24J7.30"
                     /number=1
     exon            complement(6115. .6524)
                     /gene="F24J7.30"
```



```
                     /number=2
     intron          complement(6525. .6566)
                     /gene="F24J7.30"
                     /number=2
     exon            complement(6567. .6663)
                     /gene="F24J7.30"
                     /number=3
     gene            7104. .7640
                     /gene="F24J7.40"
     CDS             complement(join(7104. .7357,7508. .7640))
                     /gene="F24J7.40"
                     /codon_start=1
                     /product="hypothetical protein"
                     /db_xref="PID:e1250339"
                     /db_xref="PID:g2853075"
                     /translation="MTPLLTVSCVTSDTVDASTSSLDKGPSEEESPTDFKCGVSEAME
                     VSLVDLSTSSLDESPIGEENLNGFECGVSDATEEVFLYIILGFLADASTISLDEAPSEE
                     DNPSGFKCCVSEAMEDMFLCVDTNIG"
     exon            complement(7104. .7357)
                     /gene="F24J7.40"
                     /number=1
     intron          complement(7358. .7507)
                     /gene="F24J7.40"
                     /number=1
     exon            complement(7508. .7640)
                     /gene="F24J7.40"
                     /number=2
     exon            8341. .9174
                     /gene="F24J7.50"
                     /number=1
     gene            8341. .15047
                     /gene="F24J7.50"
     CDS             join(8341. .9174,9274. .9336,9426. .9505,10232. .10261,
                     10373. .10493,10612. .10686,10889. .11042,11426. .11544,
                     11759. .12015,12129. .12300,12385. .12468,12781. .13181,
                     13378. .13531,14210. .14229,14278. .14376,14650. .14717,
                     14959. .15047)
                     /gene="F24J7.50"
                     /note="similarity to probable membrane protein YDR027c,
                     Saccharomyces cerevisiae, PIR2:S50934"
                     /codon_start=1
                     /product="putative protein"
                     /db_xref="PID:e1250340"
                     /db_xref="PID:g2853076"
                     /translation="MDSHPSLMGRSITNSNRSSLDLGRPSSSSSSSPSPLTKSISDAS
                     SQSLSSILNNPHGGKSGVYGSDASWVGWWSSSTFVAPAEPAPVASTKLPGSELTRSDF
                     HGYVSSISESHGRPEDIRKHTREESCGFDQESHVSGLAACLREVPSLYFKEDPALEDG
                     ATFRSACPFSSLNENLAILQEKLSQYLDVVELHLVKEISVRSDSFFEAQGQLQDLNVKI
                     VEGCSRIRELKETIRLLDRNLVDSARQIQELSSTRINMLELQRKLRLILYVNQALSAL
                     KLLVASADCAGALDITDDLQNLLAGDELTGLYCFRHLRDHVTSSIDSINSELFRENPT
                     IIILTSEPMRISIHDTGEIDVLILSAANIRGSISSNGNTGEEVKLEEEDTSTLCDRLL
```



```
                     PLVIGLLRTAKFPSILRMYRDTLTSEMKNAIKKAVADLLPILVARSLESDFSHGERSV
                     DDETILTPSAYTGGGLSLASKLRTLSSEAFVNLLTAIFKIVQAHLVRASEVKKAIEWI
                     LCNIDGHYAADSVAAAIAVGAVAAETAQEIGFQGGSLVSSPLGKATSKAPPLQGKSSD
                     ASSLMNMSRNFRADVLRENTEAVFAACEVTHGRWAKLLGVRALLHPKLKLQEFMSIYD
                     LTQEFITSTEKIGGRLGSSIRGTLQSQAKAFVDSQHESRMTKLKAVLDQETWDEIDVP
                     EEPQSIISSLFASQRLISGKVDDADLNSYHSNRLPLNGSLTSGSGDQNSELRNEKSES
                     SEGSVVSDAQVTKTVSPESLERSKAGVSSATNNQSNQKAHGKSNFLFTQGVCYHMVNCG
                     LILLKMLSEYIDMNNSLPALSSEIVLRVVEVLRFFNTRTCQLVLGAGAMQRLIKLLSS
                     SGPGGRCTSNKRIVGKKFTPVGHLLLAIAVLMKLATFTVCCLRHCTKPMFKQSSGDFN
                     YPHTNLSNAYELGDKLNRGKEKVTL"
     intron          9175. .9273
                     /gene="F24J7.50"
                     /number=1
     exon            9274. .9336
                     /gene="F24J7.50"
                     /number=2
     intron          9337. .9425
                     /gene="F24J7.50"
                     /number=2
     exon            9426. .9505
...
Note: remainder of annotations omitted.

  Query Match           19.5%;  Score 17;  DB 23;  Length 86554;
  Best Local Similarity 100.0%;  Pred. No. 1.92e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  24477 ATTTTTCCTTTCTAAAA 24493
          |||||||||||||||||
Cp     39 ATTTTTCCTTTCTAAAA 23


RESULT   44
LOCUS       HS481A17   107428 bp    DNA             PRI       07-JUL-1998
DEFINITION  Human DNA sequence from PAC 481A17 on chromosome X contains ESTs.
ACCESSION   Z82212
NID         g1841914
KEYWORDS    X.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 107428)
  AUTHORS   McDougall,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-FEB-1997) E-mail enquires: humquery@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Feb 17, 1997 this sequence version replaced gi:1780978.
            IMPORTANT:  This sequence is the entire insert of clone 481A17. The
            true left end of clone 481A17 is at 1 in this sequence. The true
            right end of clone 481A17 is at 107428.
```

```
            481A17 is from the library RPCI3 constructed at the Roswell Park
            Cancer Institute by the group of Pieter de Jong. For further
            details see http://bacpac.med.buffalo.edu/ VECTOR: pCYPAC2.
FEATURES             Location/Qualifiers
     source          1. .107428
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="X"
                     /clone="481A17"
                     /clone_lib="RPCI3"
     repeat_region   1. .1883
                     /note="L1 repeat: matches 3306. .5184 of consensus"
     repeat_region   1878. .2428
                     /note="L1PA7 repeat: matches 322. .888 of consensus"
     repeat_region   2643. .3258
                     /note="L1 repeat: matches 4707. .5390 of consensus"
     repeat_region   3110. .3530
                     /note="L1MA1 repeat: matches 1. .421 of consensus"
     repeat_region   3528. .3832
                     /note="L1PA9 repeat: matches 546. .860 of consensus"
     repeat_region   3841. .4131
                     /note="L1PA7 repeat: matches 888. .595 of consensus"
     repeat_region   4131. .5271
                     /note="L1 repeat: matches 5356. .4209 of consensus"
     repeat_region   5274. .5550
                     /note="AluSx repeat: matches 4. .289 of consensus"
     repeat_region   5561. .6716
                     /note="L1 repeat: matches 4202. .3064 of consensus"
     repeat_region   6738. .7490
                     /note="L1PA14 repeat: matches 896. .139 of consensus"
     repeat_region   7496. .9368
                     /note="L1 repeat: matches 3074. .1214 of consensus"
     repeat_region   8977. .9561
                     /note="MER25 repeat: matches 2101. .1519 of consensus"
     repeat_region   10608. .10999
                     /note="L1 repeat: matches 1968. .1576 of consensus"
     repeat_region   10942. .11233
                     /note="MER25 repeat: matches 2135. .1839 of consensus"
     repeat_region   11236. .11384
                     /note="L1MC1 repeat: matches 1079. .934 of consensus"
     repeat_region   11424. .11514
                     /note="L1PA16 repeat: matches 904. .814 of consensus"
     repeat_region   11533. .12422
                     /note="L1PA2 repeat: matches 891. .1 of consensus"
     repeat_region   12273. .16771
                     /note="L1 repeat: matches 5390. .896 of consensus"
     repeat_region   16777. .17545
                     /note="L1 repeat: matches 763. .4 of consensus"
     repeat_region   17548. .17897
                     /note="L1PA15 repeat: matches 830. .509 of consensus"
```



```
     repeat_region   17900. .17969
                     /note="L1MC1 repeat: matches 948. .884 of consensus"
     repeat_region   19143. .20202
                     /note="L1MC3 repeat: matches 2486. .1410 of consensus"
     repeat_region   20736. .20850
                     /note="MIR2 repeat: matches 145. .22 of consensus"
     repeat_region   21129. .21521
                     /note="MSTA repeat: matches 3. .426 of consensus"
     repeat_region   21522. .23132
                     /note="MST-INTERNAL repeat: matches 1. .1638 of consensus"
     repeat_region   23145. .23549
                     /note="MSTA repeat: matches 1. .410 of consensus"
     repeat_region   23536. .24420
                     /note="L1PB3 repeat: matches 897. .1 of consensus"
     repeat_region   24276. .25658
                     /note="L1 repeat: matches 5389. .3999 of consensus"
     repeat_region   25898. .26288
                     /note="L1PA12 repeat: matches 912. .513 of consensus"
     repeat_region   26307. .26997
                     /note="L1PB2 repeat: matches 887. .206 of consensus"
     repeat_region   26993. .28004
                     /note="L1 repeat: matches 5057. .4047 of consensus"
     repeat_region   28004. .28336
                     /note="L1 repeat: matches 5056. .5389 of consensus"
     repeat_region   28189. .29080
                     /note="L1PA2 repeat: matches 1. .893 of consensus"
     repeat_region   29090. .29228
                     /note="L1PA3 repeat: matches 218. .83 of consensus"
     repeat_region   29229. .29452
                     /note="L1PA8 repeat: matches 687. .910 of consensus"
     repeat_region   29468. .29949
                     /note="L1 repeat: matches 5335. .4846 of consensus"
     repeat_region   29937. .30078
                     /note="L1 repeat: matches 4269. .4410 of consensus"
     repeat_region   30077. .31119
                     /note="L1 repeat: matches 4342. .5390 of consensus"
     repeat_region   30974. .31950
                     /note="L1PB3 repeat: matches 1. .897 of consensus"
     repeat_region   31969. .32437
                     /note="L1 repeat: matches 4388. .4855 of consensus"
     repeat_region   32610. .32714
                     /note="MIR repeat: matches 201. .93 of consensus"
     repeat_region   32841. .33194
                     /note="THE1A repeat: matches 1. .354 of consensus"
     repeat_region   33416. .38808
                     /note="L1 repeat: matches 4. .5390 of consensus"
     repeat_region   38659. .39549
                     /note="L1PA2 repeat: matches 1. .891 of consensus"
     repeat_region   39688. .39743
                     /note="MLT1A1 repeat: matches 365. .309 of consensus"
     repeat_region   40432. .40524
```



```
                     /note="MIR repeat: matches 106. .14 of consensus"
     repeat_region   41263. .41424
                     /note="MIR repeat: matches 96. .260 of consensus"
     repeat_region   42832. .46108
                     /note="L1 repeat: matches 1844. .5173 of consensus"
     repeat_region   46095. .46268
                     /note="L1MA3 repeat: matches 261. .430 of consensus"
     repeat_region   46280. .46611
                     /note="L1 repeat: matches 6. .363 of consensus"
     repeat_region   46824. .47054
                     /note="L1 repeat: matches 381. .586 of consensus"
     repeat_region   47131. .52047
                     /note="L1 repeat: matches 589. .5390 of consensus"
     repeat_region   51888. .52800
                     /note="L1PA10 repeat: matches 1. .906 of consensus"
     repeat_region   52795. .53255
                     /note="L1PA16 repeat: matches 420. .902 of consensus"
     prim_transcript <54102. .>56467
                     /note="match: multiple ESTs; match X78604 ARF-like protein
                     5; match: N49401 R77692 R29677 W06984 H98153; match:
                     N40498 N49490 N25929 H13517 T36067; match: N49468 Z32790"
     repeat_region   54497. .54783
                     /note="AluSg repeat: matches 290. .1 of consensus"
     repeat_region   55015. .55050
                     /note="18 copies of 2 mer 81 % conserved"
     repeat_region   56782. .57427
                     /note="L1PA4 repeat: matches 891. .220 of consensus"
     repeat_region   57437. .59159
                     /note="L1 repeat: matches 2465. .761 of consensus"
     repeat_region   60070. .60182
                     /note="L1 repeat: matches 116. .5 of consensus"
     repeat_region   60588. .60885
                     /note="AluSg repeat: matches 1. .297 of consensus"
     repeat_region   61001. .61304
                     /note="AluSx repeat: matches 302. .1 of consensus"
     repeat_region   61345. .61455
                     /note="MER5A repeat: matches 1. .111 of consensus"
     repeat_region   61556. .61622
                     /note="MER5A repeat: matches 122. .188 of consensus"
     repeat_region   62992. .63372
                     /note="L1PB1 repeat: matches 902. .514 of consensus"
     repeat_region   63495. .63800
                     /note="L1 repeat: matches 1265. .967 of consensus"
     repeat_region   63806. .63837
                     /note="16 copies of 2 mer 91 % conserved"
     repeat_region   63840. .64163
                     /note="L1PB2 repeat: matches 538. .207 of consensus"
     repeat_region   64185. .65064
                     /note="L1PA2 repeat: matches 891. .3 of consensus"
     repeat_region   64923. .65856
                     /note="L1 repeat: matches 5390. .4445 of consensus"
```



```
     repeat_region   65850. .66226
                     /note="L1 repeat: matches 4502. .4125 of consensus"
     repeat_region   66228. .66444
                     /note="L1PB1 repeat: matches 221. .1 of consensus"
     repeat_region   66299. .70317
                     /note="L1 repeat: matches 5390. .1340 of consensus"
     repeat_region   70018. .70998
                     /note="MER25 repeat: matches 2136. .1163 of consensus"
     repeat_region   70985. .71385
                     /note="MER25 repeat: matches 1031. .582 of consensus"
     repeat_region   71493. .72050
                     /note="MER25 repeat: matches 576. .12 of consensus"
     repeat_region   73650. .73951
                     /note="AluSx repeat: matches 302. .1 of consensus"
     repeat_region   74943. .75245
                     /note="MER11A repeat: matches 654. .358 of consensus"
     repeat_region   75537. .75578
                     /note="14 copies of 3 mer 83 % conserved"
     repeat_region   75579. .75802
                     /note="MIR repeat: matches 260. .32 of consensus"
     repeat_region   76800. .78171
                     /note="L1 repeat: matches 834. .2213 of consensus"
     repeat_region   78173. .78951
                     /note="L1 repeat: matches 4. .772 of consensus"
     repeat_region   78947. .80804
                     /note="L1 repeat: matches 896. .2753 of consensus"
     repeat_region   80786. .82175
                     /note="L1 repeat: matches 4002. .5390 of consensus"
     repeat_region   82027. .82918
...
Note: remainder of annotations omitted.

  Query Match           19.5%;  Score 17;  DB 26;  Length 107428;
  Best Local Similarity 100.0%;  Pred. No. 1.92e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  60504 TATTTTTCCTTTCTAAA 60520
          |||||||||||||||||
Cp     40 TATTTTTCCTTTCTAAA 24


RESULT   45
LOCUS       AC003092   108878 bp    DNA             PRI       06-NOV-1997
DEFINITION  Human BAC clone RG335O22 from 7q21-q22, complete sequence.
ACCESSION   AC003092
NID         g2588611
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE   1  (bases 1 to 108878)
  AUTHORS   Lacy,M and Harper,M.
  TITLE     The sequence of H. sapiens BAC clone RG335O22
  JOURNAL   Unpublished (1997)
REFERENCE   2  (bases 1 to 108878)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-NOV-1997) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     SUBMITTED BY:
            Genome Sequencing Center
            Department of Genetics
            Washington University
            St. Louis MO 63108, USA
            http://genome.wustl.edu/gsc
            mailto:sapiens@watson.wustl.edu

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded or sequenced with an alternate
            chemistry; an attempt was made to resolve all sequencing problems,
            such as compressions and repeats; all regions were covered by
            sequence from more than one subclone; and the assembly was
            confirmed by restriction digest.

            MAPPING INFORMATION:
            The sequence of this clone was established as part of a mapping and
            sequencing collaboration between the NHGRI Chromosome 7 Mapping
            Project (Eric D. Green, Director), John D. McPherson in the
            Department of Genetics (Washington University), and the Washington
            University Genome Sequencing Center.  For additional information
            about the map position of this sequence, see
            http://www.nhgri.nih.gov/DIR/GTB/CHR7 or send
            mailto:egreen@nhgri.nih.gov

            SOURCE INFORMATION:
            Clone RG335O22 is from a release of the human BAC library
            CITB-HS-A.  The library contains cloned DNA from human sperm.  See:
             Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J.
            Kim et al., Genomics 34:213-8 (1996).  The clone is available from
            Research Genetics, Inc. (http://www.resgen.com).
            VECTOR:  pBeloBAC11
            Selection:  chloramphenicol


            NEIGHBORING SEQUENCE INFORMATION:
            The actual start of this clone is at base position 1 of RG335O22;
```



```
            the actual end is at base position 108878 of RG335O22.

            This clone contains STS sWSS3514 (NID:g1222801).
FEATURES             Location/Qualifiers
     source          1. .108878
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /clone="RG335O22"
                     /clone_lib="CITB-HS-A"
                     /map="7q21-q22"
     repeat_region   563. .794
                     /rpt_family="L1"
     repeat_region   1243. .1272
                     /rpt_family="L1"
     repeat_region   6077. .6102
                     /rpt_family="L1"
     repeat_region   13493. .13826
                     /rpt_family="L1"
     repeat_region   15133. .15424
                     /rpt_family="ALU"
     repeat_region   20066. .20485
                     /rpt_family="L1"
     repeat_region   25026. .25305
                     /rpt_family="ALU"
     misc_feature    29019. .29203
                     /note="match to EST D20877 (NID:g504697)"
     repeat_region   36139. .36166
                     /rpt_family="L1"
     repeat_region   complement(37310. .37421)
                     /rpt_family="ALU"
     repeat_region   37435. .37522
                     /rpt_family="ALU"
     repeat_region   37570. .37667
                     /rpt_family="ALU"
     repeat_region   complement(37670. .37708)
                     /rpt_family="ALU"
     repeat_region   complement(38413. .38687)
                     /rpt_family="ALU"
     repeat_region   complement(40439. .40727)
                     /rpt_family="ALU"
     repeat_region   42843. .42878
                     /rpt_family="L1"
     misc_feature    43587. .43791
                     /note="match to EST D20858 (NID:g504678)"
     repeat_region   complement(45232. .45303)
                     /rpt_family="PTR"
     repeat_region   complement(50142. .50216)
                     /rpt_family="PTR"
     repeat_region   51259. .51297
                     /rpt_family="L1"
```



```
     repeat_region   complement(60928. .61221)
                     /rpt_family="ALU"
     repeat_region   62463. .62800
                     /rpt_family="L1"
     repeat_region   complement(65195. .66897)
                     /rpt_family="PTR"
     repeat_region   65222. .66897
                     /rpt_family="PTR"
     repeat_region   71038. .71064
                     /rpt_family="L1"
     repeat_region   complement(75188. .75478)
                     /rpt_family="ALU"
     repeat_region   complement(81096. .81385)
                     /rpt_family="ALU"
     repeat_region   82157. .82212
                     /rpt_family="L1"
     repeat_region   92199. .92490
                     /rpt_family="ALU"
     repeat_region   95111. .95142
                     /rpt_family="L1"
     repeat_region   95520. .95820
                     /rpt_family="ALU"
     repeat_region   96048. .96393
                     /rpt_family="MER"
     repeat_region   96861. .96929
                     /rpt_family="ALU"
     repeat_region   96932. .96977
                     /rpt_family="L1"
     repeat_region   102590. .102621
                     /rpt_family="L1"
     repeat_region   complement(104717. .104998)
                     /rpt_family="ALU"
BASE COUNT    34576 a  20570 c  20710 g  33022 t
ORIGIN

  Query Match           19.5%;  Score 17;  DB 26;  Length 108878;
  Best Local Similarity 100.0%;  Pred. No. 1.92e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  24888 TTTTCCTTTCTAAAACC 24904
          |||||||||||||||||
Cp     37 TTTTCCTTTCTAAAACC 21


RESULT   46
LOCUS       AC005153   124191 bp    DNA             PRI       20-JUN-1998
DEFINITION  Homo sapiens PAC clone DJ0537P09 from 7p11.2-p12, complete
            sequence.
ACCESSION   AC005153
NID         g3242766
KEYWORDS    HTG.
```



```
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 124191)
  AUTHORS   Dauphin,S. and Biewald,T.
  TITLE     The sequence of Homo sapiens PAC clone DJ0537P09
  JOURNAL   Unpublished (1998)
REFERENCE   2  (bases 1 to 124191)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUN-1998) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     SUBMITTED BY:  WUGSC
            Genome Sequencing Center
            Department of Genetics
            Washington University
            St. Louis MO 63108, USA
            http://genome.wustl.edu/gsc
            mailto:sapiens@watson.wustl.edu

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded or sequenced with an alternate
            chemistry; an attempt was made to resolve all sequencing problems,
            such as compressions and repeats; all regions were covered by
            sequence from more than one subclone; and the assembly was
            confirmed by restriction digest.

            MAPPING INFORMATION:
            The sequence of this clone was established as part of a mapping and
            sequencing collaboration between the NHGRI Chromosome 7 Mapping
            Project (Eric D. Green, Director), John D. McPherson in the
            Department of Genetics (Washington University), and the Washington
            University Genome Sequencing Center.  For additional information
            about the map position of this sequence, see
            http://www.nhgri.nih.gov/DIR/GTB/CHR7 , send
            mailto:egreen@nhgri.nih.gov , or see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            This clone was derived from human PAC library RPCI-4, prepared by
            Pieter de Jong and coworkers at Roswell Park Cancer Institute,
            using the method described by Ioannou et al., Nature Genetics
            6:84-9 (1994).  The library is from one male donor.  For further
            details, see http://bacpac.med.buffalo.edu/
            The clone is available from Genome Systems, Inc.
            (http://www.genomesystems.com).
```

```
            VECTOR:  pCYPAC2
            NEIGHBORING SEQUENCE INFORMATION:
            The actual start of this clone is at base position 1 of DJ0537P09;
            actual end is at 124191 of DJ0537P09.
FEATURES              Location/Qualifiers
     source           1. .124191
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="7"
                      /map="7p11.2-p12"
                      /clone_lib="RPCI-4"
                      /clone="DJ0537P09"
     repeat_region    11. .173
                      /rpt_family="MER4-group"
     repeat_region    177. .795
                      /rpt_family="L1"
     repeat_region    1579. .1883
                      /rpt_family="Retroviral"
     repeat_region    2223. .2623
                      /rpt_family="MaLR"
     gene             3290. .>122871
                      /gene="GRB10"
     gene             3290. .>122871
                      /gene="GRB10"
     CDS              join(3290. .3340,31704. .31791,60954. .61176,65749. .65890,
                      108634. .108790,116327. .116442,117473. .117541,
                      119265. .119402,120732. .120842,122773. .>122871)
                      /gene="GRB10"
                      /note="match to protein AF0000 (PID:g2130617);
                      H_DJ0537P09.1a"
                      /codon_start=1
                      /product="insulin receptor binding protein GRB10"
                      /db_xref="PID:g3242768"
                      /translation="MALAGCPDSFLHHPYYQDKVEQTPRSQQDPAGPGLPAQSDRLAN
                      HQEDDVDLEALVNDMNASLESLYSACSMQSDTVPLLQNGQHARSQPRASGPPRSIQPQ
                      VSPRQRVQRSQPVHILAVRRLQEEDQQFRTSSLPAIPNPFPELCGPGSPPVLTPGSLP
                      PSQAAAKQDVKVFSEDGTSKVVEILADMTARDLCQLLVYKSHCVDDNSWTLVEHHPHL
                      GLERCLEDRELVVQVESTMASESKFLFRKNYAKYEFFKNPMNFFPEQMVTWCQQSNGS
                      QTQLLQNFLNSSSCPEIQGFLHVKELGKKSWKKLYVCLRRSGLYCSTKGTSKEPRHLQ
                      LLADLEDSNIFSLIAGRKQYNAPTDHGLCIKPNKVRNETKELRLLCAEDEQTRTCWMT
                      AFRLLK"
     CDS              join(3290. .3340,31704. .31791,60954. .61176,65749. .65890,
                      108634. .108790,116327. .116442,117473. .117541,
                      120732. .120842,122773. .>122871)
                      /gene="GRB10"
                      /note="match to protein I39175 (PID:g2136153);
                      H_DJ0537P09.1b"
                      /codon_start=1
                      /product="alternatively spliced variant of insulin
                      receptor binding protein GRB10"
                      /db_xref="PID:g3242767"
```



```
                      /translation="MALAGCPDSFLHHPYYQDKVEQTPRSQQDPAGPGLPAQSDRLAN
                      HQEDDVDLEALVNDMNASLESLYSACSMQSDTVPLLQNGQHARSQPRASGPPRSIQPQ
                      VSPRQRVQRSQPVHILAVRRLQEEDQQFRTSSLPAIPNPFPELCGPGSPPVLTPGSLP
                      PSQAAAKQDVKVFSEDGTSKVVEILADMTARDLCQLLVYKSHCVDDNSWTLVEHHPHL
                      GLERCLEDRELVVQVESTMASESKFLFRKNYAKYEFFKNPMNFFPEQMVTWCQQSNGS
                      QTQLLQEPRHLQLLADLEDSNIFSLIAGRKQYNAPTDHGLCIKPNKVRNETKELRLLC
                      AEDEQTRTCWMTAFRLLK"
     repeat_region    5671. .5848
                      /rpt_family="L2"
     repeat_region    6642. .6939
                      /rpt_family="Alu"
     repeat_region    6940. .7015
                      /rpt_family="MER1_type"
     repeat_region    8145. .8295
                      /rpt_family="L2"
     repeat_region    8483. .8552
                      /rpt_family="MIR"
     repeat_region    8556. .8655
                      /rpt_family="MIR"
     repeat_region    9195. .9500
                      /rpt_family="Alu"
     repeat_region    9575. .9696
                      /rpt_family="MIR"
     repeat_region    10143. .10517
                      /rpt_family="MaLR"
     repeat_region    14204. .14468
                      /rpt_family="MER2_type"
     repeat_region    15921. .16228
                      /rpt_family="L1"
     repeat_region    16381. .16626
                      /rpt_family="L2"
     repeat_region    18709. .18781
                      /rpt_family="L1"
     repeat_region    18893. .19442
                      /rpt_family="L1"
     repeat_region    19581. .19797
                      /rpt_family="L1"
     repeat_region    19979. .20033
                      /rpt_family="MIR"
     repeat_region    20932. .21230
                      /rpt_family="Alu"
     repeat_region    22451. .22490
                      /rpt_family="MIR"
     repeat_region    22642. .22787
                      /rpt_family="L1"
     repeat_region    23587. .23789
                      /rpt_family="MIR"
     repeat_region    25713. .25819
                      /rpt_family="MER1_type"
     misc_feature     25820. .26073
                      /gene="GRB10"
```



```
                      /note="match to EST AA129408 (NID:g1689294) zn83c02.r1"
     misc_feature     26129. .26219
                      /gene="GRB10"
                      /note="match to EST AA129408 (NID:g1689294) zn83c02.r1"
     repeat_region    26640. .26937
                      /rpt_family="Alu"
     repeat_region    27447. .27623
                      /rpt_family="MER53"
     repeat_region    28767. .28844
                      /rpt_family="MIR"
     repeat_region    28921. .29309
                      /rpt_family="L1"
     repeat_region    34010. .34397
                      /rpt_family="Alu"
     repeat_region    34861. .34997
                      /rpt_family="L1"
     repeat_region    35458. .35758
                      /rpt_family="Alu"
     repeat_region    35872. .37519
                      /rpt_family="L1"
     repeat_region    38300. .38850
                      /rpt_family="L1"
     repeat_region    39174. .39265
                      /rpt_family="MIR"
     repeat_region    39607. .39664
                      /rpt_family="L1"
     repeat_region    41280. .41542
                      /rpt_family="Alu"
     repeat_region    42628. .42737
                      /rpt_family="MIR"
     repeat_region    43326. .43618
                      /rpt_family="MER2_type"
     repeat_region    44122. .44319
                      /rpt_family="MIR"
     repeat_region    45297. .45597
                      /rpt_family="Alu"
     repeat_region    47428. .47569
                      /rpt_family="MIR"
     repeat_region    48557. .48601
                      /rpt_family="L2"
     repeat_region    48807. .48949
...
Note: remainder of annotations omitted.

  Query Match           19.5%;  Score 17;  DB 26;  Length 124191;
  Best Local Similarity 100.0%;  Pred. No. 1.92e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  96486 TTTCTAAAACCCTCACC 96502
          |||||||||||||||||
Cp     31 TTTCTAAAACCCTCACC 15
```



```
RESULT   47
LOCUS       AC004547   130632 bp    DNA             PRI       10-APR-1998
DEFINITION  Homo sapiens PAC clone DJ0982C16 from 7q11.23-q21.1, complete
            sequence.
ACCESSION   AC004547
NID         g3046309
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 130632)
  AUTHORS   Rohlfing,T., Klinke,B. and Ames,H.
  TITLE     The sequence of Homo sapiens PAC clone DJ0982C16
  JOURNAL   Unpublished (1998)
REFERENCE   2  (bases 1 to 130632)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-APR-1998) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     SUBMITTED BY:  WUGSC
            Genome Sequencing Center
            Department of Genetics
            Washington University
            St. Louis MO 63108, USA
            http://genome.wustl.edu/gsc
            mailto:sapiens@watson.wustl.edu

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded or sequenced with an alternate
            chemistry; an attempt was made to resolve all sequencing problems,
            such as compressions and repeats; all regions were covered by
            sequence from more than one subclone; and the assembly was
            confirmed by restriction digest.

            MAPPING INFORMATION:
            The sequence of this clone was established as part of a mapping and
            sequencing collaboration between the NHGRI Chromosome 7 Mapping
            Project (Eric D. Green, Director), John D. McPherson in the
            Department of Genetics (Washington University), and the Washington
            University Genome Sequencing Center.  For additional information
            about the map position of this sequence, see
            http://www.nhgri.nih.gov/DIR/GTB/CHR7 , send
            mailto:egreen@nhgri.nih.gov , or see http://genome.wustl.edu/gsc
```

```
            SOURCE INFORMATION:
            This clone was derived from human PAC library RPCI-5, prepared by
            Pieter de Jong and coworkers at Roswell Park Cancer Institute,
            using the method described by Ioannou et al., Nature Genetics
            6:84-9 (1994).  The library is from one male donor.  For further
            details, see http://bacpac.med.buffalo.edu/
            The clone is available from Genome Systems, Inc.
            (http://www.genomesystems.com).
            VECTOR:  pCYPAC2
            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is DJ0731H19.  The actual start of
            this clone is at base position 1 of DJ0982C16; actual end is at
            130632 of DJ0982C16.

            This clone contains STS sWSS2130 (NID:g1113390).
FEATURES            Location/Qualifiers
     source         1. .130632
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="7"
                    /map="7q11.23-q21.1"
                    /clone_lib="RPCI-5"
                    /clone="DJ0982C16"
     repeat_region  865. .1347
                    /rpt_family="L1"
     repeat_region  2510. .2845
                    /rpt_family="L2"
     repeat_region  3106. .3507
                    /rpt_family="L1"
     repeat_region  4552. .4726
                    /rpt_family="MIR"
     repeat_region  5432. .5608
                    /rpt_family="MIR"
     repeat_region  6205. .6305
                    /rpt_family="MIR"
     repeat_region  6371. .6669
                    /rpt_family="Alu"
     repeat_region  7353. .7662
                    /rpt_family="Alu"
     repeat_region  8821. .9024
                    /rpt_family="MIR"
     repeat_region  9670. .9759
                    /rpt_family="MER2_type"
     repeat_region  10018. .10379
                    /rpt_family="MaLR"
     repeat_region  10543. .10825
                    /rpt_family="Alu"
     repeat_region  10957. .11047
                    /rpt_family="MER2_type"
     repeat_region  13882. .14020
```



```
                    /rpt_family="MIR"
     repeat_region  14222. .14569
                    /rpt_family="L2"
     repeat_region  15328. .16201
                    /rpt_family="L1"
     repeat_region  16204. .16292
                    /rpt_family="Alu"
     repeat_region  16294. .16442
                    /rpt_family="L1"
     repeat_region  16443. .16501
                    /rpt_family="MER1_type?"
     repeat_region  16530. .16896
                    /rpt_family="MaLR"
     repeat_region  17554. .17906
                    /rpt_family="MaLR"
     repeat_region  19933. .19994
                    /rpt_family="MIR"
     repeat_region  20157. .20213
                    /rpt_family="MIR"
     repeat_region  20390. .20626
                    /rpt_family="L2"
     repeat_region  20762. .20906
                    /rpt_family="L2"
     repeat_region  21577. .21684
                    /rpt_family="MIR"
     repeat_region  24696. .24777
                    /rpt_family="L2"
     repeat_region  28593. .28666
                    /rpt_family="Unknown"
     repeat_region  28944. .29040
                    /rpt_family="MIR"
     repeat_region  29050. .29236
                    /rpt_family="MIR"
     repeat_region  29754. .30275
                    /rpt_family="MER1_type"
     repeat_region  31689. .31921
                    /rpt_family="MIR"
     repeat_region  34794. .34853
                    /rpt_family="MER1_type"
     repeat_region  35412. .35542
                    /rpt_family="LOR1"
     repeat_region  36942. .37236
                    /rpt_family="Alu"
     repeat_region  39222. .39491
                    /rpt_family="L2"
     repeat_region  39989. .40289
                    /rpt_family="Alu"
     repeat_region  40955. .41080
                    /rpt_family="L2"
     repeat_region  43998. .44105
                    /rpt_family="L1"
```



```
     repeat_region  46263. .46411
                    /rpt_family="MaLR"
     repeat_region  47962. .49333
                    /rpt_family="MER1_type"
     repeat_region  51302. .51604
                    /rpt_family="Alu"
     repeat_region  52802. .52853
                    /rpt_family="(CA)n"
     repeat_region  53101. .53211
                    /rpt_family="MIR"
     repeat_region  54115. .54604
                    /rpt_family="MaLR"
     repeat_region  54972. .55103
                    /rpt_family="L2"
     repeat_region  56378. .56559
                    /rpt_family="MIR"
     repeat_region  56584. .56744
                    /rpt_family="MIR"
     repeat_region  60723. .60903
                    /rpt_family="L2"
     repeat_region  61071. .61234
                    /rpt_family="L2"
     repeat_region  61320. .61915
                    /rpt_family="L1"
     repeat_region  63072. .63393
                    /rpt_family="L1"
     repeat_region  63746. .64039
                    /rpt_family="Alu"
     repeat_region  64217. .65036
                    /rpt_family="L1"
     repeat_region  65073. .65142
                    /rpt_family="HY3"
     repeat_region  65147. .65535
                    /rpt_family="L1"
     repeat_region  65540. .65814
                    /rpt_family="Alu"
     repeat_region  68351. .69414
                    /rpt_family="L2"
     repeat_region  69562. .69780
                    /rpt_family="L2"
     repeat_region  71272. .71559
                    /rpt_family="Alu"
     repeat_region  71785. .71987
                    /rpt_family="Alu"
     repeat_region  72371. .72556
                    /rpt_family="MIR"
     repeat_region  72892. .73177
                    /rpt_family="Alu"
     repeat_region  73184. .73271
                    /rpt_family="MER1_type"
```



```
Note: remainder of annotations omitted.

  Query Match          19.5%;  Score 17;  DB 26;  Length 130632;
  Best Local Similarity 100.0%;  Pred. No. 1.92e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  44007 AAATATAGTAAAATGTG 44023
          |||||||||||||||||
Qy     36 AAATATAGTAAAATGTG 52


RESULT   48
LOCUS       HS82J11    132981 bp    DNA             PRI       13-JAN-1998
DEFINITION  Human DNA sequence from PAC 82J11 and cosmid U134E6 on chromosome
            Xq22. Contains NIK like and Thyroxin-binding globulin precursor
            (T4-binding globulin, TBG) genes, ESTs and STSs.
ACCESSION   Z83850
NID         g2780172
KEYWORDS    globulin; NIK; T4; TBG; Thyroxin; Xq22.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 132981)
  AUTHORS   Ho,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JAN-1998) sanger.ac.uk/HGP/ChrX/) Sanger Centre,
            Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquires:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jan 16, 1998 this sequence version replaced gi:2467172.
            IMPORTANT: This sequence is the entire insert of clone 82J11.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variations annotated may not be found in the sequence submission
            corresponding to the overlapping clone as we submit sequences with
            only a small overlap as described above.
            This sequence was generated from part of bacterial clone contigs of
            human chromosome X, constructed by the Sanger Centre chromosome X
            mapping group. Further information can be found at
            http://www.sanger.ac.uk/HGP/ChrX/
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            The true left end of clone 82J11 is at 1 in this sequence. The true
            left end of clone U134E6 is at 62897.
            The true right end of clone U134E6 is at 102829.
            The true right end of clone 82J11 is at 132981.
```

82J11 is from the library RPCI1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see http://bacpac.med.buffalo.edu/ U134E6 is from the Lawrence Livermore National Laboratory flow-sorted X chromosome cosmid library LLOXNC01.

FEATURES Location/Qualifiers

source 1. .132981
/organism="Homo sapiens"
/db_xref="taxon:9606"

source 1. .92644
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="82J11"
/clone_lib="RPCI1"

repeat_region 1150. .1340
/note="LTR8 repeat: matches 561. .383 of consensus"

repeat_region 9510. .9531
/note="11 copies of 2 mer 100 % conserved"

repeat_region 9923. .10133
/note="MER3 repeat: matches 208. .1 of consensus"

repeat_region 10261. .11135
/note="L1 repeat: matches 5281. .4409 of consensus"

repeat_region 11133. .11992
/note="L1PA2 repeat: matches 34. .893 of consensus"

repeat_region 14392. .14433
/note="21 copies of 2 mer 93 % conserved"

repeat_region 15824. .16707
/note="L1MA4 repeat: matches 1039. .66 of consensus"

repeat_region 16790. .16855
/note="L1PA5 repeat: matches 890. .825 of consensus"

repeat_region 16857. .20874
/note="L1 repeat: matches 5299. .1333 of consensus"

repeat_region 20593. .21155
/note="MER25 repeat: matches 2132. .1556 of consensus"

repeat_region 22337. .22908
/note="L1MB4 repeat: matches 931. .333 of consensus"

repeat_region 24279. .24518
/note="MIR repeat: matches 4. .254 of consensus"

repeat_region 25316. .25482
/note="AluJb repeat: matches 302. .134 of consensus; incomplete repeat"

mRNA join(<25698. .25832,27946. .28089,30381. .30521,34448. .34548, 36822. .36971,40106. .40241,43541. .43680,46012. .>46099)
/gene="dJ82J11.1"
/note="mouse NIK serine threonine protein kinase like; match ESTs R98571 AA333635"
/evidence=not_experimental

gene 25698. .46099
/gene="dJ82J11.1"

CDS join(25698. .25832,27946. .28089,30381. .30521,34448. .34548, 36822. .36971,40106. .40241,43541. .43680)

/partial
/gene="dJ82J11.1"
/note="mouse NIK serine threonine protein kinase like; match: proteins P97820 CE02384"
/codon_start=3
/evidence=not_experimental
/db_xref="PID:e1231427"
/db_xref="PID:g2780173"
/translation="PEESPKQPSEVNVNPLYVSPACKKPLIHMYEKEFTSEICCGSLW GVNLLLGTRSNLYLMDRSGKADITKLIRRRPFRQIQVLEPLNLLITISGHKNRLRVYH LTWLRNKILNNDPESKRRQEEMLKTEEACKAIDKLTGFQHEETTYIAIALKSSIHLYA WAPKSFDESTAIKVPPTLDHKPVTVDLAIGSEKRLKIFPSSADGYHLIDAESEVMSDV TLPKNNIIILPDCLGIGMMLTFNAEALSVEANEQLFKKILEMWKDIPSSIAFECTQRT TGWGQKAIEVRSLQSRVLESELKRRSIKKLRPLCTRGDK"

repeat_region 26070. .26296
/note="AluSg repeat: matches 51. .279 of consensus; incomplete repeat"

repeat_region 29261. .29598
/note="THE1C repeat: matches 371. .36 of consensus"

repeat_region 30655. .30695
/note="5S repeat: matches 41. .1 of consensus"

repeat_region 32254. .33127
/note="L1MB8 repeat: matches 920. .1 of consensus"

repeat_region 33130. .33431
/note="AluSx repeat: matches 296. .1 of consensus"

repeat_region 33435. .34221
/note="L1 repeat: matches 5248. .4454 of consensus"

repeat_region 34694. .34898
/note="AluJo repeat: matches 301. .100 of consensus; incomplete repeat"

repeat_region 35592. .35734
/note="MER5A repeat: matches 189. .43 of consensus"

repeat_region 37678. .37969
/note="AluYa5 repeat: matches 10. .301 of consensus"

repeat_region 41035. .41084
/note="25 copies of 2 mer 100 % conserved"

repeat_region 41321. .41354
/note="17 copies of 2 mer 91 % conserved"

repeat_region 45370. .45533
/note="L1MA8 repeat: matches 862. .1031 of consensus"

prim_transcript 46370. .49114
/note="match: ESTs N30117 W60729 N33939 R98572 R28658 R70758 R62577 H47233 H47234 R62576 H06699 R70808 R28510 H06749 AA331981 AA335043 AA249339 AA247203 H03220 AA331981"

misc_feature complement(47194. .47616)
/note="match STSs G28348 G26027"

misc_feature complement(49515. .49722)
/note="match STS AF020167"

unsure 50129. .50131

repeat_region 50194. .50574

/note="L1 repeat: matches 7. .388 of consensus"

repeat_region 50573. .55498
/note="L1 repeat: matches 479. .5390 of consensus"

repeat_region 55353. .56235
/note="L1PA2 repeat: matches 1. .891 of consensus"

repeat_region 57119. .57154
/note="18 copies of 2 mer 81 % conserved"

repeat_region 58395. .58799
/note="MLT1A1 repeat: matches 362. .4 of consensus"

repeat_region 59693. .59994
/note="AluSx repeat: matches 302. .1 of consensus"

unsure 59837. .59917

repeat_region 60007. .60053
/note="MSTD repeat: matches 48. .1 of consensus"

repeat_region 60065. .61105
/note="MST-INTERNAL repeat: matches 1621. .628 of consensus"

repeat_region 61117. .61960
/note="L1 repeat: matches 4530. .5388 of consensus"

repeat_region 61829. .62702
/note="L1PA7 repeat: matches 1. .887 of consensus"

repeat_region 62721. .62912
/note="MST-INTERNAL repeat: matches 620. .433 of consensus"

variation 63677. .63679
/note="clone U134E6; GGA in this entry; insertion"
/replace="ga"

repeat_region 64668. .64889
/note="L1ME3A repeat: matches 349. .565 of consensus"

repeat_region 65224. .65456
/note="L1ME3A repeat: matches 910. .691 of consensus"

repeat_region 65303. .65456
/note="L1ME3 repeat: matches 839. .691 of consensus"

repeat_region 66620. .66721
/note="MER20 repeat: matches 23. .125 of consensus"

repeat_region 66973. .67866
/note="L1PA15 repeat: matches 897. .1 of consensus"

repeat_region 67723. .72165
/note="L1 repeat: matches 5390. .987 of consensus"

repeat_region 73816. .74035
/note="MIR repeat: matches 249. .18 of consensus"

repeat_region 74040. .74096
/note="L1MC1 repeat: matches 1079. .1022 of consensus"

repeat_region 74155. .74188
/note="17 copies of 2 mer 82 % conserved"

repeat_region 74288. .74681
/note="MLT1A1 repeat: matches 362. .2 of consensus"

repeat_region 76306. .76443
/note="MIR2 repeat: matches 138. .1 of consensus"

repeat_region 77147. .77434
/note="AluJo repeat: matches 16. .302 of consensus; incomplete repeat"

repeat_region 78277. .78947
/note="L1 repeat: matches 3606. .2975 of consensus"

repeat_region 79099. .79390
/note="AluSx repeat: matches 302. .1 of consensus"

repeat_region 79413. .79776
/note="L1 repeat: matches 2512. .2123 of consensus"

repeat_region 79882. .80048
/note="AluJo repeat: matches 302. .132 of consensus;

...

Note: remainder of annotations omitted.

Query Match 19.5%; Score 17; DB 26; Length 132981;
Best Local Similarity 100.0%; Pred. No. 1.92e+01;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 39068 AGATTTTATTAAGACAA 39084
||||||||||||||||
Cp 73 AGATTTTATTAAGACAA 57

RESULT 49
LOCUS AC004866 136503 bp DNA HTG 12-JUN-1998
DEFINITION *** SEQUENCING IN PROGRESS *** Homo sapiens clone DJ0728H09; HTGS phase 1, 2 unordered pieces.
ACCESSION AC004866
NID g3213131
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 136503)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 136503)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (12-JUN-1998) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
COMMENT ***
*** WARNING: Phase 1 High Throughput Genome Sequence ***
***
* This sequence is unfinished. It consists of 2 contigs for
* which the order is not known; their order in this record is
* arbitrary. In some cases, the exact lengths of the gaps
* between the contigs are also unknown; these gaps are presented
* as runs of N as a convenience only. When sequencing is complete,
* the sequence data presented in this record will be replaced
*by a single finished sequence with the same accession number.

```
            *        1    38905: contig of 38905 bp in length
            *    38906    38923: gap of unknown length
            *    38924   136503: contig of 97580 bp in length.
FEATURES             Location/Qualifiers
     source          1. .136503
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="DJ0728H09"
BASE COUNT    43283 a   24518 c   24256 g   44428 t     18 others
ORIGIN

  Query Match           19.5%;  Score 17;  DB 17;  Length 136503;
  Best Local Similarity 100.0%;  Pred. No. 1.92e+01;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  55476 TAGTAAAATGTGTAATT 55492
          |||||||||||||||||.
Qy     41 TAGTAAAATGTGTAATT 57


RESULT   50
LOCUS       AC004033   145857 bp    DNA              HTG       01-OCT-1998
DEFINITION  *** SEQUENCING IN PROGRESS *** Homo sapiens Chromosome 22q11.2 PAC
            Clone p_m11 In BCRL2-GGT Region; HTGS phase 1, 4 unordered pieces.
ACCESSION   AC004033
NID         g3687253
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 145857)
  AUTHORS   Zhang,G., Lao,V., Zhan,M. and Roe,B.A.
  TITLE     Homo sapiens Chromosome 22q11.2 BAC Clone p_m11 In BCRL2-GGT Region
  JOURNAL   Unpublished (1998)
REFERENCE   2  (bases 1 to 145857)
  AUTHORS   Budarf,M.L. and Emanuel,B.S.
  JOURNAL   Unpublished (1998)
REFERENCE   3  (bases 1 to 145857)
  AUTHORS   Zhang,G., Lao,V., Zhan,M., Budarf,M.L., Emanuel,B.S. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JAN-1998) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On Oct 1, 1998 this sequence version replaced gi:3643612.
            ***                                                        ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                        ***
            * This sequence is unfinished. It consists of 4 contigs for
            * which the order is not known; their order in this record is
            * arbitrary. In some cases, the exact lengths of the gaps
```



```
            * between the contigs are also unknown; these gaps are  presented
            * as runs of N as a convenience only. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
            *        1     5868: contig of 5868 bp in length
            *     5869     5948: gap of unknown length
            *     5949    16950: contig of 11002 bp in length
            *    16951    17030: gap of unknown length
            *    17031    35357: contig of 18327 bp in length
            *    35358    35437: gap of unknown length
            *    35438   145857: contig of 110420 bp in length.
FEATURES             Location/Qualifiers
     source          1. .145857
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="22"
                     /map="22q11.2"
                     /clone="p_m11"
BASE COUNT    37157 a   37695 c   36323 g   34442 t    240 others
ORIGIN

  Query Match           19.5%;  Score 17;  DB 17;  Length 145857;
  Best Local Similarity 100.0%;  Pred. No. 1.92e+01;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2552 ATATTTTTCCTTTCTAA 2568
          |||||||||||||||||
Cp     41 ATATTTTTCCTTTCTAA 25


Search completed: Tue Dec 22 22:04:25 1998
Job time : 421 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Tue Dec 22 21:14:51 1998; MasPar time 130.33 Seconds
851.916 Million cell updates/sec

Tabular output not generated.

Title: >US-08-530-112A-1449
Description: (1-62) from US08530112A.seq
Perfect Score: 61
N.A. Sequence: 1 GATCCCCATTCTTTTCAAAA.....GCAAGGAAGACTGANGGAAA 62
Comp: CTAGGGGTAAGAAAAGTTTT.....CGTTCCTTCTGACTNCCTTT

Scoring table: TABLE jmetric
Gap 60

Rmatch STD : Dbase 0; Query 0

Searched: 2275026 seqs, 895388244 bases x 2

Post-processing: Minimum Match 0%
Listing first 1000 summaries

Database: embl-est55
1:em_est1 2:em_gss1 3:em_gss2 4:em_gss3
Database: genbank-est107
5:gb_est1 6:gb_est10 7:gb_est11 8:gb_est12 9:gb_est13
10:gb_est14 11:gb_est15 12:gb_est16 13:gb_est17
14:gb_est18 15:gb_est19 16:gb_est2 17:gb_est20
18:gb_est21 19:gb_est3 20:gb_est4 21:gb_est5 22:gb_est6
23:gb_est7 24:gb_est8 25:gb_est9 26:gb_gss1 27:gb_gss2
28:gb_gss3 29:gb_gss4

Statistics: Mean 7.432; Variance 1.618; scale 4.595

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
	1	61	100.0	62	7	HUMGS01603	Human HL60 3'directed	9.07e-72
	2	31	50.8	58	7	HUMGS01613	Human HL60 3'directed	1.37e-21
	3	22	36.1	407	13	AA844688	ai70b09.s1 Soares test	1.64e-08
c	4	20	32.8	390	14	AA848939	EST191701 Normalized r	7.15e-06
c	5	20	32.8	406	28	AQ220224	HS_3248_B1_A07_MR CIT	7.15e-06
c	6	20	32.8	415	19	N32666	yx77c12.r1 Homo sapien	7.15e-06
c	7	20	32.8	419	18	AI170294	EST216220 Normalized r	7.15e-06
c	8	20	32.8	503	15	AA926956	om68h05.s1 NCI_CGAP_GC	7.15e-06
c	9	20	32.8	642	13	AA747933	nx97h10.s1 NCI_CGAP_GC	7.15e-06
c	10	20	32.8	716	26	AG002485	Homo sapiens genomic D	7.15e-06
	11	19	31.1	136	15	AA904400	ok07b09.s1 Soares_NFL_	1.33e-04
	12	19	31.1	229	20	HSC05B102	H. sapiens partial cDN	1.33e-04
	13	19	31.1	259	5	R38922	yd06a09.s1 Homo sapien	1.33e-04
	14	19	31.1	294	14	AA899734	UI-R-E0-dg-d-09-0-UI.s	1.33e-04
	15	19	31.1	316	15	AA878981	oj26e09.s1 NCI_CGAP_Ki	1.33e-04
c	16	19	31.1	328	8	AA465962	ve90b04.s1 Knowles Sol	1.33e-04
	17	19	31.1	330	19	N22475	yw34g10.s1 Homo sapien	1.33e-04
	18	19	31.1	332	28	AQ094353	HS_3025_B2_G04_MR CIT	1.33e-04
	19	19	31.1	341	17	AU021150	Mus musculus 8-cell em	1.33e-04
	20	19	31.1	342	17	AU021148	Mus musculus 8-cell em	1.33e-04
c	21	19	31.1	353	10	AA601083	no53e04.s1 NCI_CGAP_SS	1.33e-04
	22	19	31.1	373	17	AU020149	Mus musculus 8-cell em	1.33e-04
	23	19	31.1	376	21	W42451	zc22e06.s1 Soares sene	1.33e-04
	24	19	31.1	379	13	AA859452	UI-R-A0-bf-b-05-0-UI.s	1.33e-04
	25	19	31.1	382	8	R76563	yi72h02.s1 Soares plac	1.33e-04
c	26	19	31.1	384	20	W03777	za65b10.r1 Soares feta	1.33e-04
	27	19	31.1	400	13	AA874518	vx03c09.r1 Soares 2NbM	1.33e-04
	28	19	31.1	409	17	AU019558	Mus musculus 8-cell em	1.33e-04
	29	19	31.1	413	10	AA614837	nn90c12.s1 NCI_CGAP_Br	1.33e-04
c	30	19	31.1	415	5	R07069	yf13c05.r1 Homo sapien	1.33e-04
	31	19	31.1	419	28	AQ113106	CIT-HSP-2379L5.TR CIT-	1.33e-04
	32	19	31.1	422	26	B38229	HS-1047-A2-A03-MR.abi	1.33e-04
	33	19	31.1	424	19	N23446	yv99d01.s1 Homo sapien	1.33e-04
	34	19	31.1	429	16	H27546	yl61d01.s1 Homo sapien	1.33e-04
	35	19	31.1	446	19	N27212	yw69g10.s1 Homo sapien	1.33e-04
	36	19	31.1	449	19	N32190	yy25e11.s1 Homo sapien	1.33e-04
	37	19	31.1	450	19	N27215	yw69h10.s1 Homo sapien	1.33e-04
	38	19	31.1	472	18	AI151182	qc73e06.x1 Soares_plac	1.33e-04
	39	19	31.1	473	13	AA875122	UI-R-E0-bu-c-01-0-UI.s	1.33e-04
	40	19	31.1	487	17	AU015009	Mus musculus 2-cell em	1.33e-04
	41	19	31.1	509	19	N32188	yy25e09.s1 Homo sapien	1.33e-04
	42	19	31.1	544	17	AU021048	Mus musculus 8-cell em	1.33e-04

c	43	19	31.1	580	28	AQ160686	mgxb0006L08r CUGI Rice	1.33e-04
	44	19	31.1	582	17	AU018946	Mus musculus 8-cell em	1.33e-04
	45	19	31.1	720	27	AQ001905	CIT-HSP-2290L15.TF CIT	1.33e-04
	46	18	29.5	122	15	AA946594	oq38b08.s1 NCI_CGAP_Ki	2.26e-03
	47	18	29.5	129	11	AA702751	zi82f11.s1 Soares feta	2.26e-03
	48	18	29.5	145	21	AA059119	zf64g04.s1 Soares reti	2.26e-03
c	49	18	29.5	180	17	AI066361	TENU2477 T. cruzi epim	2.26e-03
	50	18	29.5	188	19	H98966	yx11b08.s1 Homo sapien	2.26e-03
c	51	18	29.5	192	22	AA191084	zq43e01.r1 Stratagene	2.26e-03
	52	18	29.5	193	14	AA808122	oc39e04.s1 NCI_CGAP_GC	2.26e-03
	53	18	29.5	194	6	AA034946	zk25e03.s1 Soares preg	2.26e-03
	54	18	29.5	215	23	AA250734	zt32g12.s1 Soares ovar	2.26e-03
	55	18	29.5	222	15	AI013420	EST208095 Normalized r	2.26e-03
	56	18	29.5	238	13	AA877400	ob33d06.s1 NCI_CGAP_Ki	2.26e-03
	57	18	29.5	239	19	N75988	za43a07.s1 Homo sapien	2.26e-03
	58	18	29.5	259	16	H39064	yp65b10.s1 Homo sapien	2.26e-03
	59	18	29.5	273	15	AA970907	oo82c03.s1 NCI_CGAP_Ki	2.26e-03
	60	18	29.5	291	15	AI009414	EST203865 Normalized r	2.26e-03
	61	18	29.5	318	9	AA576261	nm62c09.s1 NCI_CGAP_Br	2.26e-03
	62	18	29.5	331	18	AI078731	oy12f12.s1 Soares_sene	2.26e-03
	63	18	29.5	332	15	AA995622	os28h02.s1 NCI_CGAP_Ki	2.26e-03
	64	18	29.5	333	15	AI013223	EST207898 Normalized r	2.26e-03
	65	18	29.5	337	14	AA899245	UI-R-E0-cs-f-09-0-UI.s	2.26e-03
	66	18	29.5	340	21	AA051594	mj35f08.r1 Soares mous	2.26e-03
	67	18	29.5	347	9	AA579568	nm71e01.s1 NCI_CGAP_Co	2.26e-03
c	68	18	29.5	352	11	AA671405	vn81e01.r1 Knowles Sol	2.26e-03
	69	18	29.5	353	15	AI011398	EST205849 Normalized r	2.26e-03
	70	18	29.5	358	16	R77513	yi79h03.s1 Homo sapien	2.26e-03
	71	18	29.5	359	14	AA901199	UI-R-A1-dp-e-02-0-UI.s	2.26e-03
	72	18	29.5	362	14	AA910650	ok65b09.s1 NCI_CGAP_GC	2.26e-03
	73	18	29.5	364	18	AI042594	oy14g05.x1 Soares_sene	2.26e-03
	74	18	29.5	377	15	AI012125	EST206576 Normalized r	2.26e-03
	75	18	29.5	378	11	AA707033	zj32c02.s1 Soares feta	2.26e-03
	76	18	29.5	379	28	AQ131502	HS_3043_A1_B02_MR CIT	2.26e-03
	77	18	29.5	381	23	AA137359	mq77c02.r1 Stratagene	2.26e-03
	78	18	29.5	387	20	N89742	zb14b06.s1 Soares feta	2.26e-03
	79	18	29.5	388	26	AQ023175	HS_2177_B1_D02_MF CIT	2.26e-03
	80	18	29.5	390	5	R36643	yh89c06.s1 Homo sapien	2.26e-03
	81	18	29.5	393	7	AA454158	zx46f09.s1 Soares test	2.26e-03
	82	18	29.5	398	10	AA652996	ns62h07.s1 NCI_CGAP_Pr	2.26e-03
	83	18	29.5	401	16	H29754	yp02f11.s1 Homo sapien	2.26e-03
	84	18	29.5	404	12	AA775632	zf26b05.s1 Soares feta	2.26e-03
	85	18	29.5	413	13	AA844004	ai91d05.s1 Soares para	2.26e-03
	86	18	29.5	414	13	AA861696	ak30f07.s1 Soares test	2.26e-03
	87	18	29.5	420	14	AA901332	UI-R-A1-dv-a-07-0-UI.s	2.26e-03
c	88	18	29.5	421	21	W56706	zd14e04.r1 Soares feta	2.26e-03
	89	18	29.5	426	20	W52001	zc37h04.s1 Soares sene	2.26e-03
	90	18	29.5	430	19	N34333	yy12a02.s1 Homo sapien	2.26e-03
	91	18	29.5	433	26	AQ018802	CIT-HSP-2302G14.TR CIT	2.26e-03
	92	18	29.5	439	10	AA430524	zw21b08.s1 Soares ovar	2.26e-03
	93	18	29.5	440	10	AA647218	vn39g12.r1 Stratagene	2.26e-03

	94	18	29.5	445	13	AA160328	zo56d03.s1 Stratagene	2.26e-03
	95	18	29.5	448	15	AA943390	EST198889 Normalized r	2.26e-03
	96	18	29.5	448	6	AA011693	zi03h01.s1 Soares feta	2.26e-03
	97	18	29.5	451	22	H86596	yw07a09.s1 Soares mela	2.26e-03
	98	18	29.5	451	18	AI052491	oz19g06.x1 Soares_feta	2.26e-03
	99	18	29.5	455	14	AA893971	EST197774 Normalized r	2.26e-03
	100	18	29.5	455	10	AA477759	zu34c11.s1 Soares ovar	2.26e-03
	101	18	29.5	456	19	N20836	yx49b03.s1 Homo sapien	2.26e-03
	102	18	29.5	458	15	AA944321	EST199820 Normalized r	2.26e-03
	103	18	29.5	464	11	AA677260	zj61a03.s1 Soares feta	2.26e-03
c	104	18	29.5	468	26	AQ018515	CIT-HSP-2311C20.TR CIT	2.26e-03
	105	18	29.5	470	27	AQ037393	CIT-HSP-2335N4.TF CIT-	2.26e-03
	106	18	29.5	470	17	AI034439	ow24b01.x1 Soares_para	2.26e-03
c	107	18	29.5	472	26	B33715	HS-1029-A1-H02-MF.abi	2.26e-03
	108	18	29.5	474	17	AI088462	qb07f09.x1 Soares_preg	2.26e-03
	109	18	29.5	478	15	AA943299	EST198798 Normalized r	2.26e-03
c	110	18	29.5	482	29	AQ235741	HS_2015_B2_F05_T7 CIT	2.26e-03
	111	18	29.5	484	16	R70016	yi48c02.s1 Homo sapien	2.26e-03
	112	18	29.5	485	17	AI087842	oo24d05.x1 Soares_NSF_	2.26e-03
	113	18	29.5	487	15	AA942703	EST198202 Normalized r	2.26e-03
	114	18	29.5	491	16	R92002	yp96b04.s1 Homo sapien	2.26e-03
	115	18	29.5	493	12	AA781354	aj25c02.s1 Soares test	2.26e-03
c	116	18	29.5	500	28	AQ175902	HS_3213_B2_H05_MR CIT	2.26e-03
	117	18	29.5	504	22	AA189403	mt87b05.r1 Soares mous	2.26e-03
	118	18	29.5	519	9	AA535808	nf93g07.s1 NCI_CGAP_Co	2.26e-03
c	119	18	29.5	605	15	AA990917	BSBmMFSZ02B20SK Brugia	2.26e-03
	120	18	29.5	611	18	AI171273	EST217228 Normalized r	2.26e-03
c	121	18	29.5	621	27	AQ075404	CIT-HSP-2351O10.TF CIT	2.26e-03
	122	18	29.5	627	8	AA513733	nh88d01.s1 NCI_CGAP_Br	2.26e-03
	123	18	29.5	678	13	AA628885	af30b07.s1 Soares tota	2.26e-03
c	124	18	29.5	723	26	B20533	F7A4-Sp6 IGF Arabidops	2.26e-03
	125	18	29.5	724	26	AG005695	Homo sapiens genomic D	2.26e-03
	126	18	29.5	724	28	AG013372	Homo sapiens genomic D	2.26e-03
c	127	17	27.9	80	22	N98189	0218C3 crapPFDd2.1, De	3.50e-02
	128	17	27.9	164	23	AA064471	ml49d05.r1 Stratagene	3.50e-02
c	129	17	27.9	166	25	AA369218	EST80595 Placenta II H	3.50e-02
	130	17	27.9	205	5	T01815	wEST02536 Caenorhabdit	3.50e-02
c	131	17	27.9	224	26	HSU82602	Human retinoschisis re	3.50e-02
c	132	17	27.9	261	20	W92581	zd99c11.s1 Soares feta	3.50e-02
	133	17	27.9	263	21	AA038973	mi97h01.r1 Soares mous	3.50e-02
	134	17	27.9	272	9	AA580437	nn21c10.s1 NCI_CGAP_Co	3.50e-02
	135	17	27.9	285	16	T28195	EST31500 Homo sapiens	3.50e-02
c	136	17	27.9	290	22	AA025543	ze85b01.s1 Soares feta	3.50e-02
	137	17	27.9	293	6	C16941	Human placenta cDNA 5'	3.50e-02
c	138	17	27.9	294	26	B74931	RPC111-11O8.TV RPC111	3.50e-02
	139	17	27.9	300	15	C94869	Sus scrofa mRNA; expre	3.50e-02
c	140	17	27.9	303	19	HSC2WH032	H. sapiens partial cDN	3.50e-02
c	141	17	27.9	303	16	H22564	ym53f12.s1 Homo sapien	3.50e-02
c	142	17	27.9	303	26	B56492	CIT-HSP-2007A23.TF CIT	3.50e-02
c	143	17	27.9	304	9	AA570530	nk65e01.s1 NCI_CGAP_Sc	3.50e-02
c	144	17	27.9	307	22	AA184030	ms53f06.r1 Life Tech m	3.50e-02

	145	17	27.9	322	5	R08193	yf18f08.s1 Homo.sapien	3.50e-02
	146	17	27.9	327	5	T73838	yc56c10.r1 Homo sapien	3.50e-02
	147	17	27.9	335	26	B76754	RPCI11-16010.TP RPCI11	3.50e-02
	148	17	27.9	336	20	W36027	mc62b12.r1 Soares mous	3.50e-02
	149	17	27.9	343	8	AA504237	aa57d02.r1 NCI_CGAP_GC	3.50e-02
c	150	17	27.9	345	19	HSC1KH112	H. sapiens partial cDN	3.50e-02
c	151	17	27.9	350	5	T09923	0470m7 Plasmodium falc	3.50e-02
	152	17	27.9	354	7	AA474770	vd57d12.r1 Knowles Sol	3.50e-02
c	153	17	27.9	356	22	AA024980	ze80b06.s1 Soares feta	3.50e-02
	154	17	27.9	362	17	AI069694	TENU2738 T. cruzi epim	3.50e-02
	155	17	27.9	368	16	H52214	yr60g11.r1 Homo sapien	3.50e-02
c	156	17	27.9	369	16	R88864	ym96e05.r1 Homo sapien	3.50e-02
	157	17	27.9	370	20	W97483	mf97c05.r1 Soares mous	3.50e-02
c	158	17	27.9	372	29	HSU84062	Homo sapiens PAC clone	3.50e-02
	159	17	27.9	374	27	FR0033057	Fugu rubripes GSS sequ	3.50e-02
	160	17	27.9	377	13	AA863812	vx09c07.r1 Soares 2NbM	3.50e-02
c	161	17	27.9	377	15	AI029767	UI-R-C0-ir-b-04-0-UI.s	3.50e-02
c	162	17	27.9	380	17	AI004349	ot95b05.s1 Soares_tota	3.50e-02
c	163	17	27.9	382	8	AA278337	zs93g03.r1 NCI_CGAP_GC	3.50e-02
	164	17	27.9	385	22	AA023551	mh76f03.r1 Soares mous	3.50e-02
c	165	17	27.9	385	19	H67193	yu67f07.r1 Homo sapien	3.50e-02
	166	17	27.9	386	12	AA809473	ob85b06.s1 NCI_CGAP_GC	3.50e-02
c	167	17	27.9	387	7	AA448301	zw96f09.s1 Soares tota	3.50e-02
	168	17	27.9	389	18	AI168249	oo10b02.x1 Soares_NSF_	3.50e-02
	169	17	27.9	390	23	AA068002	mm25e11.r1 Stratagene	3.50e-02
	170	17	27.9	394	7	C24350	Dictyostelium discoide	3.50e-02
c	171	17	27.9	394	13	AA644333	ab64e04.s1 Stratagene	3.50e-02
c	172	17	27.9	395	11	AA726277	vu89d08.r1 Stratagene	3.50e-02
	173	17	27.9	396	10	AA655975	vs45b05.r1 Stratagene	3.50e-02
c	174	17	27.9	398	11	AA707483	ah40d08.s1 Soares test	3.50e-02
	175	17	27.9	399	10	AA649328	ns36c07.s1 NCI_CGAP_GC	3.50e-02
c	176	17	27.9	403	23	AA236896	zs43f12.s1 Soares NhHM	3.50e-02
c	177	17	27.9	405	5	R41812	yg04f06.s1 Homo sapien	3.50e-02
c	178	17	27.9	409	12	AA776745	ah49g04.s1 Soares test	3.50e-02
c	179	17	27.9	409	23	AA142663	ms12e11.r1 Stratagene	3.50e-02
c	180	17	27.9	411	19	H67181	yu67d07.r1 Homo sapien	3.50e-02
c	181	17	27.9	412	29	AQ136151	HS_3060_B1_G02_MR CIT	3.50e-02
	182	17	27.9	415	11	AA734421	vr90e02.r1 Barstead mo	3.50e-02
	183	17	27.9	417	23	AA244515	mx02b06.r1 Soares mous	3.50e-02
c	184	17	27.9	417	5	R41743	yg11d05.s1 Homo sapien	3.50e-02
c	185	17	27.9	418	26	B45064	HS-1060-B1-D01-MR.abi	3.50e-02
c	186	17	27.9	420	16	R44742	yg30e04.s1 Homo sapien	3.50e-02
c	187	17	27.9	421	10	AA402959	zu54d11.s1 Soares ovar	3.50e-02
c	188	17	27.9	422	13	AA863001	og99d04.s1 NCI_CGAP_Ki	3.50e-02
c	189	17	27.9	422	5	R44099	yg27b10.s1 Homo sapien	3.50e-02
c	190	17	27.9	424	18	AI081492	oz53g04.x1 Soares_sene	3.50e-02
c	191	17	27.9	425	27	AQ041756	CIT-HSP-2328M12.TR CIT	3.50e-02
	192	17	27.9	429	20	W62810	md71f12.r1 Soares mous	3.50e-02
c	193	17	27.9	432	16	H29321	ym60e03.s1 Homo sapien	3.50e-02
c	194	17	27.9	436	27	AQ058028	CIT-HSP-2337L13.TF CIT	3.50e-02
	195	17	27.9	439	12	AA771132	vt18a06.r1 Barstead mo	3.50e-02

c	196	17	27.9	439	6	AA010231	zi08a06.r1 Soares feta	3.50e-02
	197	17	27.9	441	27	AQ043807	CIT-HSP-2329K9.TR CIT-	3.50e-02
c	198	17	27.9	443	16	R60356	yh04h05.s1 Homo sapien	3.50e-02
c	199	17	27.9	444	26	B46128	HS-1063-A1-F12-MR.abi	3.50e-02
c	200	17	27.9	444	17	AU017285	Mus musculus 2-cell em	3.50e-02
c	201	17	27.9	447	13	AA487178	ab21c10.s1 Stratagene	3.50e-02
c	202	17	27.9	448	12	AA768393	oa63b03.s1 NCI_CGAP_GC	3.50e-02
c	203	17	27.9	448	9	AA564509	nh21d08.s1 NCI_CGAP_Pr	3.50e-02
c	204	17	27.9	448	16	H05071	yi69g04.s1 Homo sapien	3.50e-02
c	205	17	27.9	449	27	AQ055422	CIT-HSP-2344P18.TF CIT	3.50e-02
	206	17	27.9	453	5	T54617	yb41b11.s1 Homo sapien	3.50e-02
c	207	17	27.9	455	6	AA410187	zv23d07.s1 Soares NhHM	3.50e-02
c	208	17	27.9	455	8	AA210724	zr88e09.s1 NCI_CGAP_GC	3.50e-02
	209	17	27.9	456	29	AQ253223	HS_2008_B2_B12_MR CIT	3.50e-02
	210	17	27.9	457	9	AA596627	vm56e04.r1 Stratagene	3.50e-02
	211	17	27.9	458	19	N74465	za54b06.s1 Homo sapien	3.50e-02
	212	17	27.9	458	22	AA049860	mj14c11.r1 Soares mous	3.50e-02
c	213	17	27.9	458	16	H08197	yi92b02.s1 Homo sapien	3.50e-02
c	214	17	27.9	459	16	H06375	yi79b12.s1 Homo sapien	3.50e-02
	215	17	27.9	460	8	AA243678	zr67h02.s1 Soares NhHM	3.50e-02
c	216	17	27.9	462	17	AI056101	ow33b02.s1 Soares_para	3.50e-02
c	217	17	27.9	463	18	AI056506	oy53d06.x1 NCI_CGAP_Br	3.50e-02
c	218	17	27.9	465	16	H23096	ym51d10.s1 Homo sapien	3.50e-02
c	219	17	27.9	467	16	R60520	yh13f09.s1 Homo sapien	3.50e-02
	220	17	27.9	469	21	W24968	zb66d05.r1 Soares feta	3.50e-02
c	221	17	27.9	469	15	C77306	Mus musculus 3.5-dpc b	3.50e-02
c	222	17	27.9	470	9	AA568316	nf15c04.s1 NCI_CGAP_Pr	3.50e-02
	223	17	27.9	472	22	AA049838	mj14b10.r1 Soares_mous	3.50e-02
c	224	17	27.9	472	15	AI065358	TENU2245 T. cruzi epim	3.50e-02
c	225	17	27.9	474	27	B59458	CIT-HSP-2017K22.TR CIT	3.50e-02
c	226	17	27.9	476	18	AI073652	ox98b09.s1 Soares_sene	3.50e-02
c	227	17	27.9	478	5	R35404	yg65a09.r1 Homo sapien	3.50e-02
c	228	17	27.9	479	16	R61356	yh15b07.s1 Homo sapien	3.50e-02
c	229	17	27.9	480	5	R50822	yg62c01.s1 Homo sapien	3.50e-02
c	230	17	27.9	482	17	AI128551	qc68d03.x1 Soares_plac	3.50e-02
	231	17	27.9	485	27	AQ057551	CIT-HSP-2345I22.TR CIT	3.50e-02
	232	17	27.9	485	23	AA144324	mr75d05.r1 Stratagene	3.50e-02
c	233	17	27.9	490	17	AI005247	ot98h04.x1 Soares_tota	3.50e-02
	234	17	27.9	493	12	AA763672	vp09g08.r1 Soares mous	3.50e-02
	235	17	27.9	495	13	AA885196	am34b02.s1 Soares NFL	3.50e-02
	236	17	27.9	500	26	B19409	pEMS730-F23 Flow Sorte	3.50e-02
	237	17	27.9	503	7	AA448438	zw96f09.r1 Soares tota	3.50e-02
	238	17	27.9	505	24	AA266035	mz53b03.r1 Barstead mo	3.50e-02
	239	17	27.9	505	28	AQ112248	CIT-HSP-2371N2.TF CIT-	3.50e-02
	240	17	27.9	506	27	B69715	CIT-HSP-2028E7.TP CIT-	3.50e-02
c	241	17	27.9	510	28	AQ217944	HS_2001_B1_F12_MR CIT	3.50e-02
	242	17	27.9	511	20	W48190	me87d10.r1 Soares mous	3.50e-02
	243	17	27.9	512	22	AA172841	ms58b07.r1 Life Tech m	3.50e-02
c	244	17	27.9	512	22	AA172841	ms58b07.r1 Life Tech m	3.50e-02
	245	17	27.9	514	28	AQ183670	HS_3184_B2_D01_T7 CIT	3.50e-02
	246	17	27.9	515	10	AA616328	vn65a03.r1 Barstead mo	3.50e-02

c	247	17	27.9	518	8	AA562135	vk97f03.r1 Knowles Sol	3.50e-02
	248	17	27.9	520	28	AQ114708	CIT-HSP-2371D18.TF CIT	3.50e-02
c	249	17	27.9	549	18	AI086125	os40a06.s1 Soares_NhHM	3.50e-02
c	250	17	27.9	556	19	N29731	yw78h05.s1 Homo sapien	3.50e-02
c	251	17	27.9	557	12	AA745821	ob18a09.s1 NCI_CGAP_Ki	3.50e-02
c	252	17	27.9	566	27	B59561	CIT-HSP-342C17.TPB CIT	3.50e-02
c	253	17	27.9	577	13	AA487237	ab21e10.s1 Stratagene	3.50e-02
	254	17	27.9	589	17	AU019729	Mus musculus 8-cell em	3.50e-02
c	255	17	27.9	589	10	AA429437	zw47b01.r1 Soares tota	3.50e-02
c	256	17	27.9	630	7	AA537298	vk46d02.r1 Soares mous	3.50e-02
	257	17	27.9	640	27	AQ053341	CIT-HSP-2341I7.TR CIT-	3.50e-02
	258	17	27.9	664	8	AA550431	1583m3 gmbPfHB3.1, G.	3.50e-02
	259	17	27.9	714	27	AQ029273	RPCI11-37D17.TV RPCI11	3.50e-02
c	260	17	27.9	730	27	AQ055699	CIT-HSP-2342D12.TR CIT	3.50e-02
c	261	17	27.9	752	27	B67088	CIT-HSP-2017K22.TRB CI	3.50e-02
	262	17	27.9	843	28	AF010632	Homo sapiens chromosom	3.50e-02
c	263	17	27.9	1130	26	B09925	T18E3-T7 TAMU Arabidop	3.50e-02
	264	17	27.9	1174	26	B08789	F12G22-Sp6 IGF Arabido	3.50e-02
	265	16	26.2	43	7	MUSGS00488	Mouse 3'-directed cDN	4.85e-01
c	266	16	26.2	50	23	AA166134	ms28a01.r1 Stratagene	4.85e-01
c	267	16	26.2	106	17	AI004492	ot57c07.s1 Soares_test	4.85e-01
	268	16	26.2	122	9	AA571226	vm02e06.r1 Knowles Sol	4.85e-01
c	269	16	26.2	129	12	AA782981	ai63e01.s1 Soares test	4.85e-01
	270	16	26.2	150	21	AA012667	RPU1703AC Rat myometri	4.85e-01
	271	16	26.2	151	9	AA608142	vm81b01.r1 Knowles Sol	4.85e-01
	272	16	26.2	180	20	HSC0ND072	H. sapiens partial cDN	4.85e-01
	273	16	26.2	182	11	AA734309	vt26f05.r1 Barstead mo	4.85e-01
	274	16	26.2	192	23	AA217992	mv57h01.r1 Soares mous	4.85e-01
	275	16	26.2	203	13	AA824795	CT095.UNI Tomato Leaf	4.85e-01
	276	16	26.2	205	5	T84733	yd54b11.r1 Homo sapien	4.85e-01
	277	16	26.2	209	17	AI087692	RRAMCA975SR Brugia mal	4.85e-01
	278	16	26.2	211	9	AA590812	vm25b09.r1 Knowles Sol	4.85e-01
	279	16	26.2	212	11	AA683746	vr04c10.r1 Knowles Sol	4.85e-01
	280	16	26.2	218	9	AA608196	vm85d11.r1 Knowles Sol	4.85e-01
	281	16	26.2	219	6	AA036894	zk29c11.r1 Soares preg	4.85e-01
c	282	16	26.2	222	25	AA355451	EST63906 Jurkat T-cell	4.85e-01
	283	16	26.2	226	12	AA748122	nx79g02.s1 NCI_CGAP_Ew	4.85e-01
	284	16	26.2	226	26	B55933	CIT-HSP-2005I17.TF CIT	4.85e-01
c	285	16	26.2	232	12	AA709237	a121b08.s1 Soares test	4.85e-01
c	286	16	26.2	233	22	N44349	SW3ICA1185SK Brugia ma	4.85e-01
c	287	16	26.2	234	5	R46458	yj52h08.r1 Homo sapien	4.85e-01
	288	16	26.2	236	22	AA117322	mn26a12.r1 Beddington	4.85e-01
c	289	16	26.2	237	20	HSC0NF041	H. sapiens partial cDN	4.85e-01
c	290	16	26.2	241	28	AQ104509	HS_3127_A1_B08_MR CIT	4.85e-01
c	291	16	26.2	251	5	T89476	ye04h06.s1 Homo sapien	4.85e-01
c	292	16	26.2	251	16	R59942	yh11e10.r1 Homo sapien	4.85e-01
c	293	16	26.2	252	24	AA274752	vc04h07.r1 Soares mous	4.85e-01
c	294	16	26.2	263	18	AI175832	EST219403 Normalized r	4.85e-01
	295	16	26.2	269	16	H49044	yo21b08.s1 Homo sapien	4.85e-01
	296	16	26.2	271	10	AA623344	vn25b03.r1 Knowles Sol	4.85e-01
	297	16	26.2	272	11	AA692896	vr56a12.s1 Knowles Sol	4.85e-01

c	298	16	26.2	277	5	R22506	yh23a09.s1 Homo sapien	4.85e-01
	299	16	26.2	281	7	AA466909	vf10e04.r1 Knowles Sol	4.85e-01
c	300	16	26.2	281	19	HSC2UH102	H. sapiens partial cDN	4.85e-01
	301	16	26.2	287	22	AA013786	mh13c04.r1 Soares mous	4.85e-01
	302	16	26.2	289	9	AA575193	vg12h04.r1 Soares mous	4.85e-01
	303	16	26.2	300	9	C60027	C.elegans cDNA clone y	4.85e-01
	304	16	26.2	301	11	AA693527	zi51a05.s1 Soares feta	4.85e-01
c	305	16	26.2	303	25	AA333065	EST37099 Embryo, 8 wee	4.85e-01
c	306	16	26.2	306	25	AA365726	EST76560 Pineal gland	4.85e-01
	307	16	26.2	318	18	AI156218	ud95h09.r1 Soares mous	4.85e-01
c	308	16	26.2	319	17	AA409855	EST01584 Mouse 7.5 dpc	4.85e-01
	309	16	26.2	323	18	AI145362	UI-R-BT0-qe-a-08-0-UI.	4.85e-01
c	310	16	26.2	323	15	AI045642	UI-R-C1-ka-g-12-0-UI.s	4.85e-01
c	311	16	26.2	325	10	AA584087	no04g11.s1 NCI_CGAP_Ph	4.85e-01
	312	16	26.2	326	10	AA516784	vn68a04.r1 Barstead mo	4.85e-01
	313	16	26.2	330	9	AA575025	vm38b05.r1 Knowles Sol	4.85e-01
	314	16	26.2	335	7	AA465802	ve02d06.r1 Soares mous	4.85e-01
c	315	16	26.2	337	20	DY1B74A08	H. sapiens partial cDN	4.85e-01
c	316	16	26.2	340	5	R01564	ye75a09.r1 Homo sapien	4.85e-01
c	317	16	26.2	342	22	N97866	1413C3 czapPFDd2.1, De	4.85e-01
	318	16	26.2	343	21	AA008987	mg99e03.r1 Soares mous	4.85e-01
c	319	16	26.2	344	23	AA177520	mt25a11.r1 Soares mous	4.85e-01
c	320	16	26.2	345	10	AA665635	ag55d05.s1 Gessler Wil	4.85e-01
	321	16	26.2	346	14	AA930498	vo45d01.r1 Barstead mo	4.85e-01
c	322	16	26.2	349	24	AA320152	EST22469 Adipose tissu	4.85e-01
c	323	16	26.2	351	14	AA899496	UI-R-E0-de-d-09-0-UI.s	4.85e-01
c	324	16	26.2	354	8	AA489882	ab01d05.r1 Stratagene	4.85e-01
c	325	16	26.2	355	5	R08087	yf17e05.r1 Homo sapien	4.85e-01
c	326	16	26.2	355	16	H12174	ym17f08.r1 Homo sapien	4.85e-01
	327	16	26.2	358	27	AQ046667	RPCI11-31L8.TJ RPCI11	4.85e-01
	328	16	26.2	358	11	AA122192	zn28e03.s1 Stratagene	4.85e-01
	329	16	26.2	359	19	H88078	yw20d10.r1 Homo sapien	4.85e-01
c	330	16	26.2	359	16	R67019	yi26e06.s1 Homo sapien	4.85e-01
c	331	16	26.2	361	5	T07569	EST05459 Homo sapiens	4.85e-01
c	332	16	26.2	361	5	R11144	yf39c12.r1 Homo sapien	4.85e-01
c	333	16	26.2	362	5	R05606	ye92c06.r1 Homo sapien	4.85e-01
c	334	16	26.2	363	10	AA666783	vm51h05.r1 Stratagene	4.85e-01
c	335	16	26.2	365	22	AA099212	zl78g02.s1 Stratagene	4.85e-01
c	336	16	26.2	366	14	AA965182	UI-R-C0-hc-c-01-0-UI.s	4.85e-01
	337	16	26.2	367	24	AA305250	EST176463 Colon carcin	4.85e-01
	338	16	26.2	368	15	AA942865	EST198364 Normalized r	4.85e-01
	339	16	26.2	372	15	AI043444	TENU1003 T. cruzi epim	4.85e-01
	340	16	26.2	377	16	H26686	yl14g06.s1 Homo sapien	4.85e-01
c	341	16	26.2	380	13	AA496942	ae32d10.s1 Gessler Wil	4.85e-01
	342	16	26.2	382	17	AI121587	uc43f03.r1 Soares mous	4.85e-01
	343	16	26.2	382	26	B27083	T27L1TF TAMU Arabidops	4.85e-01
	344	16	26.2	382	11	AA115751	zn30a04.r1 Stratagene	4.85e-01
	345	16	26.2	385	26	AQ023189	HS_2177_B1_H03_MF CIT	4.85e-01
c	346	16	26.2	386	16	H15710	ym28d07.s1 Homo sapien	4.85e-01
c	347	16	26.2	387	27	AQ044273	CIT-HSP-2331L9.TF CIT-	4.85e-01
	348	16	26.2	392	11	AA704944	zj82e12.s1 Soares feta	4.85e-01

c	349	16	26.2	394	15	AA263141	PMY0505 KG1-a Lambda Z	4.85e-01
c	350	16	26.2	396	23	AA227840	zr28h01.s1 Stratagene	4.85e-01
	351	16	26.2	397	21	H49403	yq19f03.s1 Soares feta	4.85e-01
	352	16	26.2	400	10	T04355	402 Lambda-PRL2 Arabid	4.85e-01
c	353	16	26.2	405	22	AA114653	nn04e08.r1 Beddington	4.85e-01
	354	16	26.2	406	11	AA723079	zg83d09.s1 Soares feta	4.85e-01
	355	16	26.2	406	22	AA025783	ze91d08.s1 Soares feta	4.85e-01
c	356	16	26.2	406	27	B88384	RPCI11-19P9.TP RPCI11	4.85e-01
	357	16	26.2	408	6	AA432950	ve91b05.r1 Knowles Sol	4.85e-01
c	358	16	26.2	409	28	AQ111397	CIT-HSP-2374N14.TF CIT	4.85e-01
c	359	16	26.2	409	7	AA448156	zw82h11.r1 Soares test	4.85e-01
c	360	16	26.2	409	12	AA724996	ah97f04.s1 Soares NFL	4.85e-01
c	361	16	26.2	413	24	AA308836	EST179779 Colon carcin	4.85e-01
c	362	16	26.2	414	5	R10491	yf37d07.r1 Homo sapien	4.85e-01
	363	16	26.2	415	21	W30236	mc26h09.r1 Soares mous	4.85e-01
	364	16	26.2	416	13	AA822392	vw37b12.r1 Soares mous	4.85e-01
	365	16	26.2	418	29	AQ146320	HS_2231_B1_D06_MF CIT	4.85e-01
c	366	16	26.2	420	28	AQ177619	HS_2207_A2_D11_MF CIT	4.85e-01
	367	16	26.2	423	29	AQ136044	HS_3062_A1_D01_MR CIT	4.85e-01
	368	16	26.2	425	10	AA624683	vn79b02.r1 Knowles Sol	4.85e-01
	369	16	26.2	426	18	AI182865	uc66g09.r1 Soares mous	4.85e-01
	370	16	26.2	429	10	AA641744	ns19b08.r1 NCI_CGAP_GC	4.85e-01
c	371	16	26.2	429	5	R08098	yf17g02.r1 Homo sapien	4.85e-01
c	372	16	26.2	430	14	AA835983	oc80h12.s1 NCI_CGAP_GC	4.85e-01
c	373	16	26.2	430	13	AA609170	af12a12.s1 Soares test	4.85e-01
c	374	16	26.2	431	19	N73803	ys80f06.s1 Homo sapien	4.85e-01
	375	16	26.2	432	11	AA694007	zi48d01.s1 Soares feta	4.85e-01
c	376	16	26.2	434	26	B78268	T32A18TF TAMU Arabidop	4.85e-01
	377	16	26.2	436	7	HUM088A04B	Human fetal brain cDNA	4.85e-01
c	378	16	26.2	436	6	C20127	Rice cDNA, partial seq	4.85e-01
c	379	16	26.2	437	26	FR0009416	F.rubripes GSS sequenc	4.85e-01
c	380	16	26.2	438	28	AQ223106	HS_2010_B1_H09_T7 CIT	4.85e-01
	381	16	26.2	439	28	AQ196578	CIT-HSP-2380D10.TF CIT	4.85e-01
	382	16	26.2	439	23	AA142636	ma12d09.s1 Stratagene	4.85e-01
	383	16	26.2	440	10	T43041	6304 Lambda-PRL2 Arabi	4.85e-01
c	384	16	26.2	441	22	N97924	1548C3 czappFPDd2.1, De	4.85e-01
	385	16	26.2	443	7	AA446636	zw88g12.r1 Soares tota	4.85e-01
	386	16	26.2	444	28	AQ091191	HS_3009_B1_G01_T7 CIT	4.85e-01
	387	16	26.2	445	23	T03511	IB395 Infant brain, Be	4.85e-01
	388	16	26.2	447	21	W33689	mc28a11.r1 Soares mous	4.85e-01
c	389	16	26.2	447	15	AI031009	UI-R-C0-je-f-12-0-UI.s	4.85e-01
c	390	16	26.2	448	5	R54680	yj74f09.r1 Homo sapien	4.85e-01
c	391	16	26.2	448	15	AI044776	UI-R-C1-jt-b-02-0-UI.s	4.85e-01
c	392	16	26.2	448	6	AA126884	zl87a07.s1 Stratagene	4.85e-01
	393	16	26.2	450	28	AQ208955	HS_3230_A2_F04_MR CIT	4.85e-01
	394	16	26.2	452	22	W95298	ze05e03.r1 Soares feta	4.85e-01
	395	16	26.2	454	27	B80387	CIT-HSP-2045J17.TR CIT	4.85e-01
c	396	16	26.2	455	17	AI084867	qf17b11.x1 NCI_CGAP_Br	4.85e-01
	397	16	26.2	456	16	H24452	ym53a07.s1 Homo sapien	4.85e-01
c	398	16	26.2	457	28	AQ225080	HS_2017_A2_F11_MR CIT	4.85e-01
	399	16	26.2	459	23	AA203888	mu28h02.r1 Soares 2NbM	4.85e-01

c	400	16	26.2	460	8	AA393269	zt70c09.r1 Soares test	4.85e-01
c	401	16	26.2	463	27	AQ041356	CIT-HSP-2328J17.TR CIT	4.85e-01
	402	16	26.2	464	8	AA549164	vk85a01.s1 Knowles Sol	4.85e-01
c	403	16	26.2	465	7	AA447911	zx11b11.r1 Soares tota	4.85e-01
c	404	16	26.2	469	15	AI059595	UI-R-C1-ij-f-12-0-UI.s	4.85e-01
	405	16	26.2	474	28	AQ167534	HS_3070_B1_C07_MF CIT	4.85e-01
c	406	16	26.2	475	26	B19212	F8N4-T7 IGF Arabidopsi	4.85e-01
	407	16	26.2	477	10	AA666157	ac41e11.s1 Stratagene	4.85e-01
c	408	16	26.2	478	5	T78101	yd79h07.s1 Homo sapien	4.85e-01
c	409	16	26.2	479	21	W08634	mb42b12.r1 Soares mous	4.85e-01
c	410	16	26.2	482	26	B55653	CIT-HSP-388N11.TF CIT-	4.85e-01
c	411	16	26.2	483	26	B26740	T18E7TF TAMU Arabidops	4.85e-01
	412	16	26.2	484	29	AQ230964	HS_2037_B2_B08_MR CIT	4.85e-01
	413	16	26.2	486	17	AI024244	ov71b05.s1 Soares_test	4.85e-01
	414	16	26.2	487	6	W87684	zh68b07.s1 Soares feta	4.85e-01
	415	16	26.2	489	15	AA996729	UI-R-C0-hf-g-06-0-UI.s	4.85e-01
c	416	16	26.2	489	24	AA266891	mz98g12.r1 Soares mous	4.85e-01
	417	16	26.2	492	14	C90605	Dictyostelium discoide	4.85e-01
	418	16	26.2	495	29	AQ143290	HS_3067_B1_A02_MF CIT	4.85e-01
c	419	16	26.2	497	29	AQ252616	HS_2010_B2_B05_MR CIT	4.85e-01
	420	16	26.2	498	14	C92913	Dictyostelium discoide	4.85e-01
c	421	16	26.2	499	26	B36158	HS-1038-A1-E09-MR.abi	4.85e-01
	422	16	26.2	501	18	AI149442	qc72h01.x1 Soares_plac	4.85e-01
	423	16	26.2	502	14	C92796	Dictyostelium discoide	4.85e-01
	424	16	26.2	503	14	C90570	Dictyostelium discoide	4.85e-01
	425	16	26.2	508	13	AA838070	oe93g12.s1 NCI_CGAP_Co	4.85e-01
c	426	16	26.2	511	26	B48471	RPCI11-2C13.TP RPCI11	4.85e-01
c	427	16	26.2	513	15	AI045220	UI-R-C1-ju-h-08-0-UI.s	4.85e-01
c	428	16	26.2	513	28	AQ171805	HS_3179_A1_F12_T7 CIT	4.85e-01
	429	16	26.2	514	18	AI180782	ub76c04.r1 Soares mous	4.85e-01
	430	16	26.2	516	28	AQ214453	HS_3243_B1_F04_T7 CIT	4.85e-01
	431	16	26.2	516	28	AQ209205	HS_3236_B1_B05_T7 CIT	4.85e-01
c	432	16	26.2	516	23	AA170787	ms94g10.r1 Soares mous	4.85e-01
c	433	16	26.2	517	9	AA533821	nj94b07.s1 NCI_CGAP_Pr	4.85e-01
c	434	16	26.2	522	11	AA728497	33676 CD4-6 Arabidopsi	4.85e-01
c	435	16	26.2	523	23	AA161887	mr91a04.r1 Stratagene	4.85e-01
c	436	16	26.2	527	8	AA101245	zn43b08.s1 Stratagene	4.85e-01
c	437	16	26.2	530	15	C81181	Mus musculus 3.5-dpc b	4.85e-01
c	438	16	26.2	530	27	B94996	CIT-HSP-2167H6.TF CIT-	4.85e-01
	439	16	26.2	531	14	AA898719	NCC3A3T7 Conidial Neur	4.85e-01
c	440	16	26.2	536	14	AA955618	UI-R-E1-fb-b-07-0-UI.s	4.85e-01
c	441	16	26.2	540	28	AQ198987	RPCI11-63B23.TK RPCI11	4.85e-01
c	442	16	26.2	542	15	C81013	Mus musculus 3.5-dpc b	4.85e-01
	443	16	26.2	547	25	AA260584	mx76c09.r1 Soares mous	4.85e-01
c	444	16	26.2	554	29	AQ237321	RPCI11-66L7.TK RPCI11	4.85e-01
c	445	16	26.2	555	11	AA081852	zm64a12.r1 Stratagene	4.85e-01
	446	16	26.2	559	14	AA897908	NCC1B8T7 Conidial Neur	4.85e-01
c	447	16	26.2	559	21	W60744	zd26h06.r1 Soares feta	4.85e-01
	448	16	26.2	561	14	AA848331	EST191091 Normalized r	4.85e-01
	449	16	26.2	564	11	AA700465	zj74e02.s1 Soares feta	4.85e-01
	450	16	26.2	571	26	B27582	F27C1TFB IGF Arabidops	4.85e-01

c	451	16	26.2	571	27	B57921	CIT-HSP-2010P17.TF CIT	4.85e-01
	452	16	26.2	572	14	C92907	Dictyostelium discoide	4.85e-01
c	453	16	26.2	572	28	AQ156286	nbxb0007H05r CUGI Rice	4.85e-01
c	454	16	26.2	573	19	T23157	C2H2-218T Homo sapiens	4.85e-01
	455	16	26.2	575	8	AA549915	0988m3 gmbPfHB3.1, G.	4.85e-01
c	456	16	26.2	575	15	C81445	Mus musculus 3.5-dpc b	4.85e-01
c	457	16	26.2	576	26	FR0020386	F.rubripes GSS sequenc	4.85e-01
c	458	16	26.2	577	8	AA402512	zt63g11.r1 Soares test	4.85e-01
c	459	16	26.2	578	23	AA105699	mm66a08.r1 Stratagene	4.85e-01
	460	16	26.2	580	10	AA648356	ns21a08.s1 NCI_CGAP_GC	4.85e-01
c	461	16	26.2	583	23	AA086854	mm86d04.r1 Stratagene	4.85e-01
	462	16	26.2	586	11	AA728662	33716 CD4-6 Arabidopsi	4.85e-01
	463	16	26.2	590	11	AA690565	wu53a09.r1 Soares mous	4.85e-01
c	464	16	26.2	594	15	C77665	Mus musculus 3.5-dpc b	4.85e-01
	465	16	26.2	601	26	FR0018606	F.rubripes GSS sequenc	4.85e-01
	466	16	26.2	601	26	B62216	T20N18TF TAMU Arabidop	4.85e-01
	467	16	26.2	610	10	AA654350	nt12c04.s1 NCI_CGAP_Pr	4.85e-01
c	468	16	26.2	616	28	AQ156628	nbxb0008M23r CUGI Rice	4.85e-01
	469	16	26.2	619	26	FR0003415	F.rubripes GSS sequenc	4.85e-01
c	470	16	26.2	624	7	HUM000S166	Human HepG2 3'-directe	4.85e-01
c	471	16	26.2	634	15	C77784	Mus musculus 3.5-dpc b	4.85e-01
c	472	16	26.2	638	27	AG011129	Homo sapiens genomic D	4.85e-01
c	473	16	26.2	638	27	AG011130	Homo sapiens genomic D	4.85e-01
	474	16	26.2	642	27	AG011141	Homo sapiens genomic D	4.85e-01
	475	16	26.2	642	27	AG011143	Homo sapiens genomic D	4.85e-01
	476	16	26.2	649	15	AA991100	BSBmMFSZ03M20SK Brugia	4.85e-01
c	477	16	26.2	654	27	AQ049365	cLM-14a11-u cLM Giardi	4.85e-01
c	478	16	26.2	663	6	W87854	zh68b07.r1 Soares feta	4.85e-01
	479	16	26.2	668	26	B15474	345D13.TP CIT978SKA1 H	4.85e-01
c	480	16	26.2	669	27	AQ053276	CIT-HSP-2341E7.TR CIT-	4.85e-01
	481	16	26.2	680	26	B18299	F1C15-Sp6 IGF Arabidop	4.85e-01
c	482	16	26.2	711	8	AA550557	1720m3 gmbPfHB3.1, G.	4.85e-01
	483	16	26.2	720	28	AG013373	Homo sapiens genomic D	4.85e-01
	484	16	26.2	720	26	AG005696	Homo sapiens genomic D	4.85e-01
	485	16	26.2	732	27	B66071	CIT-HSP-2023O15.TF CIT	4.85e-01
c	486	16	26.2	732	28	AF010781	Homo sapiens chromosom	4.85e-01
c	487	16	26.2	754	26	B06928	cSRL-88E12-u cSRL flow	4.85e-01
c	488	16	26.2	815	15	AI035120	TENG0143 T. cruzi epim	4.85e-01
c	489	16	26.2	817	26	B11106	T4K8-T7.1 TAMU Arabido	4.85e-01
	490	16	26.2	840	26	B21143	F17K13-T7 IGF Arabidop	4.85e-01
c	491	16	26.2	842	26	B18532	T3B5-Sp6 TAMU Arabidop	4.85e-01
c	492	16	26.2	897	26	B08003	F2O7-T7 IGF Arabidopsi	4.85e-01
c	493	15	24.6	71	10	AA426827	vf20d10.s1 Knowles Sol	5.94e+00
c	494	15	24.6	72	10	AA426826	vf20d09.s1 Knowles Sol	5.94e+00
c	495	15	24.6	99	7	MUSGS01674	Mouse 3'-directed cDN	5.94e+00
	496	15	24.6	114	18	AI076562	os17f02.x1 Soares_feta	5.94e+00
c	497	15	24.6	120	19	N39435	yy49e07.s1 Homo sapien	5.94e+00
c	498	15	24.6	135	25	AA348563	EST55030 Hippocampus I	5.94e+00
c	499	15	24.6	135	7	AA493098	vi71h11.r1 Stratagene	5.94e+00
c	500	15	24.6	136	25	AA348565	EST55032 Hippocampus I	5.94e+00
c	501	15	24.6	136	25	AA348558	EST55025 Hippocampus I	5.94e+00

c	502	15	24.6	141	16	H07596	crn176 Brassica napus	5.94e+00
	503	15	24.6	145	23	AA234999	zr50h05.s1 Soares NhHM	5.94e+00
	504	15	24.6	148	15	AA250328	UI-R-C1-jy-d-02-0-UI.s	5.94e+00
	505	15	24.6	151	22	AA104421	mo42d05.r1 Life Tech m	5.94e+00
c	506	15	24.6	157	25	AA348468	EST54923 Hippocampus I	5.94e+00
c	507	15	24.6	158	22	AA162228	mn44g03.r1 Beddington	5.94e+00
c	508	15	24.6	159	15	AU010654	Schizosaccharomyces po	5.94e+00
c	509	15	24.6	159	15	AU010750	Schizosaccharomyces po	5.94e+00
c	510	15	24.6	160	14	AA956758	UI-R-E1-fr-e-03-0-UI.s	5.94e+00
c	511	15	24.6	162	9	AA589137	vi53h03.r1 Beddington	5.94e+00
	512	15	24.6	163	23	W36238	249 Mouse VM cDNA libr	5.94e+00
c	513	15	24.6	167	25	AA348473	EST54931 Hippocampus I	5.94e+00
c	514	15	24.6	170	12	AA813627	ah93e12.s1 Soares NFL	5.94e+00
c	515	15	24.6	176	17	AA923438	ol47c03.s1 Soares_NFL_	5.94e+00
c	516	15	24.6	180	5	R41165	HX662-f Homo sapiens c	5.94e+00
c	517	15	24.6	182	21	W87044	mf38d10.r1 Soares mous	5.94e+00
	518	15	24.6	183	5	R40293	yf74a05.s1 Homo sapien	5.94e+00
	519	15	24.6	188	23	AA233290	zr30f01.s1 Stratagene	5.94e+00
c	520	15	24.6	190	20	T15433	IB1181 Infant brain, B	5.94e+00
c	521	15	24.6	193	10	AA294109	SWOv3MCA1506SK Onchoce	5.94e+00
	522	15	24.6	195	20	HSC07F092	H. sapiens partial cDN	5.94e+00
c	523	15	24.6	196	25	AA348500	EST54961 Hippocampus I	5.94e+00
c	524	15	24.6	198	25	AA348368	EST54878 Hippocampus I	5.94e+00
c	525	15	24.6	201	25	AA348363	EST54872 Hippocampus I	5.94e+00
c	526	15	24.6	204	19	H72194	ys02c03.r1 Homo sapien	5.94e+00
c	527	15	24.6	205	25	AA348228	EST54886 Hippocampus I	5.94e+00
c	528	15	24.6	206	23	AA220091	mv63a10.r1 Soares mous	5.94e+00
c	529	15	24.6	210	17	AU011312	Schizosaccharomyces po	5.94e+00
c	530	15	24.6	213	16	H38429	yp66e12.s1 Homo sapien	5.94e+00
	531	15	24.6	219	23	AA213117	mw79f10.r1 Soares mous	5.94e+00
c	532	15	24.6	223	23	AA254197	va07g03.r1 Soares mous	5.94e+00
c	533	15	24.6	223	12	AA736745	oa19a10.s1 NCI_CGAP_GC	5.94e+00
c	534	15	24.6	224	14	AA985142	am75e03.s1 Stratagene	5.94e+00
c	535	15	24.6	224	25	AA348444	EST55085 Hippocampus I	5.94e+00
c	536	15	24.6	226	19	HSC1QCO42	H. sapiens partial cDN	5.94e+00
c	537	15	24.6	228	23	AA257587	MBAPCZ1F8T3 Brugia mal	5.94e+00
	538	15	24.6	232	18	AI101054	EST210343 Normalized r	5.94e+00
c	539	15	24.6	233	23	AA255429	zr85f05.s1 Soares NhHM	5.94e+00
c	540	15	24.6	234	14	AA964305	UI-R-C0-gu-b-10-0-UI.s	5.94e+00
c	541	15	24.6	234	10	AA636902	vn77d06.r1 Knowles Sol	5.94e+00
	542	15	24.6	236	23	AA065836	mm07d09.r1 Stratagene	5.94e+00
	543	15	24.6	241	26	B28888	T21K13TF TAMU Arabidop	5.94e+00
	544	15	24.6	244	17	AI112910	UI-R-A1-el-d-02-0-UI.s	5.94e+00
c	545	15	24.6	249	13	H33429	EST109421 Rat PC-12 ce	5.94e+00
	546	15	24.6	250	13	AA822126	vp35b06.r1 Barstead mo	5.94e+00
c	547	15	24.6	254	20	T15858	IB1996 Infant brain, B	5.94e+00
c	548	15	24.6	254	25	AA350222	EST57561 Infant brain	5.94e+00
	549	15	24.6	255	9	AA597130	vo35e02.r1 Barstead mo	5.94e+00
	550	15	24.6	263	15	AA953665	oc02d03.s1 Soares_NFL_	5.94e+00
	551	15	24.6	264	17	AA408810	EST03292 Mouse 7.5 dpc	5.94e+00
	552	15	24.6	265	17	AA408433	EST02903 Mouse 7.5 dpc	5.94e+00

c	553	15	24.6	267	5	T09921	0469m7 Plasmodium falc 5.94e+00
c	554	15	24.6	267	14	AA904126	og20c10.s1 NCI_CGAP_PN 5.94e+00
	555	15	24.6	268	7	HUM331F12B	Human aorta cDNA 5'-en 5.94e+00
c	556	15	24.6	268	19	H91660	ys80f01.s1 Homo sapien 5.94e+00
	557	15	24.6	273	5	R11145	yf39c12.s1 Homo sapien 5.94e+00
	558	15	24.6	273	12	C84830	Dictyostelium discoide 5.94e+00
c	559	15	24.6	274	14	AA837703	oe06c04.s1 NCI_CGAP_Ov 5.94e+00
c	560	15	24.6	279	14	AA961766	or60f03.s1 NCI_CGAP_GC 5.94e+00
c	561	15	24.6	279	6	AA382063	EST95234 Activated T-c 5.94e+00
c	562	15	24.6	282	17	AI023016	ow57f05.s1 Soares_para 5.94e+00
c	563	15	24.6	282	8	AA211665	zn57a02.s1 Stratagene 5.94e+00
	564	15	24.6	283	14	AA937372	oj03b01.s1 NCI_CGAP_Me 5.94e+00
c	565	15	24.6	283	20	T15614	IB1633 Infant brain, B 5.94e+00
	566	15	24.6	287	14	AA991483	os91h01.s1 NCI_CGAP_GC 5.94e+00
	567	15	24.6	287	17	AU023212	Mus musculus unfertili 5.94e+00
c	568	15	24.6	289	27	AQ089150	HS_2207_B2_G10_MF CIT 5.94e+00
	569	15	24.6	291	23	AA119414	mp81c10.r1 Soares 2NbM 5.94e+00
	570	15	24.6	292	20	HSC2VB112	H. sapiens partial cDN 5.94e+00
c	571	15	24.6	295	5	T06536	EST04425 Homo sapiens 5.94e+00
	572	15	24.6	297	20	ATTS3779	A. thaliana transcribe 5.94e+00
c	573	15	24.6	299	16	R92818	yq17f10.s1 Homo sapien 5.94e+00
	574	15	24.6	300	9	C53610	C.elegans cDNA clone y 5.94e+00
	575	15	24.6	300	9	C35126	C.elegans cDNA clone y 5.94e+00
c	576	15	24.6	302	14	AA916743	oh84d03.s1 NCI_CGAP_Co 5.94e+00
c	577	15	24.6	303	25	AA369632	EST81066 Placenta II H 5.94e+00
c	578	15	24.6	304	11	AA708657	zf92b02.s1 Soares pine 5.94e+00
	579	15	24.6	305	8	AA559553	MC070552.NA3 TBRI-Sabe 5.94e+00
c	580	15	24.6	305	20	W78389	me78c06.r1 Soares mous 5.94e+00
c	581	15	24.6	305	9	AA606362	vo47g09.r1 Barstead mo 5.94e+00
c	582	15	24.6	308	16	T30706	EST21114 Homo sapiens 5.94e+00
	583	15	24.6	310	9	AA569194	nm30e11.s1 NCI_CGAP_Li 5.94e+00
c	584	15	24.6	310	16	R46021	yg47b06.s1 Homo sapien 5.94e+00
	585	15	24.6	312	15	AI044045	UI-R-C1-js-d-02-0-UI.s 5.94e+00
	586	15	24.6	313	19	H90132	yu88a08.s1 Homo sapien 5.94e+00
	587	15	24.6	314	14	H31536	EST105658 Rat PC-12 ce 5.94e+00
	588	15	24.6	315	17	AA409788	EST01508 Mouse 7.5 dpc 5.94e+00
	589	15	24.6	316	9	AA608461	vo52c05.r1 Barstead mo 5.94e+00
c	590	15	24.6	316	20	W57289	md68g09.r1 Soares mous 5.94e+00
	591	15	24.6	318	19	H69276	EST00074 Homo sapiens 5.94e+00
	592	15	24.6	318	27	AQ065918	HS_2240_A1_C01_MF CIT 5.94e+00
c	593	15	24.6	320	23	AA145912	mq49h11.r1 Soares 2NbM 5.94e+00
	594	15	24.6	322	18	AI140584	ow84e08.s1 Soares_feta 5.94e+00
	595	15	24.6	323	23	AA087630	mn97e08.r1 Stratagene 5.94e+00
	596	15	24.6	324	5	T64490	yc12g12.s1 Homo sapien 5.94e+00
c	597	15	24.6	328	24	AA321162	EST23616 Fronta lobe 5.94e+00
	598	15	24.6	330	14	AA900761	UI-R-E0-di-d-09-0-UI.s 5.94e+00
c	599	15	24.6	330	25	AA348822	EST55334 Right hemisph 5.94e+00
c	600	15	24.6	331	20	DY1B81C11	H. sapiens partial cDN 5.94e+00
c	601	15	24.6	331	25	AA348816	EST55328 Right hemisph 5.94e+00
	602	15	24.6	333	13	AA805091	oa89c10.s1 NCI_CGAP_GC 5.94e+00
	603	15	24.6	333	15	AI059038	UI-R-C1-lq-h-12-0-UI.s 5.94e+00

c	604	15	24.6	333	25	AA350006	EST57282 Infant brain 5.94e+00
	605	15	24.6	335	7	CELK101F8R	C.elegans cDNA clone y 5.94e+00
	606	15	24.6	336	6	CELK022DYR	C.elegans cDNA clone y 5.94e+00
c	607	15	24.6	337	5	T92994	ye23c12.r1 Homo sapien 5.94e+00
c	608	15	24.6	337	18	AI080303	os51h01.x1 Soares_sene 5.94e+00
	609	15	24.6	339	27	B74709	CIT-HSP-2034G23.TF CIT 5.94e+00
	610	15	24.6	340	7	CELK072F3R	C.elegans cDNA clone y 5.94e+00
c	611	15	24.6	340	19	N66698	yy71h04.s1 Homo sapien 5.94e+00
c	612	15	24.6	340	25	AA348806	EST55313 Right hemisph 5.94e+00
	613	15	24.6	341	7	CELK123D6R	C.elegans cDNA clone y 5.94e+00
	614	15	24.6	343	8	AA563048	vl45c09.r1 Stratagene 5.94e+00
	615	15	24.6	344	24	AA290463	vb18g04.r1 Soares mous 5.94e+00
	616	15	24.6	344	14	AA955058	UI-R-A1-ey-g-02-0-UI.s 5.94e+00
	617	15	24.6	346	16	R63682	yi10f01.s1 Homo sapien 5.94e+00
c	618	15	24.6	347	14	AA850851	EST193619 Normalized r 5.94e+00
	619	15	24.6	349	21	W06704	T2304 MVAT4 bloodstrea 5.94e+00
c	620	15	24.6	349	19	N49042	yy78c09.r1 Homo sapien 5.94e+00
c	621	15	24.6	349	19	H57712	yr27g12.s1 Homo sapien 5.94e+00
	622	15	24.6	350	5	T18254	0851c3 Plasmodium falc 5.94e+00
c	623	15	24.6	350	7	AA456598	zx75a05.s1 Soares ovar 5.94e+00
	624	15	24.6	351	17	AU018690	Mus musculus 8-cell em 5.94e+00
	625	15	24.6	351	18	AI077578	oz33h12.x1 Soares_tota 5.94e+00
c	626	15	24.6	352	13	AA884357	al60b02.s1 Soares NFL 5.94e+00
c	627	15	24.6	352	5	R10306	yf36g11.r1 Homo sapien 5.94e+00
	628	15	24.6	353	7	AA498939	vi91d02.r1 Stratagene 5.94e+00
	629	15	24.6	353	16	H04284	yj50c04.s1 Homo sapien 5.94e+00
	630	15	24.6	353	20	N81184	yw36f08.r1 Homo sapien 5.94e+00
	631	15	24.6	353	14	C91127	Dictyostelium discoide 5.94e+00
	632	15	24.6	353	8	AA213568	zr93a12.s1 NCI_CGAP_GC 5.94e+00
	633	15	24.6	354	14	AA905570	oj97d03.s1 Soares_NFL_ 5.94e+00
c	634	15	24.6	354	24	AA272166	vb63e07.r1 Barstead mo 5.94e+00
c	635	15	24.6	354	8	AA216350	nc10c10.s1 NCI_CGAP_Pr 5.94e+00
	636	15	24.6	355	22	N98120	2212c3 crapPFDd2.1, De 5.94e+00
c	637	15	24.6	355	8	AA493177	ne56e01.s1 NCI_CGAP_Co 5.94e+00
	638	15	24.6	356	10	AA665631	ag55c11.s1 Gessler Wil 5.94e+00
	639	15	24.6	356	15	AI053367	TENU1531 T. cruzi epim 5.94e+00
	640	15	24.6	356	14	AA959475	vw60h08.s1 Soares mous 5.94e+00
c	641	15	24.6	356	21	W88101	mf70a08.r1 Soares mous 5.94e+00
c	642	15	24.6	357	27	AQ063652	HS_2263_A1_A01_MF CIT 5.94e+00
c	643	15	24.6	360	19	H91566	yw29b07.r1 Homo sapien 5.94e+00
c	644	15	24.6	360	16	R70125	yi48e09.r1 Homo sapien 5.94e+00
c	645	15	24.6	361	24	AA270447	va68a01.r1 Soares mous 5.94e+00
	646	15	24.6	362	6	CELK039B2R	C.elegans cDNA clone y 5.94e+00
c	647	15	24.6	362	11	AA701441	zi59b01.s1 Soares feta 5.94e+00
	648	15	24.6	363	17	AA407556	EST00921 Mouse 7.5 dpc 5.94e+00
c	649	15	24.6	364	15	AI043386	BSBmMFSE04H8SK Brugia 5.94e+00
c	650	15	24.6	365	23	AA230611	mw03e09.r1 Soares mous 5.94e+00
c	651	15	24.6	366	27	AQ076592	CIT-HSP-2363A1.TR CIT- 5.94e+00
c	652	15	24.6	366	17	AA995356	or74a11.s1 NCI_CGAP_Lu 5.94e+00
c	653	15	24.6	367	14	AA947546	od87b12.s1 NCI_CGAP_Ov 5.94e+00
c	654	15	24.6	367	19	H79388	yu48d07.r1 Homo sapien 5.94e+00

	655	15	24.6	369	15	AA944717	EST200216 Normalized r 5.94e+00
	656	15	24.6	369	15	C80788	Mus musculus 3.5-dpc b 5.94e+00
	657	15	24.6	369	16	M85459	EST01975 Homo sapiens 5.94e+00
c	658	15	24.6	369	25	AA348332	EST54767 Hippocampus I 5.94e+00
c	659	15	24.6	372	24	AA267024	mz96e05.r1 Soares mous 5.94e+00
	660	15	24.6	374	29	AQ243966	HS_2058_B1_D11_MR CIT 5.94e+00
	661	15	24.6	374	26	B25103	F24N21TR IGF Arabidops 5.94e+00
c	662	15	24.6	374	13	AA760808	nz10e11.s1 NCI_CGAP_GC 5.94e+00
	663	15	24.6	375	9	C64548	C.elegans cDNA clone y 5.94e+00
	664	15	24.6	375	17	AI070388	UI-R-Y0-1t-e-09-0-UI.s 5.94e+00
	665	15	24.6	375	28	AQ099333	HS_3042_B2_D05_MF CIT 5.94e+00
c	666	15	24.6	375	19	N68312	yz58a12.s1 Homo sapien 5.94e+00
	667	15	24.6	376	5	R06325	yf08d06.r1 Homo sapien 5.94e+00
c	668	15	24.6	376	5	R41282	yf85e01.s1 Homo sapien 5.94e+00
c	669	15	24.6	376	5	R40748	yf80g11.s1 Homo sapien 5.94e+00
	670	15	24.6	377	6	CELK018P7R	C.elegans cDNA clone y 5.94e+00
c	671	15	24.6	377	15	AI026445	TENU0656 T. cruzi epim 5.94e+00
c	672	15	24.6	377	20	W02641	za52g02.r1 Soares feta 5.94e+00
c	673	15	24.6	378	14	AA958966	ub59g04.s1 Soares mous 5.94e+00
	674	15	24.6	381	16	H05441	yl81a07.s1 Homo sapien 5.94e+00
c	675	15	24.6	381	25	AA350751	EST58281 Infant brain 5.94e+00
c	676	15	24.6	383	14	AA969240	on57b04.s1 Soares_NFL_ 5.94e+00
	677	15	24.6	384	17	AI041462	ow36d09.s1 Soares_para 5.94e+00
	678	15	24.6	384	23	AA255497	zr85a03.s1 Soares NhHM 5.94e+00
c	679	15	24.6	384	11	AA712563	31696 Lambda-PRL2 Arab 5.94e+00
	680	15	24.6	385	26	AQ016336	CIT-HSP-2312G9.TF CIT- 5.94e+00
c	681	15	24.6	385	25	AA348828	EST55341 Right hemisph 5.94e+00
c	682	15	24.6	385	27	AQ070062	HS_3027_B1_H08_MF CIT 5.94e+00
	683	15	24.6	386	14	C89705	Dictyostelium discoide 5.94e+00
c	684	15	24.6	386	25	AA348360	EST54825 Hippocampus I 5.94e+00
c	685	15	24.6	386	15	AI030207	UI-R-C0-jb-a-12-0-UI.s 5.94e+00
c	686	15	24.6	387	6	AA054754	zk68h02.s1 Soares preg 5.94e+00
c	687	15	24.6	389	28	AQ096271	HS_3037_A1_H03_MF CIT 5.94e+00
c	688	15	24.6	389	12	AA796285	vs94a06.r1 Barstead mo 5.94e+00
	689	15	24.6	390	6	C16546	Human aorta cDNA 5'-en 5.94e+00
c	690	15	24.6	390	28	AQ177283	HS_3225_A1_E03_MR CIT 5.94e+00
c	691	15	24.6	391	27	AQ041141	CIT-HSP-2336J3.TR CIT- 5.94e+00
c	692	15	24.6	391	24	AA280977	va26c06.r1 GuayWoodfor 5.94e+00
c	693	15	24.6	392	7	HUM081D10B	Human fetal brain cDNA 5.94e+00
c	694	15	24.6	392	13	AA628796	af42d12.s1 Soares tota 5.94e+00
	695	15	24.6	393	6	CELK018F8R	C.elegans cDNA clone y 5.94e+00
c	696	15	24.6	393	5	T94022	ye33b03.r1 Homo sapien 5.94e+00
c	697	15	24.6	393	27	AQ044294	CIT-HSP-2330O12.TR CIT 5.94e+00
	698	15	24.6	395	17	AU018317	Mus musculus 8-cell em 5.94e+00
c	699	15	24.6	395	5	T60240	yb68f07.r1 Homo sapien 5.94e+00
c	700	15	24.6	395	28	AQ192681	HS_2251_B2_B06_MF CIT 5.94e+00
	701	15	24.6	396	10	AA421143	ze79d12.s1 Soares test 5.94e+00
	702	15	24.6	396	18	AI177768	EST222141 Normalized r 5.94e+00
	703	15	24.6	396	26	B29096	T26D18TR TAMU Arabidop 5.94e+00
	704	15	24.6	398	8	AA549944	1021m3 gmbPfHB3.1, G. 5.94e+00
	705	15	24.6	398	6	AA149178	z143d12.s1 Soares preg 5.94e+00

	706	15	24.6	399	20	W20050	zb40a04.r1 Soares para 5.94e+00
c	707	15	24.6	399	23	AA177458	mt30a09.r1 Soares mous 5.94e+00
c	708	15	24.6	399	23	AA239321	my40f05.r1 Barstead mo 5.94e+00
c	709	15	24.6	399	24	AA308825	EST179642 Retina II Ho 5.94e+00
c	710	15	24.6	400	17	AI131839	uc38f02.r1 Soares mous 5.94e+00
c	711	15	24.6	401	27	AQ031550	HS_2224_B1_B10_MF CIT 5.94e+00
c	712	15	24.6	402	11	AA709374	zl60e01.s1 Soares preg 5.94e+00
c	713	15	24.6	402	8	AA489199	aa57b03.s1 NCI_CGAP_GC 5.94e+00
c	714	15	24.6	402	17	AI024865	ov39g06.x1 Soares_test 5.94e+00
c	715	15	24.6	402	16	R99321	yq79a03.s1 Homo sapien 5.94e+00
	716	15	24.6	403	15	AI060181	UI-R-C1-li-d-06-0-UI.s 5.94e+00
c	717	15	24.6	403	13	AA164873	zp01a09.s1 Stratagene 5.94e+00
c	718	15	24.6	404	6	AA399038	zt87d07.s1 Soares test 5.94e+00
c	719	15	24.6	404	16	R81231	yj03c12.s1 Homo sapien 5.94e+00
	720	15	24.6	405	26	AQ010038	HS_2263_A2_MR_F01 CIT 5.94e+00
	721	15	24.6	407	27	AQ040046	CIT-HSP-2333J7.TF CIT- 5.94e+00
c	722	15	24.6	408	5	T54190	ya92c12.r1 Homo sapien 5.94e+00
c	723	15	24.6	408	14	AA847313	of01h09.s1 NCI_CGAP_Co 5.94e+00
	724	15	24.6	410	28	AQ198307	RPCI11-47I19.TJ RPCI11 5.94e+00
c	725	15	24.6	410	22	AA122605	mn25g05.r1 Beddington 5.94e+00
c	726	15	24.6	410	15	AI029301	UI-R-C0-iq-h-01-0-UI.s 5.94e+00
	727	15	24.6	411	28	AQ195264	RPCI11-66O9.TJ RPCI11 5.94e+00
c	728	15	24.6	412	14	AA900251	UI-R-E0-de-e-12-0-UI.s 5.94e+00
	729	15	24.6	413	5	R42991	yg16f10.s1 Homo sapien 5.94e+00
c	730	15	24.6	413	27	AQ026875	CIT-HSP-2324M13.TR CIT 5.94e+00
c	731	15	24.6	413	25	AA348321	EST54755 Hippocampus I 5.94e+00
	732	15	24.6	414	13	AA859842	UI-R-E0-cc-h-03-0-UI.s 5.94e+00
	733	15	24.6	416	6	CELK009CZR	C.elegans cDNA clone y 5.94e+00
	734	15	24.6	417	29	AQ140653	HS_3112_A2_B07_T7 CIT 5.94e+00
	735	15	24.6	417	12	AA763423	vw54b05.r1 Soares mous 5.94e+00
c	736	15	24.6	417	28	AQ168516	HS_3163_A2_H03_MR CIT 5.94e+00
	737	15	24.6	418	12	AA755017	vv75g04.r1 Stratagene 5.94e+00
	738	15	24.6	420	6	CELK038C7R	C.elegans cDNA clone y 5.94e+00
	739	15	24.6	420	27	AQ043894	CIT-HSP-2328C15.TR CIT 5.94e+00
c	740	15	24.6	420	26	AQ018860	CIT-HSP-2309A1.TR CIT- 5.94e+00
c	741	15	24.6	420	12	AA764315	vp02a12.r1 Soares mous 5.94e+00
c	742	15	24.6	421	5	R38295	yc93h10.s1 Homo sapien 5.94e+00
c	743	15	24.6	422	21	W80964	zd83a09.s1 Soares feta 5.94e+00
c	744	15	24.6	422	25	AA348356	EST54821 Hippocampus I 5.94e+00
c	745	15	24.6	422	26	B21621	T26B8-Sp6 TAMU Arabido 5.94e+00
	746	15	24.6	423	15	C78854	Mus musculus 3.5-dpc b 5.94e+00
c	747	15	24.6	423	8	AA234897	zs36c04.s1 Soares NhHM 5.94e+00
	748	15	24.6	425	16	R96599	yq54e11.s1 Homo sapien 5.94e+00
	749	15	24.6	425	15	AA983594	on41d09.s1 NCI_CGAP_Co 5.94e+00
	750	15	24.6	426	8	AA235168	zs36c04.r1 Soares NhHM 5.94e+00
	751	15	24.6	426	11	AA581635	nc85g02.s1 NCI_CGAP_GC 5.94e+00
	752	15	24.6	426	27	AQ072323	HS_3011_A2_E12_T7 CIT 5.94e+00
c	753	15	24.6	426	16	H40637	yp50h02.s1 Homo sapien 5.94e+00
	754	15	24.6	427	29	AF005797	Arabidopsis thaliana 1 5.94e+00
c	755	15	24.6	427	9	AA589703	vl78a12.r1 Stratagene 5.94e+00
c	756	15	24.6	427	19	N35623	yy25h04.s1 Homo sapien 5.94e+00

c	757	15	24.6	427	23	AA250326	mz58d04.r1 Soares mous	5.94e+00
	758	15	24.6	428	15	AA943381	EST198880 Normalized r	5.94e+00
	759	15	24.6	429	11	AA694492	ah33h10.s1 Soares test	5.94e+00
	760	15	24.6	429	19	N57227	yw65f04.r1 Homo sapien	5.94e+00
c	761	15	24.6	431	14	AA908028	EST851 Bovine lactatin	5.94e+00
c	762	15	24.6	431	28	AQ222598	HS_2007_B2_C12_T7 CIT	5.94e+00
c	763	15	24.6	431	11	AA683042	ae89e12.s1 Stratagene	5.94e+00
	764	15	24.6	433	14	AA892571	EST196374 Normalized r	5.94e+00
c	765	15	24.6	433	26	AQ015304	CIT-HSP-2311N4.TF CIT-	5.94e+00
	766	15	24.6	435	27	AQ040216	CIT-HSP-2337N6.TF CIT-	5.94e+00
c	767	15	24.6	435	29	AQ232078	HS_2016_B2_A03_T7 CIT	5.94e+00
	768	15	24.6	436	28	AQ195775	RPCI11-5903.TK RPCI11	5.94e+00
	769	15	24.6	437	15	C80921	Mus musculus 3.5-dpc b	5.94e+00
c	770	15	24.6	437	28	AQ198420	RPCI11-46C3.TK RPCI11	5.94e+00
c	771	15	24.6	438	16	R60164	yh12e09.s1 Homo sapien	5.94e+00
	772	15	24.6	439	19	H84103	ys94h09.r1 Homo sapien	5.94e+00
	773	15	24.6	440	18	AI076222	oy65b09.x1 NCI_CGAP_Cl	5.94e+00
c	774	15	24.6	440	23	T03818	IB941 Infant brain, Be	5.94e+00
c	775	15	24.6	440	8	AA410476	zt30g04.r1 Soares ovar	5.94e+00
	776	15	24.6	441	18	AI145215	UI-R-BT0-pv-g-02-0-UI.	5.94e+00
	777	15	24.6	441	7	AA512594	vj19a08.r1 Barstead mo	5.94e+00
	778	15	24.6	441	17	AU015527	Mus musculus 2-cell em	5.94e+00
	779	15	24.6	442	19	N29389	yy04a09.s1 Homo sapien	5.94e+00
c	780	15	24.6	442	23	AA186164	mt33g06.r1 Soares mous	5.94e+00
c	781	15	24.6	442	23	T03812	IB934 Infant brain, Be	5.94e+00
	782	15	24.6	443	17	AA409190	EST03672 Mouse 7.5 dpc	5.94e+00
c	783	15	24.6	443	11	AA712065	vu60e02.r1 Soares mous	5.94e+00
c	784	15	24.6	444	20	N93895	yy64d12.r1 Homo sapien	5.94e+00
c	785	15	24.6	446	5	R42846	yg05d10.s1 Homo sapien	5.94e+00
	786	15	24.6	447	29	AQ226042	HS_2017_A1_F02_T7 CIT	5.94e+00
c	787	15	24.6	447	8	AA503781	ne51a10.s1 NCI_CGAP_Co	5.94e+00
	788	15	24.6	448	26	B50001	CIT-HSP-653B4.TP CIT-H	5.94e+00
	789	15	24.6	449	28	AQ224576	HS_2012_A1_D02_T7 CIT	5.94e+00
c	790	15	24.6	449	24	AA265629	mz70f06.r1 Soares mous	5.94e+00
c	791	15	24.6	450	28	AQ184604	HS_2212_A1_A09_MR CIT	5.94e+00
c	792	15	24.6	451	21	W83135	mf01f07.r1 Soares mous	5.94e+00
c	793	15	24.6	452	5	R10483	yf37c07.r1 Homo sapien	5.94e+00
	794	15	24.6	453	15	AI047904	uh52h09.r1 Soares mous	5.94e+00
c	795	15	24.6	454	19	N63960	yz81c02.s1 Homo sapien	5.94e+00
c	796	15	24.6	454	20	T15430	IB1153 Infant brain, B	5.94e+00
c	797	15	24.6	454	5	R45492	yg43h07.s1 Homo sapien	5.94e+00
	798	15	24.6	455	15	AA827447	of28e08.s1 NCI_CGAP_Ki	5.94e+00
c	799	15	24.6	455	12	AA763085	vv87a05.r1 Stratagene	5.94e+00
c	800	15	24.6	455	8	AA262517	zs17a09.r1 NCI_CGAP_GC	5.94e+00
c	801	15	24.6	455	7	AA475274	vh20c02.r1 Soares mous	5.94e+00
	802	15	24.6	456	14	AA899974	UI-R-E0-cr-a-04-0-UI.s	5.94e+00
	803	15	24.6	456	7	AA517631	vh81a07.r1 Knowles Sol	5.94e+00
	804	15	24.6	460	15	AA891365	EST195168 Normalized r	5.94e+00
c	805	15	24.6	460	15	C81509	Mus musculus 3.5-dpc b	5.94e+00
c	806	15	24.6	461	22	AA025022	ze80a11.s1 Soares feta	5.94e+00
c	807	15	24.6	461	27	AQ087118	HS_2187_B1_A08_MF CIT	5.94e+00

	808	15	24.6	463	11	N38212	19439 Lambda-PRL2 Arab	5.94e+00
c	809	15	24.6	463	16	R98600	yq72e09.s1 Homo sapien	5.94e+00
	810	15	24.6	464	5	T79019	yd21e09.r1 Homo sapien	5.94e+00
	811	15	24.6	464	27	AQ007337	CIT-HSP-2297N22.TF CIT	5.94e+00
	812	15	24.6	464	12	AA264907	LD09584.5prime LD Dros	5.94e+00
	813	15	24.6	465	17	AI072547	UI-R-C2-ni-f-03-0-UI.s	5.94e+00
c	814	15	24.6	465	27	B86005	RPCI11-21L12.TP RPCI11	5.94e+00
c	815	15	24.6	466	21	AA025461	ze84g12.s1 Soares feta	5.94e+00
c	816	15	24.6	467	16	T28748	EST54665 Homo sapiens	5.94e+00
c	817	15	24.6	467	18	AI185986	qe34d03.s1 Soares_feta	5.94e+00
	818	15	24.6	468	8	AA542960	ni92g08.s1 NCI_CGAP_Pr	5.94e+00
	819	15	24.6	468	18	AI101701	EST210990 Normalized r	5.94e+00
c	820	15	24.6	469	16	H53496	yq86e04.r1 Homo sapien	5.94e+00
c	821	15	24.6	470	14	AA982386	uh07a04.r1 Soares mous	5.94e+00
c	822	15	24.6	472	16	H24362	ym56b06.s1 Homo sapien	5.94e+00
	823	15	24.6	473	9	AA554047	nk95h09.s1 NCI_CGAP_Co	5.94e+00
c	824	15	24.6	473	11	AA677210	zj60f10.s1 Soares feta	5.94e+00
c	825	15	24.6	473	10	AA643679	nq79g06.s1 NCI_CGAP_Co	5.94e+00
c	826	15	24.6	473	12	AA798416	vy06f02.r1 Stratagene	5.94e+00
c	827	15	24.6	473	6	AA028956	zk11f10.s1 Soares preg	5.94e+00
c	828	15	24.6	474	11	AA701877	zi56b12.s1 Soares feta	5.94e+00
c	829	15	24.6	474	5	R05546	yf87b06.r1 Homo sapien	5.94e+00
	830	15	24.6	475	17	AI082668	ox59b06.s1 Soares_NhHM	5.94e+00
c	831	15	24.6	475	21	W75971	zd58c05.r1 Soares feta	5.94e+00
c	832	15	24.6	476	29	AQ252419	HS_2010_B2_E05_T7 CIT	5.94e+00
c	833	15	24.6	476	27	B95249	CIT-HSP-2172L8.TF CIT-	5.94e+00
c	834	15	24.6	476	22	AA184998	mt88b10.r1 Soares mous	5.94e+00
c	835	15	24.6	476	7	AA455430	aa16g06.r1 Soares NhHM	5.94e+00
c	836	15	24.6	476	13	AA621659	af54c01.s1 Soares tota	5.94e+00
c	837	15	24.6	477	23	AA062588	zf68f03.r1 Soares pine	5.94e+00
c	838	15	24.6	478	16	H19974	yn53g11.s1 Homo sapien	5.94e+00
c	839	15	24.6	479	23	AA140361	mq03a08.r1 Soares 2NbM	5.94e+00
c	840	15	24.6	479	5	R33299	yh81f03.r1 Homo sapien	5.94e+00
c	841	15	24.6	480	28	AQ122271	HS_3082_A1_H09_MR CIT	5.94e+00
	842	15	24.6	481	17	AI132215	ue28g12.x1 Sugano mous	5.94e+00
	843	15	24.6	481	13	AA834964	nj49b04.s1 Soares test	5.94e+00
c	844	15	24.6	481	23	AA153847	mq46a03.r1 Soares 2NbM	5.94e+00
c	845	15	24.6	482	12	AA764229	vv45g10.r1 Soares 2NbM	5.94e+00
c	846	15	24.6	483	15	AI046401	ud69h02.y1 Sugano mous	5.94e+00
	847	15	24.6	484	10	AA631812	np77e12.s1 NCI_CGAP_Pr	5.94e+00
c	848	15	24.6	485	22	AA022490	ze70d07.s1 Soares feta	5.94e+00
c	849	15	24.6	486	15	AI029421	UI-R-C0-iw-h-11-0-UI.s	5.94e+00
c	850	15	24.6	487	5	R53532	yg84b03.s1 Homo sapien	5.94e+00
c	851	15	24.6	489	9	AA590879	vm17e03.r1 Knowles Sol	5.94e+00
c	852	15	24.6	489	22	AA164057	mt67g07.r1 Soares mous	5.94e+00
	853	15	24.6	490	18	AI177694	EST221335 Normalized r	5.94e+00
c	854	15	24.6	490	10	AA406220	zu65e10.s1 Soares test	5.94e+00
c	855	15	24.6	490	16	H10074	ym03g07.s1 Homo sapien	5.94e+00
c	856	15	24.6	490	18	AI148628	qc62g08.x1 Soares_plac	5.94e+00
	857	15	24.6	492	13	AA805737	ns43c05.s1 NCI_CGAP_GC	5.94e+00
	858	15	24.6	493	13	AA839892	vw65e02.r1 Stratagene	5.94e+00

	859	15	24.6	496	16	R46230	yg52a06.s1 Homo sapien	5.94e+00
	860	15	24.6	497	23	AA250269	mz61a02.r1 Soares mous	5.94e+00
c	861	15	24.6	497	7	AA462461	vg67g11.r1 Soares mous	5.94e+00
c	862	15	24.6	498	16	H29454	ym32c10.s1 Homo sapien	5.94e+00
	863	15	24.6	500	29	AQ149978	HS_3184_A1_A12_T7 CIT	5.94e+00
c	864	15	24.6	500	5	R24833	yg29e01.r1 Homo sapien	5.94e+00
c	865	15	24.6	500	11	AA722870	zg89f01.s1 Soares feta	5.94e+00
c	866	15	24.6	500	23	AA120034	mp93h10.r1 Soares 2NbM	5.94e+00
c	867	15	24.6	501	15	AI048140	ud67d08.y1 Sugano mous	5.94e+00
	868	15	24.6	502	15	C79294	Mus musculus 3.5-dpc b	5.94e+00
	869	15	24.6	503	7	AA512030	vj42f04.r1 Stratagene	5.94e+00
c	870	15	24.6	503	8	AA495992	zw06e04.s1 Soares NhHM	5.94e+00
	871	15	24.6	504	26	B61333	T22C9TR TAMU Arabidops	5.94e+00
	872	15	24.6	504	22	AA017704	ze39d09.s1 Soares reti	5.94e+00
c	873	15	24.6	505	16	H20818	ym48b03.s1 Homo sapien	5.94e+00
c	874	15	24.6	505	16	R49635	yg61g05.s1 Homo sapien	5.94e+00
	875	15	24.6	506	23	AA260368	vb06f04.r1 Soares mous	5.94e+00
c	876	15	24.6	507	21	AA035967	mi70g03.r1 Soares mous	5.94e+00
c	877	15	24.6	508	14	AA889675	ak54a11.s1 Soares_test	5.94e+00
	878	15	24.6	509	15	AI046038	TENU1074 T. cruzi epim	5.94e+00
	879	15	24.6	509	28	AQ154613	HS_3017_B2_D02_T7 CIT	5.94e+00
c	880	15	24.6	513	5	R42692	yg13d07.s1 Homo sapien	5.94e+00
c	881	15	24.6	516	11	AA713352	32912 CD4-16 Arabidops	5.94e+00
	882	15	24.6	518	29	AQ229859	HS_2027_A2_F02_MR CIT	5.94e+00
	883	15	24.6	519	18	AI103500	EST212780 Normalized r	5.94e+00
c	884	15	24.6	520	15	AI021730	BSBmMFS08BI5K Brugia	5.94e+00
	885	15	24.6	522	13	AA858256	oe43h08.s1 NCI_CGAP_Pr	5.94e+00
	886	15	24.6	524	26	B21550	T25M24-T7 TAMU Arabido	5.94e+00
c	887	15	24.6	526	10	AA422009	zv31c06.s1 Soares ovar	5.94e+00
	888	15	24.6	528	14	C93176	Dictyostelium discoide	5.94e+00
c	889	15	24.6	529	8	AA480802	zx87a12.r1 Soares ovar	5.94e+00
c	890	15	24.6	531	13	AA861398	ak37d02.s1 Soares test	5.94e+00
	891	15	24.6	533	8	AA469056	ne17d10.s1 NCI_CGAP_Co	5.94e+00
	892	15	24.6	535	29	AQ144159	HS_3076_B2_C09_MF CIT	5.94e+00
c	893	15	24.6	537	19	N63542	yy63a07.s1 Homo sapien	5.94e+00
	894	15	24.6	538	14	C92647	Dictyostelium discoide	5.94e+00
c	895	15	24.6	540	22	AA025288	ze74g02.r1 Soares feta	5.94e+00
c	896	15	24.6	543	26	AQ023569	CpGR0488A Cryptosporid	5.94e+00
c	897	15	24.6	545	15	AA997880	UI-R-C0-hu-g-09-0-UI.s	5.94e+00
	898	15	24.6	549	14	AA906196	oj98d10.s1 Soares_NFL_	5.94e+00
	899	15	24.6	549	18	AI177687	EST221328 Normalized r	5.94e+00
c	900	15	24.6	549	23	AA244961	mx13b01.r1 Soares mous	5.94e+00
	901	15	24.6	550	13	AA285345	HTH193 HTCDL1 Homo sap	5.94e+00
c	902	15	24.6	550	7	AA516958	vh85d10.r1 Knowles Sol	5.94e+00
c	903	15	24.6	550	7	AA494874	fa12c12.r1 Zebrafish I	5.94e+00
c	904	15	24.6	551	23	AA178441	mt16e07.r1 Soares mous	5.94e+00
c	905	15	24.6	551	28	AQ110207	CIT-HSP-2176E7.TR CIT-	5.94e+00
c	906	15	24.6	553	17	AI118215	ue34g02.x1 Sugano mous	5.94e+00
c	907	15	24.6	555	27	AQ043519	CIT-HSP-2318F14.TF CIT	5.94e+00
	908	15	24.6	557	26	B67197	CpG0014B CpIOWAgDNA1 C	5.94e+00
c	909	15	24.6	559	29	AQ238996	RPCI11-70C15.TJ RPCI11	5.94e+00

c	910	15	24.6	559	23	AA178491	mt16b06.r1 Soares mous	5.94e+00
	911	15	24.6	562	18	AI178668	EST222349 Normalized r	5.94e+00
	912	15	24.6	563	15	C78864	Mus musculus 3.5-dpc b	5.94e+00
	913	15	24.6	564	15	C77919	Mus musculus 3.5-dpc b	5.94e+00
	914	15	24.6	565	10	AA617809	nq04g07.s1 NCI_CGAP_Lu	5.94e+00
	915	15	24.6	567	26	FR0026411	F.rubripes GSS sequenc	5.94e+00
c	916	15	24.6	568	21	W72528	zd63a10.s1 Soares feta	5.94e+00
c	917	15	24.6	570	21	AA041934	mj04b06.r1 Soares mous	5.94e+00
	918	15	24.6	574	7	AA520853	TgESTzz64h01.r1 TgME49	5.94e+00
	919	15	24.6	576	18	AI123171	qa86e01.s1 Soares_feta	5.94e+00
	920	15	24.6	577	15	AI007820	EST202271 Normalized r	5.94e+00
c	921	15	24.6	577	27	B94017	CIT-HSP-2166F3.TF CIT-	5.94e+00
c	922	15	24.6	579	15	C78710	Mus musculus 3.5-dpc b	5.94e+00
	923	15	24.6	582	17	AU020966	Mus musculus 8-cell em	5.94e+00
	924	15	24.6	582	17	AU023604	Mus musculus unfertili	5.94e+00
c	925	15	24.6	582	21	AA057466	zf56g05.r1 Soares reti	5.94e+00
	926	15	24.6	583	17	AU019905	Mus musculus 8-cell em	5.94e+00
	927	15	24.6	584	17	AU020798	Mus musculus 8-cell em	5.94e+00
	928	15	24.6	584	12	AA817820	UI-R-A0-ae-d-05-0-UI.s	5.94e+00
	929	15	24.6	585	27	AQ049439	cLM-14e10-u cLM Giardi	5.94e+00
	930	15	24.6	586	18	AI078158	oz30c10.x1 Soares_tota	5.94e+00
c	931	15	24.6	588	17	AU021587	Mus musculus 8-cell em	5.94e+00
c	932	15	24.6	588	26	B48390	RPCI11-113.TVB RPCI11	5.94e+00
	933	15	24.6	589	15	C78195	Mus musculus 3.5-dpc b	5.94e+00
c	934	15	24.6	589	13	AA625206	af70e10.r1 Soares NhHM	5.94e+00
c	935	15	24.6	592	27	B90993	CIT-HSP-2163P16.TR CIT	5.94e+00
c	936	15	24.6	592	10	AA166778	so85c07.r1 Stratagene	5.94e+00
	937	15	24.6	594	28	AQ160945	nbxb0006A04r CUGI Rice	5.94e+00
	938	15	24.6	596	26	FR0024234	F.rubripes GSS sequenc	5.94e+00
	939	15	24.6	597	17	AU018632	Mus musculus 8-cell em	5.94e+00
	940	15	24.6	602	26	B62745	T21K13TF.1 TAMU Arabid	5.94e+00
	941	15	24.6	606	27	B90364	CIT-HSP-2167C7.TF CIT-	5.94e+00
c	942	15	24.6	606	26	FR0020999	F.rubripes GSS sequenc	5.94e+00
	943	15	24.6	607	27	AQ079940	CIT-HSP-2367P19.TF CIT	5.94e+00
c	944	15	24.6	607	17	AI119329	uf03c07.y1 Sugano mous	5.94e+00
c	945	15	24.6	608	17	AI117095	ue30f03.y1 Sugano mous	5.94e+00
	946	15	24.6	612	7	AA511281	vh78h04.r1 Knowles Sol	5.94e+00
	947	15	24.6	615	15	AA946348	EST201847 Normalized r	5.94e+00
	948	15	24.6	615	28	AQ109300	CIT-HSP-2380P9.TR CIT-	5.94e+00
	949	15	24.6	617	18	AI102694	EST211983 Normalized r	5.94e+00
	950	15	24.6	618	27	AQ057229	CIT-HSP-2336C23.TF CIT	5.94e+00
	951	15	24.6	618	26	FR0002793	F.rubripes GSS sequenc	5.94e+00
	952	15	24.6	619	26	FR0013249	F.rubripes GSS sequenc	5.94e+00
	953	15	24.6	619	26	FR0005075	F.rubripes GSS sequenc	5.94e+00
	954	15	24.6	619	26	FR0004097	F.rubripes GSS sequenc	5.94e+00
c	955	15	24.6	619	26	FR0017887	F.rubripes GSS sequenc	5.94e+00
	956	15	24.6	620	28	AQ174181	HS_3206_A2_G04_T7 CIT	5.94e+00
	957	15	24.6	623	15	C93967	Dictyostelium discoide	5.94e+00
	958	15	24.6	630	9	AA582014	nn44b03.s1 NCI_CGAP_GC	5.94e+00
c	959	15	24.6	632	14	AA945159	EST200658 Normalized r	5.94e+00
	960	15	24.6	640	27	B70587	CIT-HSP-2059N13.TF CIT	5.94e+00

	961	15	24.6	642	15	AI011719	EST206170 Normalized r 5.94e+00
c	962	15	24.6	646	15	AI053032	BSBmMFSZ01016SK Brugia 5.94e+00
	963	15	24.6	651	26	B52716	CIT-HSP-2006J5.TR CIT- 5.94e+00
c	964	15	24.6	652	20	W20042	zb39g07.r1 Soares para 5.94e+00
	965	15	24.6	653	26	B62713	T19J16TR.1 TAMU Arabid 5.94e+00
	966	15	24.6	655	14	C89775	Dictyostelium discoide 5.94e+00
c	967	15	24.6	655	26	B52529	CIT-HSP-2005J3.TR CIT- 5.94e+00
	968	15	24.6	660	18	AI172057	EST218052 Normalized r 5.94e+00
c	969	15	24.6	666	18	AI176856	EST220459 Normalized r 5.94e+00
c	970	15	24.6	678	27	AQ048490	cLM-8a2-t cLM Giardia 5.94e+00
c	971	15	24.6	680	29	AQ239830	RPCI11-71A12.TJ RPCI11 5.94e+00
c	972	15	24.6	683	27	AQ041466	CIT-HSP-2317D7.TR CIT- 5.94e+00
	973	15	24.6	687	26	B72165	RPCI11-7J20.TP RPCI11 5.94e+00
	974	15	24.6	691	10	AA577667	EST241 Sugarcane leaf 5.94e+00
	975	15	24.6	694	27	AG012337	Homo sapiens genomic D 5.94e+00
c	976	15	24.6	700	26	B18860	P19I13-Sp6 IGF Arabido 5.94e+00
c	977	15	24.6	704	27	AG011528	Homo sapiens genomic D 5.94e+00
	978	15	24.6	709	11	C84127	Dictyostelium discoide 5.94e+00
c	979	15	24.6	711	27	AQ049232	cLM-13a4-t cLM Giardia 5.94e+00
c	980	15	24.6	725	26	AG005672	Homo sapiens genomic D 5.94e+00
	981	15	24.6	726	27	AQ051324	nbxb0001dA10f CUGI Ric 5.94e+00
	982	15	24.6	732	14	C92828	Dictyostelium discoide 5.94e+00
c	983	15	24.6	739	10	AA658753	fa66g10.s1 zebrafish f 5.94e+00
c	984	15	24.6	750	28	AF010924	Homo sapiens chromosom 5.94e+00
c	985	15	24.6	752	17	AI116381	uf02d04.y1 Sugano mous 5.94e+00
	986	15	24.6	760	26	AQ013041	CIT-HSP-3298G11.TR CIT 5.94e+00
c	987	15	24.6	796	27	AQ048221	cLM-6d1-t cLM Giardia 5.94e+00
c	988	15	24.6	803	17	AI099212	ue40d08.y1 Sugano mous 5.94e+00
c	989	15	24.6	808	26	B18870	F21J19-Sp6 IGF Arabido 5.94e+00
c	990	15	24.6	829	26	B18963	T22D3-Sp6 TAMU Arabido 5.94e+00
c	991	15	24.6	847	26	B22001	F24J11-Sp6 IGF Arabido 5.94e+00
	992	15	24.6	856	27	AQ050999	nbxb0004dF03f CUGI Ric 5.94e+00
c	993	15	24.6	907	26	B12184	F5O23-T7 IGF Arabidops 5.94e+00
c	994	15	24.6	915	26	B13268	T1F7-Sp6.1 TAMU Arabid 5.94e+00
	995	15	24.6	920	7	C25711	Dictyostelium discoide 5.94e+00
	996	15	24.6	1036	26	B12918	T4A11-Sp6 TAMU Arabido 5.94e+00
c	997	15	24.6	1095	26	AF029546	Salmonella typhimurium 5.94e+00
c	998	15	24.6	1198	26	B08337	T19F9-Sp6.1 TAMU Arabi 5.94e+00
c	999	15	24.6	1205	26	B08894	F10C6-T7 IGF Arabidops 5.94e+00
	1000	15	24.6	1216	29	AF005801	Arabidopsis thaliana 1 5.94e+00

ALIGNMENTS

```
RESULT    1
LOCUS       HUMGS01603      62 bp    mRNA            EST       17-JUN-1996
DEFINITION  Human HL60 3'directed MboI cDNA, HUMGS01603, clone pm2646, mRNA
            sequence.
ACCESSION   D20628
NID         g501724
KEYWORDS    EST; EST(expressed sequence tag); HL60; gene signature(GS);
```



```
            granulocyte; macrophage; monocyte; promyelocyte.
SOURCE      Homo sapiens adult female promyelocyte cell_line:HL60 cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (sites)
  AUTHORS   Okubo,K., Itoh,K., Fukushima,A., Yoshii,J. and Matsubara,K.
  TITLE     Monitoring cell physiology by expression profiles and discovering
            cell type-specific genes by compiled expression profiles
  JOURNAL   Genomics 30 (2), 178-186 (1995)
  MEDLINE   96163870
REFERENCE   2  (bases 1 to 62)
  AUTHORS   Okubo,K., Fukushima,A., Yoshii,J., Niiyama,T., Kojima,Y.,
            Yoshinari,H., Arimoto,J. and Matsubara,K.
  TITLE     Gene expression of human promyelocytic cell line HL60before and
            after induction of differentiation. A new application of
            3'directedcDNA sequencing
  JOURNAL   Unpublished (1993)
REFERENCE   3  (bases 1 to 62)
  AUTHORS   Okubo,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-SEP-1993) to the DDBJ/EMBL/GenBank databases. Kousaku
            Okubo, Osaka University, Institute for Molecular and Cellular Bio;
            1-3,Yamada-oka, Suita, Osaka Pref. 565, Japan
            (E-mail:kousaku@imcb.osaka-u.ac.jp, Tel:06-877-5111(ex.3315),
            Fax:06-877-1922)
COMMENT     Submitted (10-Sep-1993) to DDBJ by:
            Kousaku Okubo
            Institute for Molecular and Cellular Biology
            Osaka University
            3-1, Yamadaoka
            Suita, Osaka, 565
            Japan
            Phone: 06-877-5111
            Fax  : 06-875-1922.
FEATURES              Location/Qualifiers
     source           1. .62
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /cell_line="HL60"
                      /cell_type="promyelocyte"
                      /dev_stage="adult"
                      /sex="female"
BASE COUNT       30 a      11 c       9 g      11 t       1 others
ORIGIN

  Query Match          100.0%;  Score 61;  DB 7;  Length 62;
  Best Local Similarity 100.0%;  Pred. No. 9.07e-72;
  Matches   62;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```



```
Db       1 GATCCCCATTCTTTTCAAAAACAAAATAAAACAATAAAGACTGCAAGGAAGACTGANGGA 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy       1 GATCCCCATTCTTTTCAAAAACAAAATAAAACAATAAAGACTGCAAGGAAGACTGANGGA 60

Db      61 AA 62
           ||
Qy      61 AA 62


RESULT    2
LOCUS       HUMGS01613      58 bp    mRNA            EST       17-JUN-1996
DEFINITION  Human HL60 3'directed MboI cDNA, HUMGS01613, clone pm1209, mRNA
            sequence.
ACCESSION   D20638
NID         g501734
KEYWORDS    EST; EST(expressed sequence tag); HL60; gene signature(GS);
            granulocyte; macrophage; monocyte; promyelocyte.
SOURCE      Homo sapiens adult female promyelocyte cell_line:HL60 cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (sites)
  AUTHORS   Okubo,K., Itoh,K., Fukushima,A., Yoshii,J. and Matsubara,K.
  TITLE     Monitoring cell physiology by expression profiles and discovering
            cell type-specific genes by compiled expression profiles
  JOURNAL   Genomics 30 (2), 178-186 (1995)
  MEDLINE   96163870
REFERENCE   2  (bases 1 to 58)
  AUTHORS   Okubo,K., Fukushima,A., Yoshii,J., Niiyama,T., Kojima,Y.,
            Yoshinari,H., Arimoto,J. and Matsubara,K.
  TITLE     Gene expression of human promyelocytic cell line HL60before and
            after induction of differentiation. A new application of
            3'directedcDNA sequencing
  JOURNAL   Unpublished (1993)
REFERENCE   3  (bases 1 to 58)
  AUTHORS   Okubo,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-SEP-1993) to the DDBJ/EMBL/GenBank databases. Kousaku
            Okubo, Osaka University, Institute for Molecular and Cellular Bio;
            1-3,Yamada-oka, Suita, Osaka Pref. 565, Japan
            (E-mail:kousaku@imcb.osaka-u.ac.jp, Tel:06-877-5111(ex.3315),
            Fax:06-877-1922)
COMMENT     Submitted (10-Sep-1993) to DDBJ by:
            Kousaku Okubo
            Institute for Molecular and Cellular Biology
            Osaka University
            3-1, Yamadaoka
            Suita, Osaka, 565
            Japan
            Phone: 06-877-5111
```



```
            Fax  : 06-875-1922.
FEATURES              Location/Qualifiers
     source           1. .58
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /cell_line="HL60"
                      /cell_type="promyelocyte"
                      /dev_stage="adult"
                      /sex="female"
BASE COUNT       30 a       9 c       8 g       9 t       2 others
ORIGIN

  Query Match          50.8%;  Score 31;  DB 7;  Length 58;
  Best Local Similarity 100.0%;  Pred. No. 1.37e-21;
  Matches   31;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      14 AAAAACAAAATAAAACAATAAAGACTGCAAG 44
           |||||||||||||||||||||||||||||||
Qy      17 AAAAACAAAATAAAACAATAAAGACTGCAAG 47


RESULT    3
LOCUS       AA844688       407 bp    mRNA            EST       04-MAR-1998
DEFINITION  ai70b09.s1 Soares testis NHT Homo sapiens cDNA clone IMAGE:1376153
            3', mRNA sequence.
ACCESSION   AA844688
NID         g2931139
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 407)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 350.
```

```
FEATURES             Location/Qualifiers
     source          1. .407
                     /organism="Homo sapiens"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from mRNA obtained from Clontech
                     Laboratories, Inc., and primed with  a Not I - oligo(dT)
                     primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3'].
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization to Cot5, and was
                     constructed  by Bento Soares and M. Fatima Bonaldo."
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1376153"
                     /clone_lib="Soares testis NHT"
                     /sex="male"
                     /lab_host="DH10B"
BASE COUNT      150 a      57 c      51 g     149 t
ORIGIN

  Query Match           36.1%;  Score 22;  DB 13;  Length 407;
  Best Local Similarity 100.0%;  Pred. No. 1.64e-08;
  Matches    22;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    137 AAAAACAAAATAAAACAATAAA 158
          ||||||||||||||||||||||
Qy     17 AAAAACAAAATAAAACAATAAA 38


RESULT    4
LOCUS       AA848939       390 bp    mRNA            EST       30-APR-1998
DEFINITION  EST191701 Normalized rat lung, Bento Soares Rattus sp. cDNA clone
            RLUAI16 3' end, mRNA sequence.
ACCESSION   AA848939
NID         g2936479
KEYWORDS    EST.
SOURCE      Rattus sp.
  ORGANISM  Rattus sp.
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 390)
  AUTHORS   Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
            Kerlavage,A.R. and Adams,M.D.
  TITLE     Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
            Gene Index
  JOURNAL   Unpublished (1998)
COMMENT
            Contact: Lee, NH
            ATCC
```



```
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
            Seq primer: M13-21.
FEATURES             Location/Qualifiers
     source          1. .390
                     /organism="Rattus sp."
                     /note="Organ: lung; Vector: pT7T3Pac; Site_1: EcoRI;
                     Site_2: NotI"
                     /db_xref="ATCC (inhost):2007967"
                     /db_xref="taxon:10118"
                     /clone="RLUAI16"
                     /clone_lib="Normalized rat lung, Bento Soares"
BASE COUNT       89 a      71 c      90 g     140 t
ORIGIN

  Query Match           32.8%;  Score 20;  DB 14;  Length 390;
  Best Local Similarity 100.0%;  Pred. No. 7.15e-06;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    242 GTTTTATTTTGTTTTTGAAA 261
          ||||||||||||||||||||
Cp     32 GTTTTATTTTGTTTTTGAAA 13


RESULT    5
LOCUS       AQ220224       406 bp    DNA             GSS       18-SEP-1998
DEFINITION  HS_3248_B1_A07_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3248 Col=13 Row=B, genomic survey
            sequence.
ACCESSION   AQ220224
NID         g3633837
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 406)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Furlong,J.,
            Shaker,R., Schmidt,S., Traicoff,R. and Hood,L.E.
  TITLE     Construction of a Characterized Clone Resource for Genomic
            Sequencing
  JOURNAL   Unpublished (1998)
COMMENT
            Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
```



```
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3248  row: B  column: 13
            Class: BAC ends
            High quality sequence stop: 406.
FEATURES             Location/Qualifiers
     source          1. .406
                     /organism="Homo sapiens"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
                     /db_xref="taxon:9606"
                     /clone="Plate=3248 Col=13 Row=B"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
BASE COUNT      133 a      56 c      73 g     144 t
ORIGIN

  Query Match           32.8%;  Score 20;  DB 28;  Length 406;
  Best Local Similarity 100.0%;  Pred. No. 7.15e-06;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    144 TATTTTGTTTTTGAAAAGAA 163
          ||||||||||||||||||||
Cp     28 TATTTTGTTTTTGAAAAGAA 9


RESULT    6
LOCUS       N32666         415 bp    mRNA            EST       10-JAN-1996
DEFINITION  yx77c12.r1 Homo sapiens cDNA clone 267766 5'.
ACCESSION   N32666
NID         g1153065
KEYWORDS    EST.
SOURCE      human clone=267766 primer=T7 library=Soares melanocyte 2NbHM
            vector=pT7T3D (Pharmacia) with a modified polylinker host=DH10B
            (ampicillin resistant) Raitel=Not I Raite2=Eco RI Male. 1st strand
            cDNA was primed with a Not I - oligo(dT) primer
            [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCAGTTTTTTTTTTTTTTTTTT-3'],
            double-stranded cDNA was size selected, ligated to Eco RI adapters
            (Pharmacia), digested with Not I and cloned into the Not I and Eco
            RI sites of a modified pT7T3 vector (Pharmacia). Library
            constructed by Bento Soares and M.Fatima Bonaldo. RNA from normal
            foreskin melanocytes (FS374) was kindly provided by Dr. Anthony P.
            Albino.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 415)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
```



```
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 349
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES             Location/Qualifiers
     source          1. .415
                     /organism="Homo sapiens"
                     /clone="267766"
     mRNA            <1. .>415
BASE COUNT       87 a      68 c      85 g     174 t      1 others
ORIGIN

  Query Match           32.8%;  Score 20;  DB 19;  Length 415;
  Best Local Similarity 100.0%;  Pred. No. 7.15e-06;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    264 TTGTTTTATTTTGTTTTTGA 283
          ||||||||||||||||||||
Cp     34 TTGTTTTATTTTGTTTTTGA 15


RESULT    7
LOCUS       AI170294       419 bp    mRNA            EST       07-OCT-1998
DEFINITION  EST216220 Normalized rat lung, Bento Soares Rattus sp. cDNA clone
            RLUCG59 3' end, mRNA sequence.
ACCESSION   AI170294
NID         g3710334
KEYWORDS    EST.
SOURCE      Rattus sp.
  ORGANISM  Rattus sp.
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 419)
  AUTHORS   Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
            Kerlavage,A.R. and Adams,M.D.
  TITLE     Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
            Gene Index
```

```
  JOURNAL   Unpublished (1998)
COMMENT     Other_ESTs: TC49104
            Contact: Lee, NH
            ATCC
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
            Seq primer: M13-21.
FEATURES              Location/Qualifiers
     source           1..419
                      /organism="Rattus sp."
                      /note="Organ: lung; Vector: pT7T3Pac; Site_1: EcoRI;
                      Site_2: NotI"
                      /db_xref="taxon:10118"
                      /clone="RLUCG59"
                      /clone_lib="Normalized rat lung, Bento Soares"
BASE COUNT       94 a     82 c     97 g    146 t
ORIGIN

  Query Match           32.8%;  Score 20;  DB 18;  Length 419;
  Best Local Similarity 100.0%;  Pred. No. 7.15e-06;
  Matches   20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    242 GTTTTATTTTGTTTTTGAAA 261
          ||||||||||||||||||||
Cp     32 GTTTTATTTTGTTTTTGAAA 13


RESULT    8
LOCUS       AA926956      503 bp    mRNA            EST       17-JUN-1998
DEFINITION  om68h05.s1 NCI_CGAP_GC4 Homo sapiens cDNA clone IMAGE:1552377 3'
            similar to contains L1.t3 L1 repetitive element ;, mRNA sequence.
ACCESSION   AA926956
NID         g3075853
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 503)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
```



```
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 1432   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 448.
FEATURES              Location/Qualifiers
     source           1..503
                      /organism="Homo sapiens"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; 1st strand cDNA was prepared from 3 pooled
                      germ cell tumors, and was then primed with a Not I -
                      oligo(dT) primer. Double-stranded cDNA was ligated to Eco
                      RI adaptors (Pharmacia), digested with Not I and cloned
                      into the Not I and Eco RI sites of the modified pT7T3
                      vector.  Library is normalized. Library was constructed by
                      Bento Soares and M. Fatima Bonaldo."
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1552377"
                      /clone_lib="NCI_CGAP_GC4"
                      /tissue_type="pooled germ cell tumors"
                      /lab_host="DH10B"
BASE COUNT      230 a     73 c     59 g    141 t
ORIGIN

  Query Match           32.8%;  Score 20;  DB 15;  Length 503;
  Best Local Similarity 100.0%;  Pred. No. 7.15e-06;
  Matches   20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     14 ATTGTTTTATTTTGTTTTTG 33
          ||||||||||||||||||||
Cp     35 ATTGTTTTATTTTGTTTTTG 16


RESULT    9
LOCUS       AA747933      642 bp    mRNA            EST       18-FEB-1998
DEFINITION  nx97h10.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1270243
            similar to TR:Q12865 Q12865 NF-ATC. ;, mRNA sequence.
ACCESSION   AA747933
NID         g2787891
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 642)
```



```
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
            Ph.D., Gerald Marti, M.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 2448   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 61.
FEATURES              Location/Qualifiers
     source           1..642
                      /organism="Homo sapiens"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                      was prepared from human tonsillar cells enriched for
                      germinal center B cells by flow sorting (CD20+, IgD-),
                      provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                      (NCI) and Dr. Gerald Marti (CBER).  cDNA synthesis was
                      primed with a Not I - oligo(dT) primer
                      [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTT-
                      3']. Double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Not I and cloned into the Not I
                      and Eco RI sites of the modified pT7T3 vector. Library
                      went through one round of normalization, and was
                      constructed by Bento Soares and M. Fatima Bonaldo."
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1270243"
                      /clone_lib="NCI_CGAP_GCB1"
                      /tissue_type="germinal center B cell"
                      /lab_host="DH10B"
BASE COUNT      126 a    146 c    195 g    174 t      1 others
ORIGIN

  Query Match           32.8%;  Score 20;  DB 13;  Length 642;
  Best Local Similarity 100.0%;  Pred. No. 7.15e-06;
  Matches   20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    122 ATTTTGTTTTTGAAAAGAAT 141
          ||||||||||||||||||||
```



```
Cp     27 ATTTTGTTTTTGAAAAGAAT 8


RESULT   10
LOCUS       AG002485      716 bp    DNA             GSS       21-NOV-1997
DEFINITION  Homo sapiens genomic DNA, 21q region, clone: 130N6A30, genomic
            survey sequence.
ACCESSION   AG002485
NID         g2641782
KEYWORDS    GSS.
SOURCE      Homo sapiens DNA, clone:130N6A30.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 716)
  AUTHORS   Hattori,M., Ishii,K., Shiba,T. and Sakaki,Y.
  TITLE     Homo sapiens genomic DNA, chromosome 21q
  JOURNAL   Published Only in DataBase (1997) In press
REFERENCE   2  (bases 1 to 716)
  AUTHORS   Hattori,M., Ishii,K., Shiba,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-NOV-1997) to the DDBJ/EMBL/GenBank databases.
            Masahira Hattori, Kitasato University, Department of Science, JST
            Sequencing Laboratory; Kitasato 1-15-1, Sagamihara 228, Japan
            (E-mail:hattori@hgc.ims.u-tokyo.ac.jp, Tel:0427-78-9732,
            Fax:0427-78-9732)
FEATURES              Location/Qualifiers
     source           1..716
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="21"
                      /clone="130N6A30"
                      /map="21q"
BASE COUNT      162 a    193 c    155 g    204 t      2 others
ORIGIN

  Query Match           32.8%;  Score 20;  DB 26;  Length 716;
  Best Local Similarity 100.0%;  Pred. No. 7.15e-06;
  Matches   20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    147 TTGTTTTATTTTGTTTTTGA 166
          ||||||||||||||||||||
Cp     34 TTGTTTTATTTTGTTTTTGA 15


RESULT   11
LOCUS       AA904400      136 bp    mRNA            EST       09-JUN-1998
DEFINITION  ok07b09.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
            IMAGE:1507097 3', mRNA sequence.
ACCESSION   AA904400
NID         g3039523
```

```
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 136)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 566  Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 113.
FEATURES            Location/Qualifiers
     source         1. .136
                    /organism="Homo sapiens"
                    /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                    a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                    Equal amounts of plasmid DNA from three normalized
                    libraries (fetal lung NbHL19W, testis NHT, and B-cell
                    NCI_CGAP_GCB1) were mixed, and ss circles were made in
                    vitro. Following HAP purification, this DNA was used as
                    tracer in a subtractive hybridization reaction. The driver
                    was PCR-amplified cDNAs from pools of 5,000 clones made
                    from the same 3 libraries. The pools consisted of
                    I.M.A.G.E. clones 297480-302087, 682632-687239,
                    726408-728711, and 729096-731399. Subtraction by Bento
                    Soares and M. Fatima Bonaldo."
                    /db_xref="taxon:9606"
                    /clone="IMAGE:1507097"
                    /clone_lib="Soares_NFL_T_GBC_S1"
                    /lab_host="DH10B"
BASE COUNT       64 a     13 c     12 g     47 t
ORIGIN

  Query Match           31.1%;  Score 19;  DB 15;  Length 136;
  Best Local Similarity 100.0%;  Pred. No. 1.33e-04;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    103 AAAAACAAAATAAAACAAT 121
          |||||||||||||||||||
Qy     17 AAAAACAAAATAAAACAAT 35


RESULT   12
```



```
LOCUS       HSC05B102     229 bp    RNA             EST       21-SEP-1995
DEFINITION  H. sapiens partial cDNA sequence; clone c-05b10, mRNA sequence.
ACCESSION   Z38278
NID         g560286
KEYWORDS    EST; partial cDNA sequence; transcribed sequence fragment.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 229)
  AUTHORS   Genexpress.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-OCT-1994) Genethon, B.P. 60, 91002 Evry Cedex France
            and Genetique Moleculaire et Biologie du developpement, CNRS UPR420
            B.P. 8, 94801 Villejuif Cedex France.E-mail: genexpress@genethon.fr
REFERENCE   2  (bases 1 to 229)
  AUTHORS   Genexpress.
  TITLE     The Genexpress cDNA program
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 229)
  AUTHORS   Auffray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C.,
            Devignes,M.D., Duprat,S., Houlgatte,R., Jumeau,M.N., Lamy,B.,
            Lorenzo,F., Mitchell,H., Mariage-Samson,R., Pietu,G., Pouliot,Y.,
            Sebastiani-Kabaktchis,C. and Tessier,A.
  TITLE     IMAGE: molecular integration of the analysis of the human genome
            and its expression
  JOURNAL   C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
  MEDLINE   95277534
COMMENT     Clone library from B.Soares, Psychiatry Dept. Columbia Universaity
            USA;

            Cloning_method: total mRNA was oligo-(dT) primed and directionally
            cloned 5' -> 3' into the HindIII -> NotI sites of the lafmid BA
            vector;
            Sequencing_method: single read, full automatic;
            Primer: (-21)M13_universal;
            cDNA sequence complementary to mRNA (3'end)
            Stretch_removed: 18 T removed at sequence 5'end
            Normalization_method: Bento Soares, P.N.A.S in press;
            Genexpress_library_idt: C;
            Genexpress_sequence_idt: alc-05b10;

            No significant homology found with :
            genbank release 81 swissprot release 28.
FEATURES            Location/Qualifiers
     source         1. .229
                    /organism="Homo sapiens"
                    /isolate="muscular atrophy patient"
                    /db_xref="taxon:9606"
                    /dev_stage="3 months old"
                    /tissue_type="total brain"
```



```
                    /clone_lib="normalized infant brain cDNA"
                    /sex="Female"
BASE COUNT      100 a     32 c     29 g     66 t      2 others
ORIGIN

  Query Match           31.1%;  Score 19;  DB 20;  Length 229;
  Best Local Similarity 100.0%;  Pred. No. 1.33e-04;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     73 AAAAACAAAATAAAACAAT 91
          |||||||||||||||||||
Qy     17 AAAAACAAAATAAAACAAT 35


RESULT   13
LOCUS       R38922       259 bp    mRNA            EST       04-MAY-1995
DEFINITION  yd06a09.s1 Homo sapiens cDNA clone 24834 3'.
ACCESSION   R38922
NID         g796378
KEYWORDS    EST.
SOURCE      human clone=24834 library=Soares infant brain 1NIB vector=Lafmid BA
            host=DH10B (ampicillin resistant) primer=-21m13 Rsitel=Not I
            Rsite2=Hind III Whole brain from a 73 days post natal female. 1st
            strand cDNA was primed with a Not I - oligo(dT) primer [5'
            AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTT 3']; double-stranded
            cDNA was ligated to Hind III adaptors (Pharmacia), digested with
            Not I and directionally cloned into the Not I and Hind III sites of
            the Lafmid BA vector. Library went through one round of
            normalization. Library constructed by Bento Soares and M.Fatima
            Bonaldo.
  ORGANISM  Homo sapiens
            Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 259)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            GDB: G00-397-181
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 144
```



```
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES            Location/Qualifiers
     source         1. .259
                    /organism="Homo sapiens"
                    /clone="24834"
BASE COUNT      109 a     31 c     30 g     87 t      2 others
ORIGIN

  Query Match           31.1%;  Score 19;  DB 5;  Length 259;
  Best Local Similarity 100.0%;  Pred. No. 1.33e-04;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    111 AAAAACAAAATAAAACAAT 129
          |||||||||||||||||||
Qy     17 AAAAACAAAATAAAACAAT 35


RESULT   14
LOCUS       AA899734      294 bp    mRNA            EST       07-APR-1998
DEFINITION  UI-R-E0-dg-d-09-0-UI.s1 UI-R-E0 Rattus norvegicus cDNA clone
            UI-R-E0-dg-d-09-0-UI 3' similar to >
            gi|2947983|gb|AA859452|AA859452 UI-R-A0-bf-b-05-0-UI.s1 UI-R-A0
            Rattus norvegicus cDNA clone UI-R-A0-bf-b-05-0-UI 3' similar to
            gb|AA465962|AA465962 ve90b04.s1 Knowles Solter mouse 2 cell Mus
            musculus cDNA clone 833455 5', mRNA sequence.
ACCESSION   AA899734
NID         g3035088
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 294)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT
            Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            cDNA Library Preparation: M. Fatima Bonaldo, Ph.D. Clone
            distribution: clones will be available through Research Genetics
            Seq primer: M13 Forward.
```

```
FEATURES             Location/Qualifiers
     source          1. .294
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: NotI; Site_2: EcoRI; This library
                     consists of a mixture of individually tagged normalized
                     libraries constructed from 8, 12 and 18-day embryo.  The
                     tag is a string  of 3-5 nucleotides present between the
                     Not I site and the oligo-dT track  which allows
                     identification of the library of origin of a clone within
                     the mixture."
                     /db_xref="taxon:10116"
                     /clone="UI-R-E0-dg-d-09-0-UI"
                     /clone_lib="UI-R-E0"
                     /dev_stage="embryonic"
                     /lab_host="DH10B (Life Technologies)"
BASE COUNT       125 a     49 c     40 g     80 t
ORIGIN

  Query Match           31.1%;  Score 19;  DB 14;  Length 294;
  Best Local Similarity 100.0%;  Pred. No. 1.33e-04;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      95 AAAAACAAAATAAAACAAT 113
           ||||||||||||||||||
Qy      17 AAAAACAAAATAAAACAAT 35


RESULT   15
LOCUS       AA878981       316 bp    mRNA            EST       09-JUN-1998
DEFINITION  oj26e09.s1 NCI_CGAP_Kid3 Homo sapiens cDNA clone IMAGE:1493320 3',
            mRNA sequence.
ACCESSION   AA878981
NID         g2987946
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 316)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
```



```
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 806   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 230.
FEATURES             Location/Qualifiers
     source          1. .316
                     /organism="Homo sapiens"
                     /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer,
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia),  digested with Not I and cloned into the Not
                     I and Eco RI sites  of the modified pT7T3 vector.  mRNA
                     source: 2 pooled kidneys. Library went through one round
                     of normalization. Library constructed by Bento Soares and
                     M. Fatima Bonaldo."
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1493320"
                     /clone_lib="NCI_CGAP_Kid3"
                     /lab_host="DH10B"
BASE COUNT       136 a     41 c     42 g     97 t
ORIGIN

  Query Match           31.1%;  Score 19;  DB 15;  Length 316;
  Best Local Similarity 100.0%;  Pred. No. 1.33e-04;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      94 AAAAACAAAATAAAACAAT 112
           ||||||||||||||||||
Qy      17 AAAAACAAAATAAAACAAT 35


RESULT   16
LOCUS       AA465962       328 bp    mRNA            EST       04-AUG-1997
DEFINITION  ve90b04.s1 Knowles Solter mouse 2 cell Mus musculus cDNA clone
            833455 5', mRNA sequence.
ACCESSION   AA465962
NID         g2192102
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae;
            Murinae; Mus.
REFERENCE   1  (bases 1 to 328)
```



```
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:493671
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 306.
FEATURES             Location/Qualifiers
     source          1. .328
                     /organism="Mus musculus"
                     /strain="B6D2 F1/J"
                     /note="Organ: embryo; Vector: pBluescribe (modified);
                     Site_1: MluI; Site_2: SalI; Cloned unidirectionally from
                     mRNA prepared from 13,500 2-cell stage  embryos.  Primer:
                     SalI(dT): 5'-CGGTCGACCGTCGACCGTTTTTTTTTTTTTTT-3'.  cDNAs
                     were cloned into the MluI/SalI sites of a modified
                     pBluescribe  vector using commercial linkers (NEB).
                     Average insert size:  1.2 kb."
                     /db_xref="taxon:10090"
                     /clone="833455"
                     /clone_lib="Knowles Solter mouse 2 cell"
                     /tissue_type="embryo"
                     /dev_stage="2-cell"
                     /lab_host="DH10B"
     mRNA            <1. .>328
BASE COUNT        86 a     52 c     69 g    121 t
ORIGIN

  Query Match           31.1%;  Score 19;  DB 8;  Length 328;
  Best Local Similarity 100.0%;  Pred. No. 1.33e-04;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     249 ATTGTTTTATTTTGTTTTT 267
           ||||||||||||||||||||
Cp      35 ATTGTTTTATTTTGTTTTT 17


RESULT   17
```



```
LOCUS       N22475         330 bp    mRNA            EST       20-DEC-1995
DEFINITION  yw34g10.s1 Homo sapiens cDNA clone 254178 3'.
ACCESSION   N22475
NID         g1128609
KEYWORDS    EST.
SOURCE      human clone=254178 primer=m13 -40 forward library=Morton Fetal
            Cochlea vector=pBluescript SK- host=SOLR cells (kanamycin
            resistant) Rsite1=EcoRI Rsite2=XhoI The cDNA was oligo (dT) primed
            with an XhoI restriction enzyme recognition site and an 18 base
            poly dT sequence.  For the 5' end, the synthesized cDNA termini
            were treated with T4 DNA polymerase and EcoRI adaptors were ligated
            to the blunt ends. adaptor linker: GAATTCGGCACGAG.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 330)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
            Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 291
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES             Location/Qualifiers
     source          1. .330
                     /organism="Homo sapiens"
                     /clone="254178"
     mRNA            <1. .>330
BASE COUNT       146 a     42 c     45 g     97 t
ORIGIN

  Query Match           31.1%;  Score 19;  DB 19;  Length 330;
  Best Local Similarity 100.0%;  Pred. No. 1.33e-04;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     100 AAAAACAAAATAAAACAAT 118
           ||||||||||||||||||||
Qy      17 AAAAACAAAATAAAACAAT 35
```

```
RESULT   18
LOCUS       AQ094353      332 bp    DNA              GSS       26-AUG-1998
DEFINITION  HS_3025_B2_G04_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3025 Col=8 Row=N, genomic survey
            sequence.
ACCESSION   AQ094353
NID         g3465800
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 332)
  AUTHORS   Mahairas,G.G., Zackrone,K.D., Smith,T., Tipton,S., Schmidt,S.,
            Traicoff,R., Abajian,C., Blanchard,A., West,A. and Hood,L.E.
  TITLE     Construction of a Characterized Clone Resource for Genomic
            Sequencing: Generation and Preliminary Analysis of 20,000 Sequence
            Tagged Connectors
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3025  row: N  column: 8
            Class: BAC ends
            High quality sequence stop: 332.
FEATURES             Location/Qualifiers
     source          1. .332
                     /organism="Homo sapiens"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
                     /db_xref="taxon:9606"
                     /clone="Plate=3025 Col=8 Row=N"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
BASE COUNT      121 a     60 c     59 g     90 t      2 others
ORIGIN

  Query Match           31.1%;  Score 19;  DB 28;  Length 332;
  Best Local Similarity 100.0%;  Pred. No. 1.33e-04;
  Matches    19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    312 TCTTTTCAAAAACAAAATA 330
          ||||||||||||||||||||
```



```
Qy     10 TCTTTTCAAAAACAAAATA 28


RESULT   19
LOCUS       AU021150      341 bp    mRNA             EST       01-AUG-1998
DEFINITION  Mus musculus 8-cell embryo cDNA 3'-end sequence, clone J0543D04,
            mRNA sequence.
ACCESSION   AU021150
NID         g3376734
KEYWORDS    EST; EST(expressed sequence tag).
SOURCE      Mus musculus (strain:C57BL/6J) eight-cell stage embryo cDNA to
            mRNA, clone:J0543D04.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 341)
  AUTHORS   Doi,H. and Ko,M.S.
  TITLE     Systematic analyses of genes expressed in eight-cell stage mouse
            embryos (The ERATO/Doi Project at Wayne State University)
  JOURNAL   Published Only in DataBase (1998) In press
REFERENCE   2  (bases 1 to 341)
  AUTHORS   Doi,H. and Ko,M.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUN-1998) to the DDBJ/EMBL/GenBank databases.
            Hirofumi Doi, Japan Science and Technology Corporation (JST), Doi
            Bioasymmetry Project, ERATO; WBG Marive East 12F, 2-6 Nakase,
            Mihama-ku, Chiba, Chiba 261-7112, Japan (E-mail:hd@bioa.jst.go.jp,
            Tel:+81-43-299-1351, Fax:+81-43-297-7530)
FEATURES             Location/Qualifiers
     source          1. .341
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="J0543D04"
                     /dev_stage="eight-cell stage embryo"
BASE COUNT      133 a     64 c     52 g     92 t
ORIGIN

  Query Match           31.1%;  Score 19;  DB 17;  Length 341;
  Best Local Similarity 100.0%;  Pred. No. 1.33e-04;
  Matches    19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     92 AAAAACAAAATAAAACAAT 110
          ||||||||||||||||||||
Qy     17 AAAAACAAAATAAAACAAT 35


RESULT   20
LOCUS       AU021148      342 bp    mRNA             EST       01-AUG-1998
DEFINITION  Mus musculus 8-cell embryo cDNA 3'-end sequence, clone J0543D02,
            mRNA sequence.
```



```
ACCESSION   AU021148
NID         g3376732
KEYWORDS    EST; EST(expressed sequence tag).
SOURCE      Mus musculus (strain:C57BL/6J) eight-cell stage embryo cDNA to
            mRNA, clone:J0543D02.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 342)
  AUTHORS   Doi,H. and Ko,M.S.
  TITLE     Systematic analyses of genes expressed in eight-cell stage mouse
            embryos (The ERATO/Doi Project at Wayne State University)
  JOURNAL   Published Only in DataBase (1998) In press
REFERENCE   2  (bases 1 to 342)
  AUTHORS   Doi,H. and Ko,M.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUN-1998) to the DDBJ/EMBL/GenBank databases.
            Hirofumi Doi, Japan Science and Technology Corporation (JST), Doi
            Bioasymmetry Project, ERATO; WBG Marive East 12F, 2-6 Nakase,
            Mihama-ku, Chiba, Chiba 261-7112, Japan (E-mail:hd@bioa.jst.go.jp,
            Tel:+81-43-299-1351, Fax:+81-43-297-7530)
FEATURES             Location/Qualifiers
     source          1. .342
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="J0543D02"
                     /dev_stage="eight-cell stage embryo"
BASE COUNT      133 a     64 c     52 g     93 t
ORIGIN

  Query Match           31.1%;  Score 19;  DB 17;  Length 342;
  Best Local Similarity 100.0%;  Pred. No. 1.33e-04;
  Matches    19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     92 AAAAACAAAATAAAACAAT 110
          ||||||||||||||||||||
Qy     17 AAAAACAAAATAAAACAAT 35


RESULT   21
LOCUS       AA601083      353 bp    mRNA             EST       01-OCT-1997
DEFINITION  no53e04.s1 NCI_CGAP_SS1 Homo sapiens cDNA clone IMAGE:1104414
            similar to contains Alu repetitive element;contains element MIR
            repetitive element ;, mRNA sequence.
ACCESSION   AA601083
NID         g2434708
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
```



```
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 353)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: Stratagene, Inc.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 1281   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 340.
FEATURES             Location/Qualifiers
     source          1. .353
                     /organism="Homo sapiens"
                     /note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; Cloned unidirectionally.  Primer: Oligo dT. Synovial
                     sarcoma. 5' adaptor sequence: 5' GAATTCGGCACGAG 3' 3'
                     adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3' Average
                     insert size:  1.4 kb."
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1104414"
                     /clone_lib="NCI_CGAP_SS1"
                     /tissue_type="synovial sarcoma"
                     /lab_host="SOLR (kanamycin resistant)"
BASE COUNT       79 a     73 c     76 g    125 t
ORIGIN

  Query Match           31.1%;  Score 19;  DB 10;  Length 353;
  Best Local Similarity 100.0%;  Pred. No. 1.33e-04;
  Matches    19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    279 TTGTTTTATTTTGTTTTTG 297
          ||||||||||||||||||||
Cp     34 TTGTTTTATTTTGTTTTTG 16


RESULT   22
LOCUS       AU020149      373 bp    mRNA             EST       01-AUG-1998
DEFINITION  Mus musculus 8-cell embryo cDNA 3'-end sequence, clone J0528A06,
            mRNA sequence.
```

```
ACCESSION   AU020149
NID         g3375733
KEYWORDS    EST; EST(expressed sequence tag).
SOURCE      Mus musculus (strain:C57BL/6J) eight-cell stage embryo cDNA to
            mRNA, clone:J0528A06.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 373)
  AUTHORS   Doi,H. and Ko,M.S.
  TITLE     Systematic analyses of genes expressed in eight-cell stage mouse
            embryos (The ERATO/Doi Project at Wayne State University)
  JOURNAL   Published Only in DataBase (1998) In press
REFERENCE   2  (bases 1 to 373)
  AUTHORS   Doi,H. and Ko,M.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUN-1998) to the DDBJ/EMBL/GenBank databases.
            Hirofumi Doi, Japan Science and Technology Corporation (JST), Doi
            Bioasymmetry Project, ERATO; WBG Marive East 12F, 2-6 Nakase,
            Mihama-ku, Chiba, Chiba 261-7112, Japan (E-mail:hd@bioa.jst.go.jp,
            Tel:+81-43-299-1351, Fax:+81-43-297-7530)
FEATURES             Location/Qualifiers
     source          1. .373
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="J0528A06"
                     /dev_stage="eight-cell stage embryo"
BASE COUNT      136 a     73 c     59 g     99 t      6 others
ORIGIN

  Query Match           31.1%;  Score 19;  DB 17;  Length 373;
  Best Local Similarity 100.0%;  Pred. No. 1.33e-04;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     92 AAAAACAAAATAAAACAAT 110
          ||||||||||||||||||
Qy     17 AAAAACAAAATAAAACAAT 35


RESULT   23
LOCUS       W42451        376 bp    mRNA            EST       10-OCT-1996
DEFINITION  zc22e06.s1 Soares senescent fibroblasts NbHSF Homo sapiens cDNA
            clone 323074 3', mRNA sequence.
ACCESSION   W42451
NID         g1326941
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```



```
REFERENCE   1  (bases 1 to 376)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1829   Std Error: 0.00
            Seq primer: mob.REGA+ET
            High quality sequence stop: 284.
FEATURES             Location/Qualifiers
     source          1. .376
                     /organism="Homo sapiens"
                     /note="Vector: pT7T3D (Pharmacia) with a modified
                     polylinker Y_TYPE: phagemid; Site_1: Not I; Site_2: Eco
                     RI; TGTTACCAATCTGAAGTGGGAGCGGCCGCATTTTTTTTTTTTTTTTT
                     3'], double-stranded cDNA was size selected, ligated to
                     Eco RI adapters (Pharmacia), digested with Not I and
                     cloned into the Not I and Eco RI sites of a modified pT7T3
                     vector (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by Bento
                     Soares and M.Fatima Bonaldo."
                     /db_xref="taxon:9606"
                     /clone="323074"
                     /clone_lib="Soares senescent fibroblasts NbHSF"
                     /tissue_type="senescent fibroblast"
                     /lab_host="DH10B (ampicillin resistant)"
     mRNA            complement(<1. .>376)
BASE COUNT      151 a     63 c     49 g     113 t
ORIGIN

  Query Match           31.1%;  Score 19;  DB 21;  Length 376;
  Best Local Similarity 100.0%;  Pred. No. 1.33e-04;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     94 AAAAACAAAATAAAACAAT 112
          ||||||||||||||||||
Qy     17 AAAAACAAAATAAAACAAT 35
```



```
RESULT   24
LOCUS       AA859452      379 bp    mRNA            EST       10-MAR-1998
DEFINITION  UI-R-A0-bf-b-05-0-UI.s1 UI-R-A0 Rattus norvegicus cDNA clone
            UI-R-A0-bf-b-05-0-UI 3' similar to gb|AA465962|AA465962 ve90b04.s1
            Knowles Solter mouse 2 cell Mus musculus cDNA clone 833455 5', mRNA
            sequence.
ACCESSION   AA859452
NID         g2947983
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 379)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT
            Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence tag present in the cDNA between the NotI site and the
            oligo-dT track (not shown) served to identify it as a clone from
            the normalized adult ovary library. cDNA Library Preparation: M.
            Fatima Bonaldo, Ph.D. Clone distribution: clones will be available
            through Research Genetics
            Seq primer: M13 Forward.
FEATURES             Location/Qualifiers
     source          1. .379
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; This library
                     consists of a mixture of individually tagged normalized
                     libraries constructed from rat placenta, adult lung,
                     brain, liver, kidney, heart, spleen, ovary, and muscle.
                     The tag is a string of 3-5 nucleotides present between the
                     Not I site and the oligo-dT track which allows
                     identification of the library of origin of a clone within
                     the mixture."
                     /db_xref="taxon:10116"
                     /clone="UI-R-A0-bf-b-05-0-UI"
                     /clone_lib="UI-R-A0"
                     /dev_stage="adult"
                     /lab_host="DH10B (Life Technologies)"
```



```
BASE COUNT      140 a     77 c     62 g     99 t      1 others
ORIGIN

  Query Match           31.1%;  Score 19;  DB 13;  Length 379;
  Best Local Similarity 100.0%;  Pred. No. 1.33e-04;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     88 AAAAACAAAATAAAACAAT 106
          ||||||||||||||||||
Qy     17 AAAAACAAAATAAAACAAT 35


RESULT   25
LOCUS       R76563        382 bp    mRNA            EST       04-AUG-1997
DEFINITION  yi72h02.s1 Soares placenta Nb2HP Homo sapiens cDNA clone 144819 3',
            mRNA sequence.
ACCESSION   R76563
NID         g851212
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 382)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
            Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 1896
            High quality sequence stops: 295
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1896   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 295.
FEATURES             Location/Qualifiers
     source          1. .382
                     /organism="Homo sapiens"
                     /note="Organ: placenta; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
```

```
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
                     /db_xref="taxon:9606"
                     /clone="144819"
                     /clone_lib="Soares placenta Nb2HP"
                     /sex="Female"
                     /dev_stage="placenta obtained at birth (full term)"
                     /lab_host="DH10B (ampicillin resistant)"
     mRNA            complement(<1. .>382)
                     /db_xref="GDB:557901"
BASE COUNT      140 a      68 c      61 g     107 t       6 others
ORIGIN

  Query Match           31.1%;  Score 19;  DB 8;  Length 382;
  Best Local Similarity 100.0%;  Pred. No. 1.33e-04;
  Matches    19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     77 AAAAACAAAATAAAACAAT 95
          |||||||||||||||||||
Qy     17 AAAAACAAAATAAAACAAT 35


RESULT   26
LOCUS       W03777          384 bp    mRNA            EST       19-APR-1996
DEFINITION  za65b10.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone
            297403 5', mRNA sequence.
ACCESSION   W03777
NID         g1275641
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 384)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,P.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
```



```
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: mob.REGA+ET
            High quality sequence stop: 224.
FEATURES             Location/Qualifiers
     source          1. .384
                     /organism="Homo sapiens"
                     /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                     with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                     1st strand cDNA was primed with a Pac I - oligo(dT) primer
                     [5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Pac I and cloned into the Pac I
                     and Eco RI sites of the modified pT7T3 vector.  Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
                     /db_xref="taxon:9606"
                     /clone="297403"
                     /clone_lib="Soares fetal liver spleen 1NFLS"
                     /sex="male"
                     /dev_stage="20 week-post conception fetus"
                     /lab_host="DH10B (ampicillin resistant)"
     mRNA            <1. .>384
BASE COUNT      108 a      67 c      68 g     140 t       1 others
ORIGIN

  Query Match           31.1%;  Score 19;  DB 20;  Length 384;
  Best Local Similarity 100.0%;  Pred. No. 1.33e-04;
  Matches    19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    185 TTGTTTTATTTTGTTTTTG 203
          |||||||||||||||||||
Cp     34 TTGTTTTATTTTGTTTTTG 16


RESULT   27
LOCUS       AA874518        400 bp    mRNA            EST       19-MAR-1998
DEFINITION  vx03c09.r1 Soares 2NbMT Mus musculus cDNA clone 1263376 5', mRNA
            sequence.
ACCESSION   AA874518
NID         g2979207
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 400)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
```



```
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,P., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:665928
            Seq primer: -28ml3 rev2 ET from Amersham
            High quality sequence stop: 399.
FEATURES             Location/Qualifiers
     source          1. .400
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGGTTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. RNA
                     provided by Dr. Bertrand Jordan. Library went through two
                     rounds of normalization, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
                     /db_xref="taxon:10090"
                     /clone="12633?6"
                     /clone_lib="Soares 2NbMT"
                     /sex="male"
                     /tissue_type="Thymus"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
BASE COUNT      127 a      98 c      88 g      87 t
ORIGIN

  Query Match           31.1%;  Score 19;  DB 13;  Length 400;
  Best Local Similarity 100.0%;  Pred. No. 1.33e-04;
  Matches    19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     91 CAAAAACAAAATAAAACAA 109
          |||||||||||||||||||
Qy     16 CAAAAACAAAATAAAACAA 34
```



```
RESULT   28
LOCUS       AU019558        409 bp    mRNA            EST       01-AUG-1998
DEFINITION  Mus musculus 8-cell embryo cDNA 3'-end sequence, clone J0519C11,
            mRNA sequence.
ACCESSION   AU019558
NID         g3375142
KEYWORDS    EST; EST(expressed sequence tag).
SOURCE      Mus musculus (strain:C57BL/6J) eight-cell stage embryo cDNA to
            mRNA, clone:J0519C11.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 409)
  AUTHORS   Doi,H. and Ko,M.S.
  TITLE     Systematic analyses of genes expressed in eight-cell stage mouse
            embryos (The ERATO/Doi Project at Wayne State University)
  JOURNAL   Published Only in DataBase (1998) In press
REFERENCE   2  (bases 1 to 409)
  AUTHORS   Doi,H. and Ko,M.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUN-1998) to the DDBJ/EMBL/GenBank databases.
            Hirofumi Doi, Japan Science and Technology Corporation (JST), Doi
            Bioasymmetry Project, ERATO; WBG Marive East 12F, 2-6 Nakase,
            Mihama-ku, Chiba, Chiba 261-7112, Japan (E-mail:hd@bioa.jst.go.jp,
            Tel:+81-43-299-1351, Fax:+81-43-297-7530)
FEATURES             Location/Qualifiers
     source          1. .409
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="J0519C11"
                     /dev_stage="eight-cell stage embryo"
BASE COUNT      148 a      80 c      66 g     114 t       1 others
ORIGIN

  Query Match           31.1%;  Score 19;  DB 17;  Length 409;
  Best Local Similarity 100.0%;  Pred. No. 1.33e-04;
  Matches    19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     92 AAAAACAAAATAAAACAAT 110
          |||||||||||||||||||
Qy     17 AAAAACAAAATAAAACAAT 35


RESULT   29
LOCUS       AA614837        413 bp    mRNA            EST       16-OCT-1997
DEFINITION  nn90c12.s1 NCI_CGAP_Br2 Homo sapiens cDNA clone IMAGE:1098454, mRNA
            sequence.
ACCESSION   AA614837
NID         g2467033
```

```
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 413)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 2222   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 350.
FEATURES              Location/Qualifiers
     source           1. .413
                      /organism="Homo sapiens"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; 1st strand cDNA was prepared from pooled bulk
                      breast tumor tissue, and was then primed with a Not I -
                      oligo(dT) primer. Double-stranded cDNA was ligated to Eco
                      RI adaptors (Pharmacia), digested with Not I and cloned
                      into the Not I and Eco RI sites of the modified pT7T3
                      vector.  This library is the normalized version of
                      NCI_CGAP_Br1.1. Library was constructed by Bento Soares
                      and M. Fatima Bonaldo."
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1098454"
                      /clone_lib="NCI_CGAP_Br2"
                      /sex="female, pooled"
                      /tissue_type="breast"
                      /lab_host="DH10B"
BASE COUNT      168 a     70 c     56 g    119 t
ORIGIN

  Query Match           31.1%;  Score 19;  DB 10;  Length 413;
  Best Local Similarity 100.0%;  Pred. No. 1.33e-04;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```



```
Db     100 AAAAACAAAATAAAACAAT 118
           |||||||||||||||||||
Qy      17 AAAAACAAAATAAAACAAT 35


RESULT   30
LOCUS       R07069         415 bp    mRNA            EST       05-APR-1995
DEFINITION  yf13c05.r1 Homo sapiens cDNA clone 126728 5'.
ACCESSION   R07069
NID         g758992
KEYWORDS    EST.
SOURCE      human clone=126728 library=Soares fetal liver spleen 1NFLS
            vector=pT7T3D (Pharmacia) with a modified polylinker host=DH10B
            (ampicillin resistant) primer=M13RP1 Rsite1=Pac I Rsite2=Eco RI
            Liver and spleen from a 20 week-post conception male fetus. 1st
            strand cDNA was primed with a Pac I - oligo(dT) primer [5'
            AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTTT 3'], double-stranded
            cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac
            I and cloned into the Pac I and Eco RI sites of the modified pT7T3
            vector.  Library went through one round of normalization. Library
            constructed by Bento Soares and M.Fatima Bonaldo.
  ORGANISM  Homo sapiens
            Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 415)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High qality sequence stops: 383
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES              Location/Qualifiers
     source           1. .415
                      /organism="Homo sapiens"
                      /clone="126728"
BASE COUNT       92 a    120 c     92 g    105 t      6 others
ORIGIN
```



```
  Query Match           31.1%;  Score 19;  DB 5;  Length 415;
  Best Local Similarity  95.0%;  Pred. No. 1.33e-04;
  Matches   19;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db     363 ATTNTTTTATTTTGTTTTTG 382
           ||| ||||||||||||||||
Cp      35 ATTGTTTTATTTTGTTTTTG 16


RESULT   31
LOCUS       AQ113106       419 bp    DNA             GSS       28-AUG-1998
DEFINITION  CIT-HSP-2379L5.TR CIT-HSP Homo sapiens genomic clone 2379L5,
            genomic survey sequence.
ACCESSION   AQ113106
NID         g3489227
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 419)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
            Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
            Venter,J.C.
  TITLE     Use of a random human BAC End Sequence Database for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1998)
COMMENT     Other_GSSs: CIT-HSP-2379L5.TF
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13 Reverse
            Class: BAC ends.
FEATURES              Location/Qualifiers
     source           1. .419
                      /organism="Homo sapiens"
                      /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                      HindIII"
                      /db_xref="taxon:9606"
                      /clone="2379L5"
                      /clone_lib="CIT-HSP"
                      /sex="Male"
                      /cell_type="Sperm"
BASE COUNT      167 a     94 c     55 g    103 t
```



```
ORIGIN

  Query Match           31.1%;  Score 19;  DB 28;  Length 419;
  Best Local Similarity 100.0%;  Pred. No. 1.33e-04;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      25 CAAAAACAAAATAAAACAA 43
           |||||||||||||||||||
Qy      16 CAAAAACAAAATAAAACAA 34


RESULT   32
LOCUS       B38229         422 bp    DNA             GSS       17-OCT-1997
DEFINITION  HS-1047-A2-A03-MR.abi CIT Human Genomic Sperm Library C Homo
            sapiens genomic clone Plate=CT 830 Col=6 Row=A, genomic survey
            sequence.
ACCESSION   B38229
NID         g2542481
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 422)
  AUTHORS   Mahairas,G.G., Zackrone,K.D., Smith,T., Tipton,S., Schmidt,S.,
            Traicoff,R., Abajian,C., Blanchard,A., West,A. and Hood,L.E.
  TITLE     Construction of a Characterized Clone Resource for Genomic
            Sequencing: Generation and Preliminary Analysis of 20,000 Sequence
            Tagged Connectors
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Mahairas GG, Zackrone KD, Hood L
            University of Washington
            Seattle, WA 98195, USA
            Tel: (206) 616-8744
            Fax: (206) 685-7301
            Email: kzackron@u.washington.edu
            Sequence Tagged Connector
            Plate: CT 830  row: A  column: 6
            Class: BAC ends
            High quality sequence stop: 422.
FEATURES              Location/Qualifiers
     source           1. .422
                      /organism="Homo sapiens"
                      /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                      E-Coli DH10B"
                      /db_xref="taxon:9606"
                      /clone="Plate=CT 830 Col=6 Row=A"
                      /clone_lib="CIT Human Genomic Sperm Library C"
                      /sex="M"
BASE COUNT      124 a     94 c     82 g    122 t
```

```
ORIGIN

  Query Match           31.1%;  Score 19;  DB 26;  Length 422;
  Best Local Similarity 100.0%;  Pred. No. 1.33e-04;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     49 CTTTTCAAAAACAAAATAA 67
          |||||||||||||||||||
Qy     11 CTTTTCAAAAACAAAATAA 29


RESULT   33
LOCUS       N23446         424 bp    mRNA            EST       28-DEC-1995
DEFINITION  yv99d01.s1 Homo sapiens cDNA clone 250849 3'.
ACCESSION   N23446
NID         g1137596
KEYWORDS    EST.
SOURCE      human clone=250849 primer=m13 -40 forward library=Soares melanocyte
            2NbHM vector=pT7T3D (Pharmacia) with a modified polylinker
            host=DH10B (ampicillin resistant) Rsitel=Not I Rsite2=Eco RI Male.
            1st strand cDNA was primed with a Not I - oligo(dT) primer
            [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCAGTTTTTTTTTTTTTTTTT-3'],
            double-stranded cDNA was size selected, ligated to Eco RI adapters
            (Pharmacia), digested with Not I and cloned into the Not I and Eco
            RI sites of a modified pT7T3 vector (Pharmacia). Library
            constructed by Bento Soares and M.Fatima Bonaldo. RNA from normal
            foreskin melanocytes (FS374) was kindly provided by Dr. Anthony P.
            Albino.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 424)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 298
            Source: IMAGE Consortium, LLNL
```



```
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES              Location/Qualifiers
     source           1. .424
                      /organism="Homo sapiens"
                      /clone="250849"
     mRNA             <1. .>424
BASE COUNT      171 a     73 c     58 g    121 t      1 others
ORIGIN

  Query Match           31.1%;  Score 19;  DB 19;  Length 424;
  Best Local Similarity 100.0%;  Pred. No. 1.33e-04;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    100 AAAAACAAAATAAAACAAT 118
          |||||||||||||||||||
Qy     17 AAAAACAAAATAAAACAAT 35


RESULT   34
LOCUS       H27546         429 bp    mRNA            EST       12-JUL-1995
DEFINITION  yl61d01.s1 Homo sapiens cDNA clone 162721 3'.
ACCESSION   H27546
NID         g897536
KEYWORDS    EST.
SOURCE      human clone=162721 library=Soares breast 3NbHBst vector=pT7T3D
            (Pharmacia) with a modified polylinker host=DH10B (ampicillin
            resistant) primer=Promega -21m13 Rsitel=Not I Rsite2=Eco RI Adult
            human. 1st strand cDNA was primed with a Not I - oligo(dT) primer
            [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTT 3'],
            double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia),
            digested with Not I and cloned into the Not I and Eco RI sites of a
            modified pT7T3 vector (Pharmacia). Library went through one round
            of normalization to a Cot = 20. Library constructed by Bento Soares
            and M.Fatima Bonaldo.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 429)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
```



```
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 123
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES              Location/Qualifiers
     source           1. .429
                      /organism="Homo sapiens"
                      /clone="162721"
BASE COUNT      132 a     75 c     83 g    125 t     14 others
ORIGIN

  Query Match           31.1%;  Score 19;  DB 16;  Length 429;
  Best Local Similarity 100.0%;  Pred. No. 1.33e-04;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     94 AAAAACAAAATAAAACAAT 112
          |||||||||||||||||||
Qy     17 AAAAACAAAATAAAACAAT 35


RESULT   35
LOCUS       N27212         446 bp    mRNA            EST       29-DEC-1995
DEFINITION  yw69g10.s1 Homo sapiens cDNA clone 257538 3'.
ACCESSION   N27212
NID         g1141560
KEYWORDS    EST.
SOURCE      human clone=257538 primer=m13 -40 forward library=Soares placenta
            8to9weeks 2NbHP8to9W vector=pT7T3D (Pharmacia) with a modified
            polylinker host=DH10B (ampicillin resistant) Rsitel=Not I
            Rsite2=Eco RI two placentae: one from 8 weeks and another from 9
            weeks post conception. 1st strand cDNA was primed with a Not I -
            oligo(dT) primer
            [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCGATTTTTTTTTTTTTTTTTT-3'],
            double-stranded cDNA was size selected, ligated to Eco RI adapters
            (Pharmacia), digested with Not I and cloned into the Not I and Eco
            RI sites of a modified pT7T3 vector (Pharmacia). Library
            constructed by Bento Soares and M. Fatima Bonaldo.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 446)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,P.,
```



```
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 235
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES              Location/Qualifiers
     source           1. .446
                      /organism="Homo sapiens"
                      /clone="257538"
     mRNA             <1. .>446
BASE COUNT      172 a     77 c     65 g    130 t      2 others
ORIGIN

  Query Match           31.1%;  Score 19;  DB 19;  Length 446;
  Best Local Similarity 100.0%;  Pred. No. 1.33e-04;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     94 AAAAACAAAATAAAACAAT 112
          |||||||||||||||||||
Qy     17 AAAAACAAAATAAAACAAT 35


RESULT   36
LOCUS       N32190         449 bp    mRNA            EST       10-JAN-1996
DEFINITION  yy25e11.s1 Homo sapiens cDNA clone 272300 3'.
ACCESSION   N32190
NID         g1152589
KEYWORDS    EST.
SOURCE      human clone=272300 primer=m13 -40 forward library=Soares melanocyte
            2NbHM vector=pT7T3D (Pharmacia) with a modified polylinker
            host=DH10B (ampicillin resistant) Rsitel=Not I Rsite2=Eco RI Male.
            1st strand cDNA was primed with a Not I - oligo(dT) primer
            [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCAGTTTTTTTTTTTTTTTTT-3'],
            double-stranded cDNA was size selected, ligated to Eco RI adapters
            (Pharmacia), digested with Not I and cloned into the Not I and Eco
            RI sites of a modified pT7T3 vector (Pharmacia). Library
            constructed by Bento Soares and M.Fatima Bonaldo. RNA from normal
            foreskin melanocytes (FS374) was kindly provided by Dr. Anthony P.
            Albino.
  ORGANISM  Homo sapiens
```

```
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 449)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 350
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES              Location/Qualifiers
     source           1. .449
                      /organism="Homo sapiens"
                      /clone="272300"
     mRNA             <1. .>449
BASE COUNT      179 a     76 c     61 g    131 t      2 others
ORIGIN

  Query Match           31.1%;  Score 19;  DB 19;  Length 449;
  Best Local Similarity 100.0%;  Pred. No. 1.33e-04;
  Matches    19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     100 AAAAACAAAATAAAACAAT 118
           ||||||||||||||||||||
Qy      17 AAAAACAAAATAAAACAAT 35


RESULT   37
LOCUS       N27215         450 bp    mRNA            EST       29-DEC-1995
DEFINITION  yw69h10.s1 Homo sapiens cDNA clone 257539 3'.
ACCESSION   N27215
NID         g1141563
KEYWORDS    EST.
SOURCE      human clone=257539 primer=m13 -40 forward library=Soares placenta
            8to9weeks 2NbHP8to9W vector=pT7T3D (Pharmacia) with a modified
            polylinker host=DH10B (ampicillin resistant) Rsite1=Not I
            Rsite2=Eco RI two placentae: one from 8 weeks and another from 9
```



```
            weeks post conception. 1st strand cDNA was primed with a Not I -
            oligo(dT) primer
            [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCGATTTTTTTTTTTTTTTTT-3'],
            double-stranded cDNA was size selected, ligated to Eco RI adapters
            (Pharmacia), digested with Not I and cloned into the Not I and Eco
            RI sites of a modified pT7T3 vector (Pharmacia). Library
            constructed by Bento Soares and M. Fatima Bonaldo.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 450)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 234
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES              Location/Qualifiers
     source           1. .450
                      /organism="Homo sapiens"
                      /clone="257539"
     mRNA             <1. .>450
BASE COUNT      157 a     89 c     73 g    122 t      9 others
ORIGIN

  Query Match           31.1%;  Score 19;  DB 19;  Length 450;
  Best Local Similarity 100.0%;  Pred. No. 1.33e-04;
  Matches    19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      94 AAAAACAAAATAAAACAAT 112
           ||||||||||||||||||||
Qy      17 AAAAACAAAATAAAACAAT 35


RESULT   38
LOCUS       AI151182       472 bp    mRNA            EST       30-SEP-1998
```



```
DEFINITION  qc73e06.x1 Soares_placenta_8to9weeks_2NbHP8to9W Homo sapiens cDNA
            clone IMAGE:1715266 3', mRNA sequence.
ACCESSION   AI151182
NID         g3679651
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 472)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 450.
FEATURES              Location/Qualifiers
     source           1. .472
                      /organism="Homo sapiens"
                      /note="Organ: placenta; Vector: pT7T3D (Pharmacia) with a
                      modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                      strand cDNA was primed with a Not I - oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCGATTTTTTTTTTTTTTTTT 3'],
                      double-stranded cDNA was size selected, ligated to Eco RI
                      adapters (Pharmacia), digested with Not I and cloned into
                      the Not I and Eco RI sites of a modified pT7T3 vector
                      (Pharmacia). Library constructed by Bento Soares and
                      M.Fatima Bonaldo."
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1715266"
                      /clone_lib="Soares_placenta_8to9weeks_2NbHP8to9W"
                      /dev_stage="two placentae: one from 8 weeks and another
                      from 9 weeks post conception"
                      /lab_host="DH10B (ampicillin resistant)"
BASE COUNT      177 a     87 c     64 g    144 t
ORIGIN

  Query Match           31.1%;  Score 19;  DB 18;  Length 472;
  Best Local Similarity 100.0%;  Pred. No. 1.33e-04;
  Matches    19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      77 AAAAACAAAATAAAACAAT 95
           ||||||||||||||||||||
Qy      17 AAAAACAAAATAAAACAAT 35
```



```
RESULT   39
LOCUS       AA875122       473 bp    mRNA            EST       20-MAR-1998
DEFINITION  UI-R-E0-bu-c-01-0-UI.s2 UI-R-E0 Rattus norvegicus cDNA clone
            UI-R-E0-bu-c-01-0-UI 3' similar to gi|2947983|gb|AA859452|AA859452
            UI-R-A0-bf-b-05-0-UI.s1 UI-R-A0 Rattus norvegicus cDNA clone
            UI-R-A0-bf-b-05-0-UI 3' similar to gb|AA465962|AA465962 ve90b04.s1
            Knowles Solter mouse 2 cell Mus musculus cDNA clone 833455 5', mRNA
            sequence.
ACCESSION   AA875122
NID         g2980070
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 473)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT
            Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence tag present in the cDNA between the NotI site and the
            oligo-dT track (not shown) served to identify it as a clone from
            the normalized adult 12-Day-Embryo library. cDNA Library
            Preparation: M. Fatima Bonaldo, Ph.D. Clone distribution: clones
            will be available through Research Genetics
            Seq primer: M13 Forward.
FEATURES              Location/Qualifiers
     source           1. .473
                      /organism="Rattus norvegicus"
                      /strain="Sprague-Dawley"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: NotI; Site_2: EcoRI; This library
                      consists of a mixture of individually tagged normalized
                      libraries constructed from 8, 12 and 18-day embryo.  The
                      tag is a string  of 3-5 nucleotides present between the
                      Not I site and the oligo-dT track  which allows
                      identification of the library of origin of a clone within
                      the mixture."
                      /db_xref="taxon:10116"
                      /clone="UI-R-E0-bu-c-01-0-UI"
                      /clone_lib="UI-R-E0"
```

```
                     /dev_stage="embryonic"
                     /lab_host="DH10B (Life Technologies)"
BASE COUNT       152 a    108 c     89 g    124 t
ORIGIN

  Query Match           31.1%;  Score 19;  DB 13;  Length 473;
  Best Local Similarity 100.0%;  Pred. No. 1.33e-04;
  Matches    19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      33 AAAAACAAAATAAAACAAT 51
           |||||||||||||||||||
Qy      17 AAAAACAAAATAAAACAAT 35


RESULT   40
LOCUS       AU015009      487 bp    mRNA            EST       30-JUL-1998
DEFINITION  Mus musculus 2-cell embryo cDNA 3'-end sequence, clone J0705H01,
            mRNA sequence.
ACCESSION   AU015009
NID         g3369800
KEYWORDS    EST; EST(expressed sequence tag).
SOURCE      Mus musculus (strain:C57BL/6J) 2-cell stage embryo cDNA to mRNA,
            clone:J0705H01.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 487)
  AUTHORS   Doi,H. and Ko,M.S.
  TITLE     Systematic analyses of genes expressed in 2-cell stage mouse
            embryos (The ERATO/Doi Project at Wayne State University)
  JOURNAL   Published Only in DataBase (1998) In press
REFERENCE   2  (bases 1 to 487)
  AUTHORS   Doi,H. and Ko,M.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUN-1998) to the DDBJ/EMBL/GenBank databases.
            Hirofumi Doi, Japan Science and Technology Corporation (JST), Doi
            Bioasymmetry Project, ERATO; WBG Marive East 12F, 2-6 Nakase,
            Mihama-ku, Chiba, Chiba 261-7112, Japan (E-mail:hd@bioa.jst.go.jp,
            Tel:+81-43-299-1351, Fax:+81-43-297-7530)
FEATURES             Location/Qualifiers
     source          1. .487
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="J0705H01"
                     /dev_stage="2-cell stage embryo"
BASE COUNT       158 a    107 c     87 g    131 t      4 others
ORIGIN

  Query Match           31.1%;  Score 19;  DB 17;  Length 487;
  Best Local Similarity 100.0%;  Pred. No. 1.33e-04;
```



```
  Matches    19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      34 AAAAACAAAATAAAACAAT 52
           |||||||||||||||||||
Qy      17 AAAAACAAAATAAAACAAT 35


RESULT   41
LOCUS       N32188        509 bp    mRNA            EST       10-JAN-1996
DEFINITION  yy25e09.s1 Homo sapiens cDNA clone 272296 3'.
ACCESSION   N32188
NID         g1152587
KEYWORDS    EST.
SOURCE      human clone=272296 primer=ml3 -40 forward library=Soares melanocyte
            2NbHM vector=pT7T3D (Pharmacia) with a modified polylinker
            host=DH10B (ampicillin resistant) Rsite1=Not I Rsite2=Eco RI Male.
            1st strand cDNA was primed with a Not I - oligo(dT) primer
            [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCAGTTTTTTTTTTTTTTTTT-3'],
            double-stranded cDNA was size selected, ligated to Eco RI adapters
            (Pharmacia), digested with Not I and cloned into the Not I and Eco
            RI sites of a modified pT7T3 vector (Pharmacia). Library
            constructed by Bento Soares and M.Fatima Bonaldo. RNA from normal
            foreskin melanocytes (FS374) was kindly provided by Dr. Anthony P.
            Albino.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 509)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 350
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES             Location/Qualifiers
     source          1. .509
```



```
                     /organism="Homo sapiens"
                     /clone="272296"
     mRNA            <1. .>509
BASE COUNT       193 a     99 c     71 g    144 t      2 others
ORIGIN

  Query Match           31.1%;  Score 19;  DB 19;  Length 509;
  Best Local Similarity 100.0%;  Pred. No. 1.33e-04;
  Matches    19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     100 AAAAACAAAATAAAACAAT 118
           |||||||||||||||||||
Qy      17 AAAAACAAAATAAAACAAT 35


RESULT   42
LOCUS       AU021048      544 bp    mRNA            EST       01-AUG-1998
DEFINITION  Mus musculus 8-cell embryo cDNA 3'-end sequence, clone J0542A01,
            mRNA sequence.
ACCESSION   AU021048
NID         g3376632
KEYWORDS    EST; EST(expressed sequence tag).
SOURCE      Mus musculus (strain:C57BL/6J) eight-cell stage embryo cDNA to
            mRNA, clone:J0542A01.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 544)
  AUTHORS   Doi,H. and Ko,M.S.
  TITLE     Systematic analyses of genes expressed in eight-cell stage mouse
            embryos (The ERATO/Doi Project at Wayne State University)
  JOURNAL   Published Only in DataBase (1998) In press
REFERENCE   2  (bases 1 to 544)
  AUTHORS   Doi,H. and Ko,M.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUN-1998) to the DDBJ/EMBL/GenBank databases.
            Hirofumi Doi, Japan Science and Technology Corporation (JST), Doi
            Bioasymmetry Project, ERATO; WBG Marive East 12F, 2-6 Nakase,
            Mihama-ku, Chiba, Chiba 261-7112, Japan (E-mail:hd@bioa.jst.go.jp,
            Tel:+81-43-299-1351, Fax:+81-43-297-7530)
FEATURES             Location/Qualifiers
     source          1. .544
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="J0542A01"
                     /dev_stage="eight-cell stage embryo"
BASE COUNT       179 a    112 c     96 g    155 t      2 others
ORIGIN

  Query Match           31.1%;  Score 19;  DB 17;  Length 544;
```



```
  Best Local Similarity 100.0%;  Pred. No. 1.33e-04;
  Matches    19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      92 AAAAACAAAATAAAACAAT 110
           |||||||||||||||||||
Qy      17 AAAAACAAAATAAAACAAT 35


RESULT   43
LOCUS       AQ160686      580 bp    DNA             GSS       08-SEP-1998
DEFINITION  mgxb0006L08r CUGI Rice Blast BAC Library Magnaporthe grisea genomic
            clone mgxb0006L08r, genomic survey sequence.
ACCESSION   AQ160686
NID         g3557675
KEYWORDS    GSS.
SOURCE      Magnaporthe grisea.
  ORGANISM  Magnaporthe grisea
            Eukaryota; Fungi; Ascomycota; Euascomycetes; Pyrenomycetes;
            Phyllachorales; Phyllachoraceae; Magnaporthe.
REFERENCE   1  (bases 1 to 580)
  AUTHORS   Yu,Y., Zhu,H., Boyd,C.A., Gaudette,B., Gayle,A., Kingsbury,R.,
            Phillips,K., Sasinowski,M, Wing,R.A. and Dean,R.A.
  TITLE     A BAC End Sequencing Framework to Sequence the Magnaporthe grisea
            Genome
  JOURNAL   Unpublished (1998)
COMMENT
            Contact: Dean RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson Universiy, Clemson, SC 29634
            Tel: 864 656 5737
            Fax: 864 656 4293
            Email: rdean@clemson.edu
            Seq primer: GGAAACAGCTATGACCATG
            Class: BAC ends
            High quality sequence stop: 313.
FEATURES             Location/Qualifiers
     source          1. .580
                     /organism="Magnaporthe grisea"
                     /strain="70-15"
                     /note="Vector: pBACWICH; Site_1: HindIII; Site_2: HindIII;
                     Rice blast is one of the most devestating fungal diseases
                     of rice  world wide.  It is a filamentous ascomycete with
                     a haploid genome (n=7)  of approximately 40 Mbp.  Rice
                     blast is an important model fungal pathogen  for studying
                     numerous aspects of the fungal-host interaction.  In
                     order to facilitate genome wide analysis, a BAC library
                     containing  9216 clones with an average insert size of 130
                     kbp was constructed.  This library represents greater
                     than 25X genome coverage.  High density  colony filters
                     are available upon request."
```

```
                    /db_xref="taxon:5138"
                    /clone="mgxb0006L08r"
                    /clone_lib="CUGI Rice Blast BAC Library"
                    /tissue_type="Protoplasts"
                    /lab_host="E. coli DH10B"
BASE COUNT      159 a    161 c    116 g    144 t
ORIGIN

  Query Match           31.1%;  Score 19;  DB 28;  Length 580;
  Best Local Similarity 100.0%;  Pred. No. 1.33e-04;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    134 TTTATTGTTTTATTTTGTT 152
          |||||||||||||||||||
Cp     38 TTTATTGTTTTATTTTGTT 20


RESULT   44
LOCUS       AU018946      582 bp    mRNA            EST       31-JUL-1998
DEFINITION  Mus musculus 8-cell embryo cDNA 3'-end sequence, clone J0510D09,
            mRNA sequence.
ACCESSION   AU018946
NID         g3374436
KEYWORDS    EST; EST(expressed sequence tag).
SOURCE      Mus musculus (strain:C57BL/6J) eight-cell stage embryo cDNA to
            mRNA, clone:J0510D09.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 582)
  AUTHORS   Doi,H. and Ko,M.S.
  TITLE     Systematic analyses of genes expressed in eight-cell stage mouse
            embryos (The ERATO/Doi Project at Wayne State University)
  JOURNAL   Published Only in DataBase (1998) In press
REFERENCE   2  (bases 1 to 582)
  AUTHORS   Doi,H. and Ko,M.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUN-1998) to the DDBJ/EMBL/GenBank databases.
            Hirofumi Doi, Japan Science and Technology Corporation (JST), Doi
            Bioasymmetry Project, ERATO; WBG Marive East 12F, 2-6 Nakase,
            Mihama-ku, Chiba, Chiba 261-7112, Japan (E-mail:hd@bioa.jst.go.jp,
            Tel:+81-43-299-1351, Fax:+81-43-297-7530)
FEATURES              Location/Qualifiers
     source           1. .582
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="J0510D09"
                      /dev_stage="eight-cell stage embryo"
BASE COUNT      189 a    122 c    101 g    158 t     12 others
ORIGIN
```



```
  Query Match           31.1%;  Score 19;  DB 17;  Length 582;
  Best Local Similarity 100.0%;  Pred. No. 1.33e-04;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     92 AAAAACAAAATAAAACAAT 110
          |||||||||||||||||||
Qy     17 AAAAACAAAATAAAACAAT 35


RESULT   45
LOCUS       AQ001905      720 bp    DNA             GSS       25-JUN-1998
DEFINITION  CIT-HSP-2290L15.TF CIT-HSP Homo sapiens genomic clone 2290L15,
            genomic survey sequence.
ACCESSION   AQ001905
NID         g3029109
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 720)
  AUTHORS   Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K.,
            Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
            Simon,M. and Venter,J.C.
  TITLE     Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: M13-21;
            Class: BAC ends.
FEATURES              Location/Qualifiers
     source           1. .720
                      /organism="Homo sapiens"
                      /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                      HindIII"
                      /db_xref="taxon:9606"
                      /clone="2290L15"
                      /clone_lib="CIT-HSP"
                      /sex="Male"
                      /cell_type="Sperm"
```



```
BASE COUNT      252 a    146 c    134 g    188 t
ORIGIN

  Query Match           31.1%;  Score 19;  DB 27;  Length 720;
  Best Local Similarity 100.0%;  Pred. No. 1.33e-04;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     70 CAAAAACAAAATAAAACAA 88
          ||||||||||||||||||
Qy     16 CAAAAACAAAATAAAACAA 34


RESULT   46
LOCUS       AA946594      122 bp    mRNA            EST       23-JUL-1998
DEFINITION  oq38b08.s1 NCI_CGAP_Kid5 Homo sapiens cDNA clone IMAGE:1588599 3',
            mRNA sequence.
ACCESSION   AA946594
NID         g3109989
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 122)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 1434   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 111.
FEATURES              Location/Qualifiers
     source           1. .122
                      /organism="Homo sapiens"
                      /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                      a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                      strand cDNA was primed with a Not I - oligo(dT) primer [5'
                      AACTGGAAGAATTCGCGGCCGCAAATATTTTTTTTTTTTTTTTT 3'],
```



```
                      double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Not I and cloned into the Not I
                      and Eco RI sites of the modified pT7T3 vector. Library
                      went through one round of normalization. Library
                      constructed by Bento Soares and M. Fatima Bonaldo."
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1588599"
                      /clone_lib="NCI_CGAP_Kid5"
                      /tissue_type="2 pooled tumors (clear cell type)"
                      /lab_host="DH10B"
BASE COUNT       36 a     18 c     15 g     53 t
ORIGIN

  Query Match           29.5%;  Score 18;  DB 15;  Length 122;
  Best Local Similarity 100.0%;  Pred. No. 2.26e-03;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     99 AAAAACAAAATAAAACAA 116
          ||||||||||||||||||
Qy     17 AAAAACAAAATAAAACAA 34


RESULT   47
LOCUS       AA702751      129 bp    mRNA            EST       19-DEC-1997
DEFINITION  zi82f11.s1 Soares fetal liver spleen 1NFLS S1 Homo sapiens cDNA
            clone 447309 3', mRNA sequence.
ACCESSION   AA702751
NID         g2705864
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 129)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
            Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
            Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 118.
```

```
FEATURES             Location/Qualifiers
     source          1. .129
                     /organism="Homo sapiens"
                     /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                     with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                     This is a subtracted version of the original Soares fetal
                     liver spleen 1NFLS library.  1st strand cDNA was primed
                     with a Pac I - oligo(dT) primer [5'
                     AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Pac I and cloned into the Pac I
                     and Eco RI sites of the modified pT7T3 vector.  Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
                     /db_xref="GDB:1351566"
                     /db_xref="taxon:9606"
                     /clone="447309"
                     /clone_lib="Soares fetal liver spleen 1NFLS S1"
                     /sex="male"
                     /dev_stage="20 week-post conception fetus"
                     /lab_host="DH10B (ampicillin resistant)"
BASE COUNT       49 a      24 c      25 g      31 t
ORIGIN

  Query Match           29.5%;  Score 18;  DB 11;  Length 129;
  Best Local Similarity 100.0%;  Pred. No. 2.26e-03;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     43 AAAAACAAAATAAAACAA 60
          ||||||||||||||||||
Qy     17 AAAAACAAAATAAAACAA 34


RESULT   48
LOCUS       AA059119       145 bp    mRNA            EST       19-SEP-1996
DEFINITION  zf64g04.s1 Soares retina N2b4HR Homo sapiens cDNA clone 381750 3',
            mRNA sequence.
ACCESSION   AA059119
NID         g1551944
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 145)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
```



```
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40M13 fwd. from Amersham
            High quality sequence stop: 123.
FEATURES             Location/Qualifiers
     source          1. .145
                     /organism="Homo sapiens"
                     /note="Organ: eye; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGCTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). The retinas were obtained from a 55 year old
                     Caucasian  and total cellular poly(A)+ RNA was extracted 6
                     hrs after their removal. The retina RNA  was kindly
                     provided by Roderick R. McInnes M.D. Ph.D. from the
                     University  of Toronto. Library constructed by Bento
                     Soares and M.Fatima Bonaldo."
                     /db_xref="taxon:9606"
                     /clone="381750"
                     /clone_lib="Soares retina N2b4HR"
                     /sex="male"
                     /tissue_type="retina"
                     /dev_stage="55 year old"
                     /lab_host="DH10B (ampicillin resistant)"
     mRNA            complement(<1. .>145)
BASE COUNT       53 a      31 c      29 g      32 t
ORIGIN

  Query Match           29.5%;  Score 18;  DB 21;  Length 145;
  Best Local Similarity 100.0%;  Pred. No. 2.26e-03;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     52 AAAAACAAAATAAAACAA 69
          ||||||||||||||||||
Qy     17 AAAAACAAAATAAAACAA 34


RESULT   49
LOCUS       AI066361       180 bp    mRNA            EST       30-JUL-1998
```



```
DEFINITION  TENU2477 T. cruzi epimastigote normalized cDNA Library Trypanosoma
            cruzi cDNA clone 26f11 5', mRNA sequence.
ACCESSION   AI066361
NID         g3366526
KEYWORDS    EST.
SOURCE      Trypanosoma cruzi.
  ORGANISM  Trypanosoma cruzi
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma; Schizotrypanum.
REFERENCE   1  (bases 1 to 180)
  AUTHORS   Porcel,B., Tran,A-N., Andersson,B., Nyarady,Z., Urmenyi,T.P.,
            Rondinelli,E., Pettersson,U. and Aslund,L.
  TITLE     Generation of Expressed Sequence Tags from Trypanosoma cruzi
  JOURNAL   Unpublished (1998)
COMMENT
            Contact: Aslund L
            Department of Medical Genetics
            Uppsala University
            Biomedical Center, Box 589, S-751 23 Uppsala, Sweden
            Tel: 46 18 471 45 85
            Fax: 46 18 52 68 49
            Email: lena.aslund@medgen.uu.se
            Seq primer: T7
            High quality sequence stop: 180.
FEATURES             Location/Qualifiers
     source          1. .180
                     /organism="Trypanosoma cruzi"
                     /strain="Cl-Brenner"
                     /note="cDNA library constructed with oligo dT primed
                     epimastigote mRNA and cloned in pt7t318D phagemid with
                     modified polylinker (PHARMACIA)"
                     /db_xref="taxon:5693"
                     /clone="26f11"
                     /clone_lib="T. cruzi epimastigote normalized cDNA Library"
                     /cell_type="epimastigote"
BASE COUNT       68 a      20 c      20 g      72 t
ORIGIN

  Query Match           29.5%;  Score 18;  DB 17;  Length 180;
  Best Local Similarity 100.0%;  Pred. No. 2.26e-03;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    121 TTGTTTTATTTTGTTTTT 138
          ||||||||||||||||||
Cp     34 TTGTTTTATTTTGTTTTT 17


RESULT   50
LOCUS       H98966         188 bp    mRNA            EST       15-DEC-1995
DEFINITION  yx11b08.s1 Homo sapiens cDNA clone 261399 3'.
ACCESSION   H98966
```



```
NID         g1123634
KEYWORDS    EST.
SOURCE      human clone=261399 primer=m13 -40 forward library=Soares melanocyte
            2NbHM vector=pT7T3D (Pharmacia) with a modified polylinker
            host=DH10B (ampicillin resistant) Rsite1=Not I Rsite2=Eco RI Male.
            1st strand cDNA was primed with a Not I - oligo(dT) primer
            (5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCAGTTTTTTTTTTTTTTTTT-3'),
            double-stranded cDNA was size selected, ligated to Eco RI adapters
            (Pharmacia), digested with Not I and cloned into the Not I and Eco
            RI sites of a modified pT7T3 vector (Pharmacia). Library
            constructed by Bento Soares and M.Fatima Bonaldo. RNA from normal
            foreskin melanocytes (FS374) was kindly provided by Dr. Anthony P.
            Albino.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 188)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 173
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES             Location/Qualifiers
     source          1. .188
                     /organism="Homo sapiens"
                     /clone="261399"
     mRNA            <1. .>188
BASE COUNT       56 a      39 c      33 g      60 t
ORIGIN

  Query Match           29.5%;  Score 18;  DB 19;  Length 188;
  Best Local Similarity 100.0%;  Pred. No. 2.26e-03;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     96 AAAAACAAAATAAAACAA 113
```

```
                  ||||||||||||||||||
Qy      17 AAAAACAAAATAAAACAA 34
```

Search completed: Tue Dec 22 21:18:53 1998
Job time : 242 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. ~ n.a. database search, using Smith-Waterman algorithm

Run on: Tue Dec 22 21:19:12 1998; MasPar time 23.89 Seconds
353.135 Million cell updates/sec

Tabular output not generated.

Title: >US-08-530-112A-1449
Description: (1-62) from US08530112A.seq
Perfect Score: 61
N.A. Sequence: 1 GATCCCCCATTCTTTTCAAAA.....GCAAGGAAGACTGANGGAAA 62
Comp: CTAGGGGGTAAGAAAAGTTTT.....CGTTCCTTCTGACTNCCTTT

Scoring table: TABLE jmetric
Gap 60

Nmatch STD : Dbase 0; Query 0

Searched: 188442 seqs, 68026449 bases x 2

Post-processing: Minimum Match 0%
Listing first 1000 summaries

Database: n-geneseq32
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39 40:part40

Statistics: Mean 5.658; Variance 3.191; scale 1.773

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description	Pred. No.
1		61	100.0	62	19	T20449	Human gene signature	2.64e-29
2	c	17	27.9	1116	32	T61096	N-acetylglucosaminyl	2.52e+00
3	c	16	26.2	625	18	T19003	Human gene signature	8.79e+00
4		16	26.2	1855	30	T66381	Human adrenal gland c	8.79e+00
5	c	16	26.2	6060	28	T48454	Human beta-globin gen	8.79e+00
6		16	26.2	14507	32	T73568	Expression augmenting	8.79e+00
7	c	15	24.6	26	23	T39432	Hel-N2 selected seque	2.97e+01
8	c	15	24.6	103	19	T20056	Human gene signature	2.97e+01
9		15	24.6	111	20	T23541	Human gene signature	2.97e+01
10	c	15	24.6	330	18	T19543	Human gene signature	2.97e+01
11		15	24.6	369	8	Q59994	Human brain Expressed	2.97e+01
12	c	15	24.6	503	1	Q02418	Promoter of pCP-A1 ex	2.97e+01
13		15	24.6	633	38	T91347	Human cystatin F poly	2.97e+01
14		15	24.6	633	37	T89171	Human cystatin F enco	2.97e+01
15		15	24.6	674	9	Q53499	Sequence encoding pro	2.97e+01
16		15	24.6	675	9	Q53500	Sequence encoding pro	2.97e+01
17		15	24.6	690	8	Q48942	Human HT6.	2.97e+01
18		15	24.6	696	2	Q13299	Growth hormone-like g	2.97e+01
19	c	15	24.6	1163	32	T75004	Human endogenous retr	2.97e+01
20		15	24.6	1227	5	Q31826	Recombinant TCS seque	2.97e+01
21		15	24.6	1227	1	Q06343	Encodes Trichosanthin	2.97e+01
22		15	24.6	1228	4	Q26499	Trichosanthin coding	2.97e+01
23		15	24.6	1233	18	T16869	Pepper Group 2 consti	2.97e+01
24		15	24.6	1400	24	T45418	B. subtilis enterobac	2.97e+01
25		15	24.6	1400	10	Q62460	Bacillus subtilis sid	2.97e+01
26		15	24.6	1510	18	T15760	Human interleukin-15	2.97e+01
27		15	24.6	1534	18	T15995	Human interleukin-15	2.97e+01
28		15	24.6	1641	18	T15758	Human interleukin-15	2.97e+01
29		15	24.6	1727	18	T16871	Pepper Group 2 gene f	2.97e+01
30		15	24.6	2161	10	Q58742	dMEF2.	2.97e+01
31		15	24.6	2623	32	T51715	cDNA for adenovirus E	2.97e+01
32		15	24.6	2655	6	Q32864	Tcr gene derived from	2.97e+01
33		15	24.6	2844	2	N71250	Sequence of Vicia fab	2.97e+01
34	c	15	24.6	2960	29	T41616	Adzuki bean endo-xylo	2.97e+01
35	c	15	24.6	3012	3	Q21188	Sequence encoding a 3	2.97e+01
36	c	15	24.6	3017	8	Q49956	Impatiens Necrotic Sp	2.97e+01
37	c	15	24.6	3207	3	Q14263	p-meta-1 metastasis-a	2.97e+01
38		15	24.6	3649	1	Q04476	Sequence derived by r	2.97e+01
39	c	15	24.6	4495	24	T32143	Helicobacter-specific	2.97e+01
40	c	15	24.6	6152	32	T78867	P. falciparum liver s	2.97e+01
41		15	24.6	7451	40	V05856	Complete sequence of	2.97e+01
42	c	15	24.6	8878	22	T03852	Human fibrinogen B-be	2.97e+01

43	c	15	24.6	9295	29	T62125	Arabidopsis thaliana	2.97e+01
44		15	24.6	10798	4	Q28505	Tomato vacuolar inver	2.97e+01
45		15	24.6	10965	4	Q28507	Tomato vacuolar inver	2.97e+01
46	c	15	24.6	40352	40	V02032	MAGE-B cluster DNA se	2.97e+01
47	c	15	24.6	133894	17	T13635	AcNPV genomic DNA clo	2.97e+01
48		15	24.6	580073	27	T58840	Mycoplasma genitalium	2.97e+01
49	c	15	24.6	580073	27	T58840	Mycoplasma genitalium	2.97e+01
50	c	14	23.0	87	6	Q41022	P11 promoter mutant P	9.69e+01
51	c	14	23.0	90	6	Q41023	P11 promoter mutant s	9.69e+01
52	c	14	23.0	152	20	T21997	Human gene signature	9.69e+01
53		14	23.0	170	23	T13570	Oligo 3 for construct	9.69e+01
54		14	23.0	170	33	T73985	Rat THY-1 synthetic g	9.69e+01
55	c	14	23.0	198	18	T14853	Barley U14 intergenic	9.69e+01
56		14	23.0	206	18	T19267	Human gene signature	9.69e+01
57	c	14	23.0	215	19	T20150	Human gene signature	9.69e+01
58	c	14	23.0	225	21	T26269	Human gene signature	9.69e+01
59		14	23.0	234	19	T21928	Human gene signature	9.69e+01
60		14	23.0	236	22	T25193	Human gene signature	9.69e+01
61	c	14	23.0	251	19	T20099	Human gene signature	9.69e+01
62		14	23.0	255	8	Q60878	Human brain Expressed	9.69e+01
63		14	23.0	277	1	Q06363	Probe for Clostridium	9.69e+01
64	c	14	23.0	278	8	Q60212	Human brain Expressed	9.69e+01
65	c	14	23.0	330	21	T24518	Human gene signature	9.69e+01
66		14	23.0	358	8	Q59477	Human brain Expressed	9.69e+01
67		14	23.0	486	22	T13558	THY-1 containing HIV	9.69e+01
68		14	23.0	486	33	T73953	Rat THY-1 synthetic g	9.69e+01
69		14	23.0	988	3	Q20500	Encodes fibrinogenoly	9.69e+01
70	c	14	23.0	1085	1	N91113	Apr gene encoding alk	9.69e+01
71	c	14	23.0	1374	40	V20668	Human ninjurin 1 vari	9.69e+01
72		14	23.0	1384	35	T80017	Human galanin recepto	9.69e+01
73	c	14	23.0	1395	11	Q66812	AmEPV spheroidin G1L	9.69e+01
74	c	14	23.0	1464	4	Q24053	FK-506 binding protei	9.69e+01
75	c	14	23.0	1472	3	Q20509	FKBP.	9.69e+01
76	c	14	23.0	1505	35	T72177	Alzheimer's disease D	9.69e+01
77	c	14	23.0	1761	16	Q74061	Human beta-actin.	9.69e+01
78	c	14	23.0	1779	34	T85394	Saccharomyces cerevis	9.69e+01
79		14	23.0	1808	9	Q55616	N.meningitidis IM2394	9.69e+01
80		14	23.0	1808	21	T11239	Neisseria meningitidi	9.69e+01
81		14	23.0	1808	36	T64590	Gene sequence for HTR	9.69e+01
82	c	14	23.0	1841	38	V05504	DNA for BPN' subtili	9.69e+01
83	c	14	23.0	1868	3	N50541	Sequence of the apr[B	9.69e+01
84	c	14	23.0	1868	18	T07077	Subtilisin BPN' gene.	9.69e+01
85	c	14	23.0	1868	22	T27626	Wild type subtilisin	9.69e+01
86		14	23.0	2021	32	T58247	CH14-2a16-1, over abu	9.69e+01
87	c	14	23.0	2026	40	V09901	cDNA encoding a mouse	9.69e+01
88		14	23.0	2299	1	N81537	New DNA sequence isol	9.69e+01
89		14	23.0	2299	1	N82404	Cowpox virus gene inv	9.69e+01
90		14	23.0	2316	15	Q90332	Exons V & VI and intr	9.69e+01
91		14	23.0	2316	15	Q89559	Human GTP7 clone, lam	9.69e+01
92	c	14	23.0	2492	37	T89890	Mouse frizzled gene 6	9.69e+01
93	c	14	23.0	2514	20	T30308	Mouse Sox-9 cDNA.	9.69e+01

94	c	14	23.0	2717	2	N70621	Pho8 promoter and gen	9.69e+01
95	c	14	23.0	2855	39	T97297	Human sex comb on mid	9.69e+01
96	c	14	23.0	2855	39	V00011	Human sex comb on mid	9.69e+01
97	c	14	23.0	2885	14	Q55136	Staphylococcus aureus	9.69e+01
98	c	14	23.0	2910	31	T62461	ThR-GPA1 fusion gene.	9.69e+01
99		14	23.0	2946	30	T68233	H. pylori secreted or	9.69e+01
100	c	14	23.0	2967	31	T62460	STE2-GPA1 fusion gene	9.69e+01
101	c	14	23.0	3090	14	Q85925	Human GLM-2 protein t	9.69e+01
102	c	14	23.0	3255	39	V00013	Human sex comb on mid	9.69e+01
103	c	14	23.0	3255	39	T97299	Human sex comb on mid	9.69e+01
104	c	14	23.0	3327	39	T97298	Human sex comb on mid	9.69e+01
105	c	14	23.0	3327	39	V00012	Human sex comb on mid	9.69e+01
106	c	14	23.0	3694	2	N70622	Pho8 promoter and gen	9.69e+01
107		14	23.0	3849	40	V12271	Temperate phage int g	9.69e+01
108	c	14	23.0	3934	20	T30309	Human SOX-9 cDNA.	9.69e+01
109	c	14	23.0	3969	38	V05762	Chicken protein tyros	9.69e+01
110	c	14	23.0	4200	35	T78912	Spinocerebellar ataxi	9.69e+01
111	c	14	23.0	4821	24	T38305	Tospovirus N-RNA sequ	9.69e+01
112		14	23.0	5166	39	V04024	Human dendritic cell	9.69e+01
113		14	23.0	5477	23	T33514	Human DEC-205 cDNA.	9.69e+01
114		14	23.0	5810	37	T91323	Arabidopsis thaliana	9.69e+01
115		14	23.0	5957	1	N90002	Nucleotide Sequence o	9.69e+01
116		14	23.0	6464	8	Q48770	Arylamine N-acetyl-tr	9.69e+01
117		14	23.0	6464	8	Q48768	Arylamine N-acetyl-tr	9.69e+01
118		14	23.0	6464	8	Q48772	Arylamine N-acetyl-tr	9.69e+01
119	c	14	23.0	6768	39	V14507	AmEPV entomopoxvirus	9.69e+01
120	c	14	23.0	6768	39	V14517	AmEPV entomopoxvirus	9.69e+01
121	c	14	23.0	6768	5	Q28301	AmEPV Spheroidin DNA.	9.69e+01
122		14	23.0	7286	18	T14547	Cytotactin gene.	9.69e+01
123	c	14	23.0	7824	16	Q74052	Human interleukin-1-b	9.69e+01
124	c	14	23.0	8457	11	Q66797	AmEPV spheroidin gene	9.69e+01
125	c	14	23.0	10140	22	T32232	Plasmid pBE92.	9.69e+01
126	c	14	23.0	10207	14	Q55134	Staphylococcus aureus	9.69e+01
127		14	23.0	43804	38	T86375	Chicken embryo lethal	9.69e+01
128		14	23.0	133894	17	T13635	AcNPV genomic DNA clo	9.69e+01
129		13	21.3	20	6	Q35381	PCR primer HIV3B2A.	3.03e+02
130		13	21.3	20	5	Q35805	HIV-1 gp160 primer HI	3.03e+02
131	c	13	21.3	26	23	T39430	Hel-N2 selected seque	3.03e+02
132		13	21.3	32	34	T73500	Human gamma transcrip	3.03e+02
133		13	21.3	32	39	V12731	Primer o-383 for huma	3.03e+02
134		13	21.3	40	37	T97402	Synthetic forty mer n	3.03e+02
135		13	21.3	40	37	T97429	Oligomer s14102 Allel	3.03e+02
136	c	13	21.3	54	14	Q81383	Forward primer to mod	3.03e+02
137	c	13	21.3	56	35	T80143	Human immunoglobulin	3.03e+02
138	c	13	21.3	56	28	T49194	Human IgG1 CH2 domain	3.03e+02
139		13	21.3	59	28	T49193	Human IgG1 CH2 domain	3.03e+02
140		13	21.3	59	14	Q81182	Reverse primer to mod	3.03e+02
141		13	21.3	59	35	T80142	Human immunoglobulin	3.03e+02
142	c	13	21.3	68	8	Q48057	Human gamma4 constant	3.03e+02
143	c	13	21.3	68	17	Q98876	Humanised antibody he	3.03e+02
144	c	13	21.3	68	8	Q48369	25D2 MAb heavy chain	3.03e+02

c	145	13	21.3	75	20	T23304	Human gene signature	3.03e+02
c	146	13	21.3	89	9	Q54235	Clone 2.5 insert.	3.03e+02
	147	13	21.3	99	14	Q81384	Reverse primer to mod	3.03e+02
	148	13	21.3	99	35	T80144	Human immunoglobulin	3.03e+02
	149	13	21.3	99	28	T49195	Human IgG1 CH2 domain	3.03e+02
c	150	13	21.3	99	35	T80145	Human immunoglobulin	3.03e+02
c	151	13	21.3	104	19	T20400	Human gene signature	3.03e+02
	152	13	21.3	142	20	T23079	Human gene signature	3.03e+02
	153	13	21.3	144	23	T26214	Human gene signature	3.03e+02
	154	13	21.3	150	21	T24590	Human gene signature	3.03e+02
c	155	13	21.3	160	21	T26538	Human gene signature	3.03e+02
	156	13	21.3	161	20	T21984	Human gene signature	3.03e+02
	157	13	21.3	162	21	T24430	Human gene signature	3.03e+02
	158	13	21.3	174	22	T25239	Human gene signature	3.03e+02
c	159	13	21.3	183	19	T20205	Human gene signature	3.03e+02
c	160	13	21.3	207	20	T21477	Human gene signature	3.03e+02
c	161	13	21.3	228	22	T25862	Human gene signature	3.03e+02
	162	13	21.3	239	35	T85702	Exon 7 frameshift var	3.03e+02
	163	13	21.3	240	35	T85701	Exon 7 of LA/SS-B aut	3.03e+02
c	164	13	21.3	240	3	N82458	Sequence of gene for	3.03e+02
c	165	13	21.3	240	3	N82457	Sequence of gene for	3.03e+02
c	166	13	21.3	240	3	N82461	Sequence of gene for	3.03e+02
c	167	13	21.3	240	3	N82460	Sequence of gene for	3.03e+02
c	168	13	21.3	240	3	N82459	Sequence of gene for	3.03e+02
c	169	13	21.3	240	3	N82462	Sequence of gene for	3.03e+02
c	170	13	21.3	246	30	T77508	H. pylori secreted or	3.03e+02
c	171	13	21.3	246	30	T67828	H. pylori secreted or	3.03e+02
c	172	13	21.3	272	18	T19583	Human gene signature	3.03e+02
	173	13	21.3	276	40	V24716	H. pylori ORF 29ge103	3.03e+02
c	174	13	21.3	277	18	T19181	Human gene signature	3.03e+02
c	175	13	21.3	278	20	T23788	Human gene signature	3.03e+02
c	176	13	21.3	287	18	T19235	Human gene signature	3.03e+02
	177	13	21.3	295	20	T23727	Human gene signature	3.03e+02
c	178	13	21.3	297	22	T26314	Human gene signature	3.03e+02
	179	13	21.3	301	18	T19558	Human gene signature	3.03e+02
	180	13	21.3	303	20	T23565	Human gene signature	3.03e+02
c	181	13	21.3	303	20	T21882	Human gene signature	3.03e+02
c	182	13	21.3	306	19	T21317	Human gene signature	3.03e+02
c	183	13	21.3	308	20	T23276	Human gene signature	3.03e+02
c	184	13	21.3	312	20	T23758	Human gene signature	3.03e+02
	185	13	21.3	314	8	Q60684	Human brain Expressed	3.03e+02
c	186	13	21.3	314	20	T23530	Human gene signature	3.03e+02
c	187	13	21.3	317	20	T24078	Human gene signature	3.03e+02
c	188	13	21.3	330	1	N81363	Encodes CH2-segment o	3.03e+02
c	189	13	21.3	330	2	N70846	Sequence of CH(2) dom	3.03e+02
c	190	13	21.3	330	2	N60256	Human Ig heavy chain	3.03e+02
	191	13	21.3	342	19	T19826	Human gene signature	3.03e+02
	192	13	21.3	345	8	Q60695	Human brain Expressed	3.03e+02
	193	13	21.3	359	8	Q61275	Human brain Expressed	3.03e+02
c	194	13	21.3	390	30	T68062	H. pylori surface or	3.03e+02
c	195	13	21.3	391	8	Q59487	Human brain Expressed	3.03e+02

	196	13	21.3	418	31	T43479	ATM gene exon 45.	3.03e+02
	197	13	21.3	423	1	Q04274	Sub 2, a HIV fusion p	3.03e+02
	198	13	21.3	423	1	N90623	Sequence encoding Sub	3.03e+02
c	199	13	21.3	423	30	T67511	H. pylori surface mem	3.03e+02
c	200	13	21.3	440	25	T43416	Rat heavy chain varia	3.03e+02
c	201	13	21.3	446	13	Q80364	C-gamma-1 hinge and C	3.03e+02
c	202	13	21.3	446	13	Q80431	C-gamma-1 hinge + CH2	3.03e+02
	203	13	21.3	513	19	T20793	Human gene signature	3.03e+02
	204	13	21.3	535	1	Q04279	gp120 HIV envelope pr	3.03e+02
	205	13	21.3	536	1	N92108	Synthetic polynucleot	3.03e+02
	206	13	21.3	536	1	N90626	N-terminal region of	3.03e+02
	207	13	21.3	537	1	N90739	DNA encoding HIV port	3.03e+02
	208	13	21.3	537	1	N92116	Polynucleotide encodi	3.03e+02
	209	13	21.3	575	14	Q81746	Seq 3, a gene region	3.03e+02
c	210	13	21.3	577	37	T96142	Human plasma platelet	3.03e+02
c	211	13	21.3	577	34	T87059	Human platelet-activa	3.03e+02
c	212	13	21.3	577	29	T66132	PAF-AH exon 8 and par	3.03e+02
c	213	13	21.3	577	33	T80575	Human plasma platelet	3.03e+02
c	214	13	21.3	577	15	Q87960	Human acetyl hydrolas	3.03e+02
	215	13	21.3	580	2	N71255	Sequence of HTLV-III	3.03e+02
c	216	13	21.3	585	1	Q05024	Sequence encoding int	3.03e+02
c	217	13	21.3	606	24	T38524	Chlamydia trachomatis	3.03e+02
	218	13	21.3	615	35	T90427	K-ras genomic fragmen	3.03e+02
c	219	13	21.3	616	24	T38525	Chlamydia trachomatis	3.03e+02
c	220	13	21.3	663	24	T38527	Chlamydia trachomatis	3.03e+02
c	221	13	21.3	683	24	T38526	Chlamydia trachomatis	3.03e+02
	222	13	21.3	687	1	N90625	Sequence encoding PB1	3.03e+02
	223	13	21.3	687	1	Q04276	PB1mn, a HIV fusion p	3.03e+02
	224	13	21.3	693	1	Q04277	PB1sc, a HIV fusion p	3.03e+02
c	225	13	21.3	693	15	Q87592	Human IgG1 antibody F	3.03e+02
c	226	13	21.3	693	24	T38522	Chlamydia trachomatis	3.03e+02
c	227	13	21.3	693	31	T71691	Human IgG1 antibody F	3.03e+02
	228	13	21.3	694	1	N90628	Sequence encoding PB1	3.03e+02
c	229	13	21.3	699	40	V10199	Stealth virus nucleic	3.03e+02
c	230	13	21.3	699	9	Q53539	cDNA sequence for a h	3.03e+02
c	231	13	21.3	699	36	T80158	Human IgG1 hinge/Fc r	3.03e+02
c	232	13	21.3	699	36	T94928	Human immunoglobulin	3.03e+02
	233	13	21.3	702	6	Q35897	Recombinant PB1 fusio	3.03e+02
	234	13	21.3	702	1	N92112	Polynucleotide encodi	3.03e+02
	235	13	21.3	702	1	N81450	Sequence encoding fus	3.03e+02
c	236	13	21.3	704	24	T13819	Chlamydia trachomatis	3.03e+02
c	237	13	21.3	705	18	Q74083	Immunoglobulin G1 Fc	3.03e+02
c	238	13	21.3	705	23	T40788	Fc region of human Ig	3.03e+02
c	239	13	21.3	705	7	Q42589	Human Fc polypeptide	3.03e+02
c	240	13	21.3	740	35	T93783	IgG1 Fc cDNA which is	3.03e+02
c	241	13	21.3	740	7	Q41508	Human IgG1 Fc.	3.03e+02
c	242	13	21.3	745	10	Q55943	Polylinker/Fc fragmen	3.03e+02
c	243	13	21.3	745	16	Q92273	Fc cDNA.	3.03e+02
c	244	13	21.3	745	9	Q58428	Human IgG1 Fc-polylin	3.03e+02
c	245	13	21.3	745	35	T91030	Human IgG1 Fc region	3.03e+02
c	246	13	21.3	745	13	Q75425	Antibody Fc mutein co	3.03e+02

	247	13	21.3	750	20	Q81224	Del-gp120-VCN region	3.03e+02
c	248	13	21.3	765	1	N82142	Immunoglobulin G Fc d	3.03e+02
c	249	13	21.3	765	2	N70257	Part of the nucleotid	3.03e+02
c	250	13	21.3	767	31	T69219	Human IgG1 Fc DNA.	3.03e+02
c	251	13	21.3	768	21	T10780	Coding sequence for I	3.03e+02
c	252	13	21.3	768	17	Q96101	IgG1 hinge, CH2, CH3	3.03e+02
c	253	13	21.3	773	4	N70211	Sequence of 5' flanki	3.03e+02
c	254	13	21.3	773	2	N71219	Sequence of the 5' fl	3.03e+02
c	255	13	21.3	773	4	N70212	Sequence of 5' flanki	3.03e+02
c	256	13	21.3	773	2	N71220	Sequence of the 5' fl	3.03e+02
	257	13	21.3	785	29	T41519	Dirofilaria immitis D	3.03e+02
c	258	13	21.3	792	30	T67429	H. pylori cytoplasmic	3.03e+02
	259	13	21.3	797	17	T01543	Retinoblastoma protei	3.03e+02
	260	13	21.3	804	40	V24742	H. pylori ORF 07ap806	3.03e+02
	261	13	21.3	852	38	T88140	E. coli AIDA-I autotr	3.03e+02
c	262	13	21.3	854	29	T41627	Tobacco plant xyloglu	3.03e+02
	263	13	21.3	856	24	T42693	Superantigen gene.	3.03e+02
c	264	13	21.3	884	2	N70398	Portion of plasmid pP	3.03e+02
	265	13	21.3	885	10	Q62008	SSO2 gene.	3.03e+02
	266	13	21.3	891	2	N71251	Sequence encoding HTL	3.03e+02
	267	13	21.3	904	38	V06139	Viral infection gene	3.03e+02
	268	13	21.3	909	19	T11233	Coding region for pro	3.03e+02
c	269	13	21.3	945	30	T68066	H. pylori secreted or	3.03e+02
	270	13	21.3	960	2	N71257	Sequence encoding HTL	3.03e+02
c	271	13	21.3	975	24	T39912	Maize acetyl CoA carb	3.03e+02
	272	13	21.3	981	4	N30045	Sequence encoding bov	3.03e+02
c	273	13	21.3	984	32	T60718	cDNA encoding IgG4 he	3.03e+02
c	274	13	21.3	1006	18	T12663	Synthetic IgG4 cDNA.	3.03e+02
c	275	13	21.3	1050	7	Q40423	DNA sequence of LFA3T	3.03e+02
c	276	13	21.3	1050	5	Q28684	LFA-3-Ig fusion gene	3.03e+02
	277	13	21.3	1056	1	N92109	Synthetic polynucleot	3.03e+02
	278	13	21.3	1059	1	N92117	Polynucleotide encodi	3.03e+02
c	279	13	21.3	1129	26	T45262	Heregulin-alpha fusio	3.03e+02
c	280	13	21.3	1135	1	N90736	DNA encoding linkered	3.03e+02
c	281	13	21.3	1136	1	N90779	Sequence of the linke	3.03e+02
c	282	13	21.3	1149	18	T12659	IL4.Y124D/IgG4 gene f	3.03e+02
c	283	13	21.3	1149	18	T12660	IL4.Y124D/IgG4 PE gen	3.03e+02
c	284	13	21.3	1164	18	T12658	IL4.Y124D/IgG1 gene f	3.03e+02
	285	13	21.3	1171	11	Q68715	GDI D4 cDNA.	3.03e+02
c	286	13	21.3	1174	9	Q51547	Human kappa immunoglo	3.03e+02
	287	13	21.3	1188	1	N82040	La gene.	3.03e+02
c	288	13	21.3	1188	32	T60717	cDNA encoding leptin	3.03e+02
c	289	13	21.3	1188	32	T60719	cDNA encoding leptin	3.03e+02
c	290	13	21.3	1196	32	T60720	cDNA encoding leptin	3.03e+02
c	291	13	21.3	1203	32	T60721	cDNA encoding leptin	3.03e+02
	292	13	21.3	1227	23	T37306	Murine gonadotropin-r	3.03e+02
	293	13	21.3	1227	9	Q54832	Sequence of clone W82	3.03e+02
c	294	13	21.3	1231	31	T71692	Human Fas/Fc fusion p	3.03e+02
c	295	13	21.3	1231	15	Q87593	Human Fas/Fc fused OR	3.03e+02
	296	13	21.3	1269	2	N90738	DNA encoding HIV port	3.03e+02
	297	13	21.3	1269	1	N92115	Polynucleotide encodi	3.03e+02

	298	13	21.3	1271	23	T34991	Vascular IBP-like gro	3.03e+02
c	299	13	21.3	1272	29	T63109	Human soluble kit lig	3.03e+02
c	300	13	21.3	1272	29	T63110	Human soluble kit lig	3.03e+02
	301	13	21.3	1276	1	N92107	Synthetic polynucleot	3.03e+02
c	302	13	21.3	1302	34	T69920	DNA encoding interfer	3.03e+02
c	303	13	21.3	1309	16	Q94779	Regulator of neuronal	3.03e+02
c	304	13	21.3	1317	17	T00829	Plasmid pDC406/OX40/F	3.03e+02
c	305	13	21.3	1329	29	T61281	Humanised anti-L-sele	3.03e+02
c	306	13	21.3	1347	12	Q67347	VCAM 2D-IgG.	3.03e+02
	307	13	21.3	1355	1	N82025	Fragment containing r	3.03e+02
c	308	13	21.3	1357	31	T71700	Human deoxycytidylate	3.03e+02
c	309	13	21.3	1374	40	V24781	H. pylori ORF hp2e109	3.03e+02
c	310	13	21.3	1374	30	T68217	H. pylori cytoplasmic	3.03e+02
c	311	13	21.3	1386	8	Q49834	Anti-HIV-1 recombinan	3.03e+02
c	312	13	21.3	1389	29	T62931	2A2 (Chimeric) human	3.03e+02
	313	13	21.3	1392	33	T80053	Rice actin 1 gene pro	3.03e+02
c	314	13	21.3	1392	30	T62936	3F4 (Chimeric) human	3.03e+02
c	315	13	21.3	1392	30	T62930	Murine anti-porcine V	3.03e+02
c	316	13	21.3	1395	30	T62935	Murine anti-porcine V	3.03e+02
c	317	13	21.3	1404	33	T62868	Human gamma-4 heavy c	3.03e+02
c	318	13	21.3	1404	33	T62869	Human gamma-4E heavy	3.03e+02
c	319	13	21.3	1404	33	T62870	Human gamma-4PE heavy	3.03e+02
c	320	13	21.3	1412	4	Q25692	Sequence of the chime	3.03e+02
c	321	13	21.3	1418	22	T26889	Anti-rhesus D recombi	3.03e+02
c	322	13	21.3	1425	7	Q41516	Human CD40-L/Fc fusio	3.03e+02
c	323	13	21.3	1428	37	T96359	Soluble B7-1-Ig cDNA.	3.03e+02
c	324	13	21.3	1428	29	T61241	Human anti-RSV monocl	3.03e+02
c	325	13	21.3	1428	29	T61279	Human anti-RSV monocl	3.03e+02
c	326	13	21.3	1428	32	T75146	Human ATP receptor cD	3.03e+02
	327	13	21.3	1431	1	N90108	Human tissue factor i	3.03e+02
c	328	13	21.3	1431	29	T62513	Primatised anti-human	3.03e+02
c	329	13	21.3	1431	20	T18059	Monoclonal antibody D	3.03e+02
c	330	13	21.3	1431	29	T62510	Primatised anti-human	3.03e+02
	331	13	21.3	1432	10	Q44494	Bacillus coagulans pr	3.03e+02
c	332	13	21.3	1437	29	T13847	Primatised anti-human	3.03e+02
c	333	13	21.3	1446	39	T97439	Human cytokine recept	3.03e+02
	334	13	21.3	1448	34	T69574	Mouse tumour suscepti	3.03e+02
c	335	13	21.3	1458	4	Q23571	Reshaped CD4 antibody	3.03e+02
c	336	13	21.3	1458	4	Q23581	Reshaped CD4 antibody	3.03e+02
c	337	13	21.3	1459	14	Q93128	XhoI-EagI DNA fragmen	3.03e+02
c	338	13	21.3	1467	4	Q23570	Reshaped CAMPATH-1 an	3.03e+02
	339	13	21.3	1484	13	Q73688	HIV virus-1(LAI) gp12	3.03e+02
	340	13	21.3	1484	13	Q73690	HIV virus-1(LAI) gp12	3.03e+02
	341	13	21.3	1497	5	Q30798	Soluble HIV gp120.	3.03e+02
c	342	13	21.3	1497	7	Q41000	Sequence encoding hem	3.03e+02
c	343	13	21.3	1498	7	Q43504	Sequence encoding hae	3.03e+02
c	344	13	21.3	1498	24	T38359	Treponema hyodysenter	3.03e+02
c	345	13	21.3	1510	31	T69217	Retrovirus vector pLN	3.03e+02
c	346	13	21.3	1528	31	T69216	Retrovirus vector pLN	3.03e+02
	347	13	21.3	1539	5	Q30971	HIV gp160-MN precurso	3.03e+02
c	348	13	21.3	1549	3	Q20066	Encodes heavy chain o	3.03e+02

c	349	13	21.3	1557	10	Q45225	Sequence encoding a r	3.03e+02
c	350	13	21.3	1560	40	V03229	DNA encoding CD80-Ig-	3.03e+02
	351	13	21.3	1568	14	Q80575	HIV-1 env gene.	3.03e+02
	352	13	21.3	1568	23	T33295	HIV-1 env gene.	3.03e+02
	353	13	21.3	1570	13	Q73692	HIV virus-1(LAI) gp12	3.03e+02
	354	13	21.3	1576	6	Q40415	cDNA clone of pTOMUC8	3.03e+02
c	355	13	21.3	1576	8	Q49944	Human anti-HBs heavy	3.03e+02
c	356	13	21.3	1587	16	T02490	P-selectin ligand/Pc	3.03e+02
c	357	13	21.3	1587	10	Q63958	P-selectin ligand-IgG	3.03e+02
	358	13	21.3	1594	29	T62392	Direct tandem repeat	3.03e+02
	359	13	21.3	1599	37	V01073	Bovine P58 gene 3' un	3.03e+02
	360	13	21.3	1605	20	T29414	DGAL4-DT-IL-2 gene in	3.03e+02
c	361	13	21.3	1617	6	Q35099	Antibody D heavy chai	3.03e+02
c	362	13	21.3	1641	9	Q54655	chiT84.12 H3 heavy ch	3.03e+02
	363	13	21.3	1671	4	Q24134	50 kD subunit of SceI	3.03e+02
c	364	13	21.3	1680	34	T90413	Human neutral amino a	3.03e+02
	365	13	21.3	1697	17	T01541	Retinoblastoma protei	3.03e+02
	366	13	21.3	1698	29	T62390	Antisense sequence to	3.03e+02
c	367	13	21.3	1705	35	T80133	CTLA4-modified IgG1 f	3.03e+02
c	368	13	21.3	1717	17	T01586	Rat neu oncogene (Erb	3.03e+02
c	369	13	21.3	1747	35	T80131	CTLA4-IgG4 fusion pro	3.03e+02
c	370	13	21.3	1747	35	T80132	CTLA4-modified IgG4 f	3.03e+02
c	371	13	21.3	1765	11	Q71873	Sequence coding human	3.03e+02
	372	13	21.3	1791	1	N92118	Sequence encoding HIV	3.03e+02
	373	13	21.3	1791	1	N90741	DNA encoding HIV port	3.03e+02
	374	13	21.3	1796	1	N92110	Synthetic polynucleot	3.03e+02
c	375	13	21.3	1796	10	Q57750	CD4-gamma 2 chimeric	3.03e+02
c	376	13	21.3	1796	4	Q28088	Encodes CD4-gamma2 ch	3.03e+02
c	377	13	21.3	1806	4	Q27830	Sequence encoding one	3.03e+02
	378	13	21.3	1816	25	T42315	Oilpalm ACP thioester	3.03e+02
c	379	13	21.3	1837	35	T72174	Alzheimer's disease D	3.03e+02
	380	13	21.3	1862	20	T29413	DGAL4-DT-scFv(FRP5) g	3.03e+02
c	381	13	21.3	1863	29	T47705	Human FRX1 gene.	3.03e+02
c	382	13	21.3	1894	28	T60740	Alpha-1-acid glycopro	3.03e+02
	383	13	21.3	1919	20	T29412	DGAL4-DT-scFv(FRP5) g	3.03e+02
	384	13	21.3	1920	1	N81452	Sequence encoding fus	3.03e+02
	385	13	21.3	1920	1	N92114	Polynucleotide encodi	3.03e+02
c	386	13	21.3	1956	35	T90512	DNA encoding chimeric	3.03e+02
c	387	13	21.3	1975	1	Q05057	DraI/DraI fragment of	3.03e+02
c	388	13	21.3	1979	3	Q20293	SEQ ID No. 4 that cod	3.03e+02
c	389	13	21.3	1996	31	T71900	Human purinergic rece	3.03e+02
c	390	13	21.3	2008	3	Q20294	SEQ ID No. 5 that cod	3.03e+02
c	391	13	21.3	2009	21	T27385	Human immunoglobulin	3.03e+02
	392	13	21.3	2020	18	T04707	ALVAC recombinant vCP	3.03e+02
	393	13	21.3	2028	18	T04708	ALVAC recombinant vCP	3.03e+02
	394	13	21.3	2028	2	N70080	Human recombinant pro	3.03e+02
c	395	13	21.3	2029	5	Q30909	pBAG101 insert.	3.03e+02
	396	13	21.3	2040	5	Q31246	Chalcone synthase pro	3.03e+02
c	397	13	21.3	2043	9	Q55340	tICAM(453)IgG immunoa	3.03e+02
	398	13	21.3	2060	18	T04709	COPAK recombinant vCP	3.03e+02
	399	13	21.3	2074	24	T41420	Yeast cell wall prote	3.03e+02

	400	13	21.3	2076	14	Q83531	Sequence encoding bon	3.03e+02
c	401	13	21.3	2079	35	T90513	DNA encoding chimeric	3.03e+02
	402	13	21.3	2084	2	Q13865	PRP 140 gene and regu	3.03e+02
c	403	13	21.3	2092	1	Q04091	2.1 kb XbaI-BamHI fra	3.03e+02
c	404	13	21.3	2099	13	Q72494	Murine tumour rejecti	3.03e+02
c	405	13	21.3	2099	5	Q32369	smage-II genomic DNA.	3.03e+02
c	406	13	21.3	2099	16	T01173	SMAGE II gene.	3.03e+02
c	407	13	21.3	2150	5	Q32368	smage-I genomic DNA.	3.03e+02
c	408	13	21.3	2150	4	N90360	Sequence of a genetic	3.03e+02
c	409	13	21.3	2150	16	T01172	SMAGE I gene.	3.03e+02
c	410	13	21.3	2150	13	Q72493	Murine tumour rejecti	3.03e+02
c	411	13	21.3	2167	37	T94055	Human multidrug resis	3.03e+02
	412	13	21.3	2199	33	T80052	Rice actin 1 gene pro	3.03e+02
	413	13	21.3	2229	3	Q21883	BamHI/KpnI pSZ-4 clon	3.03e+02
	414	13	21.3	2263	1	N90398	HIV-1 env gene segmen	3.03e+02
	415	13	21.3	2271	20	T10279	Bombyx mori LPS-bindi	3.03e+02
c	416	13	21.3	2287	28	T60739	IgG1 gene.	3.03e+02
c	417	13	21.3	2287	4	Q25443	Sequence encoding ant	3.03e+02
c	418	13	21.3	2334	29	T66428	cDNA encoding fusion	3.03e+02
c	419	13	21.3	2364	2	Q12632	CD4-specific CDR-graf	3.03e+02
	420	13	21.3	2397	1	N81441	Clone encoding bovine	3.03e+02
c	421	13	21.3	2399	13	Q75356	OKT3 monoclonal antib	3.03e+02
c	422	13	21.3	2430	2	N60127	Sequence encoding Hum	3.03e+02
c	423	13	21.3	2430	16	Q74051	Human interleukin-1-a	3.03e+02
c	424	13	21.3	2441	14	Q55144	Enterococcus faecalis	3.03e+02
	425	13	21.3	2454	23	T32864	Tomato P119 gene prom	3.03e+02
	426	13	21.3	2457	1	N81451	Sequence encoding fus	3.03e+02
c	427	13	21.3	2482	4	Q28089	Encodes CD4-IgG2 chim	3.03e+02
c	428	13	21.3	2482	10	Q57751	CD4-IgG2 chimeric hea	3.03e+02
c	429	13	21.3	2488	4	Q27831	Sequence encoding CD4	3.03e+02
	430	13	21.3	2492	3	Q14755	FUS2 gene.	3.03e+02
	431	13	21.3	2508	31	T67197	Zebrafish retinoid X	3.03e+02
	432	13	21.3	2511	1	N92113	Polynucleotide encodi	3.03e+02
	433	13	21.3	2526	29	T62395	Construct pMK10/AS15.	3.03e+02
c	434	13	21.3	2560	5	Q30910	pMDR1002 insert, pre-	3.03e+02
	435	13	21.3	2562	15	Q89817	DNA encoding the AUF1	3.03e+02
	436	13	21.3	2571	18	T09010	HIV-1 strain MN rev a	3.03e+02
	437	13	21.3	2571	14	Q76020	DNA encoding gp120 fr	3.03e+02
	438	13	21.3	2574	14	Q76019	DNA encoding gp120 fr	3.03e+02
	439	13	21.3	2581	12	Q70104	Bovine enterokinase.	3.03e+02
c	440	13	21.3	2589	4	N90357	Sequence of a genetic	3.03e+02
	441	13	21.3	2607	2	N60077	DNA sequence of the e	3.03e+02
	442	13	21.3	2622	35	T90461	Thermoanaerobacter th	3.03e+02
	443	13	21.3	2641	5	Q30970	Non-cleavable, solubl	3.03e+02
	444	13	21.3	2673	1	N81449	Sequence encoding fus	3.03e+02
c	445	13	21.3	2688	35	T72165	Alzheimer's disease D	3.03e+02
c	446	13	21.3	2700	31	T73281	S. cerevisiae origin	3.03e+02
	447	13	21.3	2712	18	T06613	Autotaxin coding sequ	3.03e+02
c	448	13	21.3	2726	36	T97875	Watermellon seed nuce	3.03e+02
	449	13	21.3	2730	32	T58778	Human eotaxin genomic	3.03e+02
c	450	13	21.3	2759	15	Q91288	ORC3 gene encoding a	3.03e+02

c	451	13	21.3	2760	28	T61883	Marek's disease virus	3.03e+02
c	452	13	21.3	2828	3	Q15171	Bacillus thuringiensi	3.03e+02
	453	13	21.3	2868	10	Q58007	Sequence of plasmid p	3.03e+02
	454	13	21.3	2945	16	T05127	HIV virus-1 gp160 env	3.03e+02
	455	13	21.3	2946	18	T06612	Autotaxin coding sequ	3.03e+02
c	456	13	21.3	3001	33	T61583	Malolactic enzyme and	3.03e+02
c	457	13	21.3	3001	6	Q34589	A6 fragment of MDV BC	3.03e+02
c	458	13	21.3	3001	17	T02338	Marek's disease type	3.03e+02
c	459	13	21.3	3076	21	T27391	Human IgG CH1-hinge-C	3.03e+02
	460	13	21.3	3084	39	V09365	HIV-1/CD4 chimeric nu	3.03e+02
c	461	13	21.3	3095	18	T02918	Human glucose-6-phosp	3.03e+02
c	462	13	21.3	3095	18	T02920	Human glucose-6-phosp	3.03e+02
c	463	13	21.3	3095	18	T02921	Human glucose-6-phosp	3.03e+02
c	464	13	21.3	3095	18	T02922	Human glucose-6-phosp	3.03e+02
c	465	13	21.3	3097	18	T02919	Human glucose-6-phosp	3.03e+02
c	466	13	21.3	3108	21	T27388	Human IgG CH1-hinge-C	3.03e+02
	467	13	21.3	3125	13	Q73686	HIV virus-1(LAI) gp12	3.03e+02
c	468	13	21.3	3127	21	T27392	Human IgG CH1-hinge-C	3.03e+02
c	469	13	21.3	3133	4	N90356	Sequence of a genetic	3.03e+02
c	470	13	21.3	3147	21	T27389	Human IgG CH1-hinge-C	3.03e+02
	471	13	21.3	3156	2	N60128	Sequence of the envel	3.03e+02
c	472	13	21.3	3282	24	T40914	DNA encoding VH regio	3.03e+02
	473	13	21.3	3290	15	Q86348	Human protein S cDNA	3.03e+02
	474	13	21.3	3312	13	Q79326	Human ARD 1.	3.03e+02
c	475	13	21.3	3312	2	N71034	pMx34 cDNA insert.	3.03e+02
	476	13	21.3	3326	29	T62397	Construct pMK10/AS15X	3.03e+02
	477	13	21.3	3344	1	N81442	Clone encoding human	3.03e+02
	478	13	21.3	3387	20	T09021	Arabidopsis thaliana	3.03e+02
c	479	13	21.3	3400	30	T62937	3F4 human G2/G4 chime	3.03e+02
c	480	13	21.3	3400	30	T62932	2A2 human G2/G4 chime	3.03e+02
	481	13	21.3	3499	29	T62393	Construct pMK10/AS3'U	3.03e+02
	482	13	21.3	3588	21	T15143	HIV-1 gp160-NT5 p55 g	3.03e+02
	483	13	21.3	3641	40	V03351	Homo sapiens interleu	3.03e+02
c	484	13	21.3	3808	18	T04701	pHIV32 fragment contg	3.03e+02
	485	13	21.3	3821	9	Q56609	Human PGF-2/NT-3 5' f	3.03e+02
	486	13	21.3	3917	22	T31723	K-ras oncogene.	3.03e+02
	487	13	21.3	3975	3	Q22999	SERP gene.	3.03e+02
	488	13	21.3	3975	1	N81157	Malaria-specific gene	3.03e+02
	489	13	21.3	4020	2	N71016	Sequence of LAV/HTLV	3.03e+02
c	490	13	21.3	4575	1	N91465	Drosophila melanogast	3.03e+02
c	491	13	21.3	4699	38	T49500	Transferrin receptor	3.03e+02
c	492	13	21.3	4699	21	Q94442	Bacterial transferrin	3.03e+02
c	493	13	21.3	4723	34	T78802	Gamma heavy chain pla	3.03e+02
	494	13	21.3	4940	15	Q91286	ORC1 gene encoding a	3.03e+02
	495	13	21.3	4940	31	T73279	S. cerevisiae origin	3.03e+02
c	496	13	21.3	4954	14	Q55753	Escherichia coli geno	3.03e+02
c	497	13	21.3	4990	27	T48800	Plasmid mtalsEpoRFc e	3.03e+02
c	498	13	21.3	5009	21	Q94444	Bacterial transferrin	3.03e+02
c	499	13	21.3	5009	38	T49502	Transferrin receptor	3.03e+02
c	500	13	21.3	5033	38	T49501	Transferrin receptor	3.03e+02
c	501	13	21.3	5033	21	Q94443	Bacterial transferrin	3.03e+02

	502	13	21.3	5072	31	T71315	cDNA encoding C. eleg	3.03e+02
	503	13	21.3	5073	31	T71314	cDNA encoding C. eleg	3.03e+02
c	504	13	21.3	5099	38	T49503	Transferrin receptor	3.03e+02
c	505	13	21.3	5099	21	Q94445	Bacterial transferrin	3.03e+02
c	506	13	21.3	5144	38	T49504	Transferrin receptor	3.03e+02
c	507	13	21.3	5144	21	Q94446	Bacterial transferrin	3.03e+02
c	508	13	21.3	5250	30	T62933	2A2 human IgG4 expres	3.03e+02
c	509	13	21.3	5283	35	T91972	DNA encoding rat thro	3.03e+02
	510	13	21.3	5300	35	T90447	Thermoanaerobacter th	3.03e+02
c	511	13	21.3	5300	30	T62938	3F4 human IgG4 expres	3.03e+02
c	512	13	21.3	5337	31	T69893	DNA for aggrecanase a	3.03e+02
c	513	13	21.3	5337	31	T69892	DNA for aggrecanase a	3.03e+02
	514	13	21.3	5408	15	Q95494	Human Cdn-3 DNA.	3.03e+02
	515	13	21.3	5476	29	T60078	Human SMCY cDNA.	3.03e+02
c	516	13	21.3	5495	35	T75590	Canine herpesvirus ge	3.03e+02
	517	13	21.3	5510	2	Q12707	Rice actin gene RAc1.	3.03e+02
	518	13	21.3	5643	33	T80055	Rice actin 1 gene.	3.03e+02
c	519	13	21.3	5648	1	Q03368	Sequence of human lys	3.03e+02
	520	13	21.3	6124	1	Q03568	Sequence encoding the	3.03e+02
c	521	13	21.3	6285	40	V03503	Plasmid pCDIL5HEHC1.0	3.03e+02
c	522	13	21.3	6285	22	T34109	Anti-IL-5 humanised a	3.03e+02
c	523	13	21.3	6327	22	T32301	Dermatomyositis speci	3.03e+02
c	524	13	21.3	6367	18	T12661	Vector COSFcLink.	3.03e+02
c	525	13	21.3	6367	27	T48803	Vector COSFcLink cont	3.03e+02
c	526	13	21.3	6474	7	Q40300	Sequence of plasmid p	3.03e+02
c	527	13	21.3	6557	18	T15932	Anti-IgE VH expressio	3.03e+02
c	528	13	21.3	6608	27	T42751	Chicken CHD-1A gene.	3.03e+02
c	529	13	21.3	6811	7	Q41004	Sequence of plasmid p	3.03e+02
c	530	13	21.3	6889	18	T15931	DHFR/intron (WTrasSD)	3.03e+02
	531	13	21.3	6926	7	Q40281	Sequence of clone pP2	3.03e+02
c	532	13	21.3	6926	18	T12662	IL-4.Y124D/IgG1 fusio	3.03e+02
c	533	13	21.3	7073	27	T50962	TF8-5G9 CDR-grafted h	3.03e+02
c	534	13	21.3	7127	36	T85609	Human obesity protein	3.03e+02
c	535	13	21.3	7127	32	T75093	Human obesity protein	3.03e+02
c	536	13	21.3	7167	40	V03228	Plasmid CD80Fcalphatp	3.03e+02
	537	13	21.3	7474	31	T71320	Plasmid pTB73 encodin	3.03e+02
	538	13	21.3	7625	31	T71323	Hybrid activation vec	3.03e+02
c	539	13	21.3	7653	9	Q64740	ced-3 (C6322T) gene.	3.03e+02
c	540	13	21.3	7653	9	Q64739	ced-3 (G6297A) gene.	3.03e+02
c	541	13	21.3	7653	9	Q64737	ced-3 (G5757A) gene.	3.03e+02
c	542	13	21.3	7653	9	Q64738	ced-3 (C5940T) gene.	3.03e+02
c	543	13	21.3	7653	9	Q64736	ced-3 (G2487A) gene.	3.03e+02
c	544	13	21.3	7653	9	Q54401	ced-3 gene.	3.03e+02
c	545	13	21.3	7653	9	Q54666	ced-3 gene.	3.03e+02
c	546	13	21.3	7653	9	Q64745	ced-3 (C7020T) gene.	3.03e+02
c	547	13	21.3	7653	9	Q64741	ced-3 (G6372A) gene.	3.03e+02
c	548	13	21.3	7653	9	Q64743	ced-3 (C6485T) gene.	3.03e+02
c	549	13	21.3	7653	9	Q64742	ced-3 (C6434T) gene.	3.03e+02
c	550	13	21.3	7653	24	T38196	Nematode Ced-3 gene.	3.03e+02
c	551	13	21.3	7653	9	Q64744	ced-3 (G6536A) gene.	3.03e+02
c	552	13	21.3	7745	28	T59308	Sh2-mlRev6 gene (shru	3.03e+02

c	553	13	21.3	7745	34	T79859	Genomic DNA encoding	3.03e+02
	554	13	21.3	8355	24	T35524	Human interferon-beta	3.03e+02
	555	13	21.3	8395	1	Q05056	BIV 106 viral genome.	3.03e+02
	556	13	21.3	8482	1	Q03642	BIV 127 viral genome.	3.03e+02
c	557	13	21.3	8540	11	Q65628	Vector contg. TCAE 8	3.03e+02
c	558	13	21.3	8654	14	Q55138	Staphylococcus epider	3.03e+02
c	559	13	21.3	8841	14	Q89299	Tryptophan-2,3-dioxyg	3.03e+02
	560	13	21.3	9088	2	N60476	Sequence of lymphaden	3.03e+02
	561	13	21.3	9193	2	N60365	Sequence of LAV virus	3.03e+02
c	562	13	21.3	9201	9	Q55003	Humanised anti-CD18 A	3.03e+02
c	563	13	21.3	9208	11	Q65629	Vector contg. TCAE 8	3.03e+02
	564	13	21.3	9213	2	N60288	Sequence of the HTLV-	3.03e+02
	565	13	21.3	9278	13	Q76124	Human MDC genomic DNA	3.03e+02
	566	13	21.3	9425	7	Q45919	HTLV-III DNA (Clone B	3.03e+02
	567	13	21.3	9425	7	Q45921	HTLV-III DNA (Clone B	3.03e+02
	568	13	21.3	9427	7	Q45920	HTLV-III DNA (Clone B	3.03e+02
c	569	13	21.3	9431	3	Q20907	DNA encoding the chim	3.03e+02
	570	13	21.3	9642	31	T71324	Plasmid pTB113 used f	3.03e+02
	571	13	21.3	9709	3	Q22488	HIV-1 proviral clone	3.03e+02
	572	13	21.3	9709	16	Q96140	HIV-1 NL4-3 genomic D	3.03e+02
	573	13	21.3	9739	27	T58549	Human Immunodeficienc	3.03e+02
	574	13	21.3	9739	3	Q14751	HIV-1(MN) env protein	3.03e+02
	575	13	21.3	9745	2	N60240	HTLV-III virus (HIV v	3.03e+02
	576	13	21.3	9746	27	T58550	Human Immunodeficienc	3.03e+02
	577	13	21.3	9746	3	Q14752	HIV-1(MN-ST1) env pro	3.03e+02
	578	13	21.3	9749	7	Q45922	HTLV-III DNA (Clone H	3.03e+02
c	579	13	21.3	9824	7	Q40763	Wild-type shrunken-2	3.03e+02
	580	13	21.3	10288	31	T71322	Plasmid pCB51 encodin	3.03e+02
c	581	13	21.3	10380	31	T67164	Human alpha-N-acetylg	3.03e+02
	582	13	21.3	10443	31	T71319	Plasmid pTB72 encodin	3.03e+02
c	583	13	21.3	10592	35	T75616	Canine herpesvirus ge	3.03e+02
c	584	13	21.3	10704	7	Q43845	Plasmid pAH4625.	3.03e+02
c	585	13	21.3	10844	7	Q43848	Plasmid pAH4808.	3.03e+02
c	586	13	21.3	11336	7	Q43497	Approximate nucleotid	3.03e+02
c	587	13	21.3	11529	7	Q43844	Plasmid pAH4602.	3.03e+02
c	588	13	21.3	12132	7	Q43847	Plasmid pAH4807.	3.03e+02
c	589	13	21.3	13254	24	T40915	Nucleotide sequence o	3.03e+02
	590	13	21.3	13414	31	T71321	Plasmid pCB50 encodin	3.03e+02
	591	13	21.3	13585	17	T11549	Tumour rejection anti	3.03e+02
c	592	13	21.3	14507	32	T73568	Expression augmenting	3.03e+02
	593	13	21.3	15581	39	V14355	Plasmid pNLnSG11 expr	3.03e+02
c	594	13	21.3	17041	3	Q21065	Genomic DNA of human	3.03e+02
c	595	13	21.3	17041	12	Q67057	PACPAP38 DNA.	3.03e+02
c	596	13	21.3	18177	1	N90490	DNA of human retinobl	3.03e+02
	597	13	21.3	21126	2	N60970	TL-DNA region of A.rh	3.03e+02
c	598	13	21.3	24417	38	T97221	Pseudomonas aeruginos	3.03e+02
	599	13	21.3	26698	20	T08126	Mouse syndecan-1 gene	3.03e+02
	600	13	21.3	26700	40	V15946	Mouse syndecan gene s	3.03e+02
	601	13	21.3	26700	11	Q67902	Syndecan gene.	3.03e+02
c	602	13	21.3	26764	31	T71696	Human deoxycytidylate	3.03e+02
c	603	13	21.3	29392	40	V15422	Mouse poly Ig recepto	3.03e+02

	604	13	21.3	29544	17	T12460	Ovine adenovirus geno	3.03e+02
	605	13	21.3	40352	40	V02032	MAGE-B cluster DNA se	3.03e+02
c	606	13	21.3	53526	40	T94101	Human PKD1 gene.	3.03e+02
c	607	13	21.3	53577	40	T94108	Human PKD1 locus betw	3.03e+02
c	608	13	21.3	53577	28	T18551	Human polycystic kidn	3.03e+02
c	609	13	21.3	134525	1	Q04525	Total base sequence o	3.03e+02
c	610	12	19.7	20	37	T97431	Oligomer sl4102 Commo	9.07e+02
c	611	12	19.7	20	37	T97419	Donor Sequence oligom	9.07e+02
c	612	12	19.7	30	17	T04616	Fibrinogen B-beta cha	9.07e+02
c	613	12	19.7	36	10	Q45296	Sequence of PCR prime	9.07e+02
c	614	12	19.7	40	37	T96164	5' intron sequence of	9.07e+02
c	615	12	19.7	41	22	T26748	Human gene signature	9.07e+02
	616	12	19.7	73	20	T23162	Human gene signature	9.07e+02
c	617	12	19.7	75	14	Q81862	Antisense strand of t	9.07e+02
c	618	12	19.7	77	21	T25380	Human gene signature	9.07e+02
	619	12	19.7	78	13	Q78302	Heat shock protein le	9.07e+02
	620	12	19.7	81	20	T22389	Human gene signature	9.07e+02
c	621	12	19.7	84	22	T26024	Human gene signature	9.07e+02
c	622	12	19.7	106	19	T20928	Human gene signature	9.07e+02
c	623	12	19.7	126	5	Q31719	Yeast artificial chro	9.07e+02
c	624	12	19.7	153	29	T61912	Rat cadherin-6 extrac	9.07e+02
	625	12	19.7	155	18	T19487	Human gene signature	9.07e+02
	626	12	19.7	156	28	T59834	Beta-casein gene frag	9.07e+02
	627	12	19.7	157	10	Q76578	Human genome fragment	9.07e+02
c	628	12	19.7	162	11	Q77255	Human genome fragment	9.07e+02
	629	12	19.7	164	5	Q33800	Sequence upstream fro	9.07e+02
c	630	12	19.7	164	20	T22276	Human gene signature	9.07e+02
	631	12	19.7	171	20	T22266	Human gene signature	9.07e+02
	632	12	19.7	173	2	Q11479	Fragment of plasmid p	9.07e+02
	633	12	19.7	184	18	T19273	Human gene signature	9.07e+02
	634	12	19.7	186	1	Q04497	Serine protease inhib	9.07e+02
c	635	12	19.7	194	4	Q24631	GSE H.	9.07e+02
	636	12	19.7	196	18	T19660	Human gene signature	9.07e+02
c	637	12	19.7	203	20	T24135	Human gene signature	9.07e+02
c	638	12	19.7	212	10	Q77454	Human genome fragment	9.07e+02
c	639	12	19.7	214	22	T24799	Human gene signature	9.07e+02
c	640	12	19.7	214	8	Q59962	Human brain Expressed	9.07e+02
c	641	12	19.7	216	20	T23448	Human gene signature	9.07e+02
c	642	12	19.7	223	6	Q39928	Expressed Sequence Ta	9.07e+02
c	643	12	19.7	223	8	Q59340	Human brain Expressed	9.07e+02
c	644	12	19.7	236	30	T43945	Sequence flanking mar	9.07e+02
	645	12	19.7	240	2	N60541	Sequence encoding the	9.07e+02
	646	12	19.7	248	21	T23344	Human gene signature	9.07e+02
c	647	12	19.7	249	19	T09728	Tomato genomic DNA, D	9.07e+02
	648	12	19.7	253	22	T25894	Human gene signature	9.07e+02
	649	12	19.7	270	1	N81568	Right half of gamma-i	9.07e+02
c	650	12	19.7	276	29	T67397	H. pylori cytoplasmic	9.07e+02
	651	12	19.7	283	22	T22933	Human gene signature	9.07e+02
c	652	12	19.7	288	20	T22186	Human gene signature	9.07e+02
	653	12	19.7	295	19	T21139	Human gene signature	9.07e+02
c	654	12	19.7	295	21	T25663	Human gene signature	9.07e+02

	655	12	19.7	307	22	T25167	Human gene signature	9.07e+02
c	656	12	19.7	313	4	Q26715	Encodes exon III of h	9.07e+02
	657	12	19.7	318	39	V09718	Chloroplast transit p	9.07e+02
	658	12	19.7	318	4	Q27200	CTP2.	9.07e+02
c	659	12	19.7	319	20	T23784	Human gene signature	9.07e+02
c	660	12	19.7	322	18	T19139	Human gene signature	9.07e+02
c	661	12	19.7	326	8	Q59874	Human brain Expressed	9.07e+02
c	662	12	19.7	327	30	T68037	H. pylori secreted or	9.07e+02
	663	12	19.7	335	20	T22250	Human gene signature	9.07e+02
	664	12	19.7	336	1	Q05969	Insert from plasmid p	9.07e+02
	665	12	19.7	336	1	N91056	Plasmid pSVMT-APPI ca	9.07e+02
c	666	12	19.7	356	40	V09977	Partial cDNA Incyte c	9.07e+02
c	667	12	19.7	356	8	Q60042	Human brain Expressed	9.07e+02
	668	12	19.7	368	23	T36780	Prostate-specific mem	9.07e+02
	669	12	19.7	372	4	Q25906	Encodes a new platele	9.07e+02
	670	12	19.7	372	4	Q25849	Anti-trypsin protein	9.07e+02
	671	12	19.7	372	3	Q14723	Amyloid precursor pro	9.07e+02
	672	12	19.7	372	4	Q25267	DNA encoding a monocl	9.07e+02
	673	12	19.7	372	3	Q14722	Amyloid precursor pro	9.07e+02
	674	12	19.7	372	3	Q14721	Amyloid precursor pro	9.07e+02
c	675	12	19.7	375	8	Q60808	Human brain Expressed	9.07e+02
	676	12	19.7	390	17	T03445	HAV VP-1 membrane pro	9.07e+02
c	677	12	19.7	413	20	T21451	Human gene signature	9.07e+02
c	678	12	19.7	423	40	V24615	H. pylori ORF 01ep305	9.07e+02
	679	12	19.7	439	14	Q85360	Platelet factor-4.	9.07e+02
	680	12	19.7	450	1	N81569	Left and right half o	9.07e+02
	681	12	19.7	457	1	N91053	Sequence encoding exo	9.07e+02
	682	12	19.7	459	19	T20506	Human gene signature	9.07e+02
	683	12	19.7	479	5	Q29876	Pheromone receptor c1	9.07e+02
c	684	12	19.7	491	35	T75665	Human preproamylin PC	9.07e+02
	685	12	19.7	494	1	Q06359	Sequence encoding sal	9.07e+02
	686	12	19.7	504	7	Q40784	DNA sequence used in	9.07e+02
c	687	12	19.7	513	22	T26145	Human gene signature	9.07e+02
	688	12	19.7	515	7	Q40786	DNA sequence used in	9.07e+02
	689	12	19.7	542	1	N81571	ClaI-BamHI restrictio	9.07e+02
c	690	12	19.7	552	34	T90023	DNA encoding light ch	9.07e+02
c	691	12	19.7	564	40	V24756	H. pylori ORF 01ce116	9.07e+02
c	692	12	19.7	564	34	T90021	DNA encoding light ch	9.07e+02
	693	12	19.7	567	19	T15469	Feline interferon-gam	9.07e+02
c	694	12	19.7	582	34	T90020	DNA encoding heavy ch	9.07e+02
c	695	12	19.7	585	34	T90022	DNA encoding heavy ch	9.07e+02
c	696	12	19.7	631	26	T28527	S. pneumoniae detecti	9.07e+02
c	697	12	19.7	648	32	T78410	Hepatocyte growth fac	9.07e+02
	698	12	19.7	681	1	Q04278	PBlwmj2, a HIV fusion	9.07e+02
c	699	12	19.7	684	4	Q25666	Sequence of the genom	9.07e+02
	700	12	19.7	687	1	Q04275	PBlrf, a HIV fusion p	9.07e+02
	701	12	19.7	705	30	T67925	H. pylori surface or	9.07e+02
	702	12	19.7	723	30	T67946	H. pylori secreted or	9.07e+02
	703	12	19.7	726	30	T67637	H. pylori transmembra	9.07e+02
	704	12	19.7	737	7	Q40778	Genomic DNA sequence	9.07e+02
c	705	12	19.7	756	12	Q73729	Rape abscission/dehis	9.07e+02

	706	12	19.7	771	1	Q06181	PR-1b cDNA cloned int	9.07e+02
	707	12	19.7	779	35	T79332	DNA encoding whale ma	9.07e+02
c	708	12	19.7	804	9	Q53898	Saporin coding sequen	9.07e+02
c	709	12	19.7	804	18	T17964	Saporin cDNA clone M1	9.07e+02
c	710	12	19.7	873	4	Q23662	Truncated hepatocyte	9.07e+02
	711	12	19.7	879	40	V03314	Methanococcus thermol	9.07e+02
c	712	12	19.7	885	30	T67976	H. pylori cytoplasmic	9.07e+02
c	713	12	19.7	893	15	Q92529	P. communis (pear) ar	9.07e+02
c	714	12	19.7	909	30	T68186	H. pylori transporter	9.07e+02
	715	12	19.7	970	35	T79969	Presenilin-interactin	9.07e+02
	716	12	19.7	980	20	T27713	Mouse pp32 cDNA clone	9.07e+02
c	717	12	19.7	1002	5	Q29855	Odorant receptor clon	9.07e+02
	718	12	19.7	1006	7	Q42493	Protein p22 cDNA.	9.07e+02
	719	12	19.7	1028	39	T73604	Tomato PR-glucanase g	9.07e+02
	720	12	19.7	1030	9	Q55113	Streptococcal fibrino	9.07e+02
c	721	12	19.7	1050	35	T79674	BRCA2 cancer suscepti	9.07e+02
	722	12	19.7	1061	23	T35200	Mouse interleukin-12	9.07e+02
c	723	12	19.7	1064	15	Q87093	Tpa-1 phage insert.	9.07e+02
c	724	12	19.7	1064	29	T63677	Peroxidase gene, Tpa-	9.07e+02
	725	12	19.7	1067	11	Q67906	Cotton fibre-specific	9.07e+02
	726	12	19.7	1067	31	T70040	Cotton fibre specific	9.07e+02
	727	12	19.7	1079	2	N60175	D-amino-acid-oxidase	9.07e+02
	728	12	19.7	1079	2	Q11478	Plasmid pUR3001 , enc	9.07e+02
	729	12	19.7	1083	28	T33166	Mutated OCIF, OCIF-DC	9.07e+02
	730	12	19.7	1098	21	T13659	AcNPV ORF 54, residue	9.07e+02
	731	12	19.7	1112	16	Q89727	Plasmodium falciparum	9.07e+02
	732	12	19.7	1119	2	N71222	Sequence of soybean l	9.07e+02
	733	12	19.7	1134	36	T85195	Plastid targeting reg	9.07e+02
	734	12	19.7	1134	36	T91893	Decaprenyl synthase g	9.07e+02
c	735	12	19.7	1144	38	T95782	Stylosanthes humilis	9.07e+02
c	736	12	19.7	1153	5	Q32848	Rat GAP-43 gene.	9.07e+02
	737	12	19.7	1166	29	T60972	Mouse lipopolysacchar	9.07e+02
	738	12	19.7	1173	28	T35475	Human tumour necrosis	9.07e+02
	739	12	19.7	1193	1	N90629	Sequence encoding bov	9.07e+02
	740	12	19.7	1206	28	T33161	Mutated OCIF, OCIF-Cl	9.07e+02
c	741	12	19.7	1233	16	Q92963	Human HRR25-like case	9.07e+02
c	742	12	19.7	1233	12	Q70828	Protein kinase (CKIga	9.07e+02
	743	12	19.7	1242	23	T36790	Prostate-specific mem	9.07e+02
c	744	12	19.7	1248	40	V24743	H. pylori ORF 07ce102	9.07e+02
	745	12	19.7	1260	1	Q03250	Sequence encoding bet	9.07e+02
c	746	12	19.7	1260	22	T17043	B.thuringiensis insec	9.07e+02
c	747	12	19.7	1287	37	T75506	P. americanus antifr	9.07e+02
	748	12	19.7	1302	24	T43045	Helicobacter pylori p	9.07e+02
	749	12	19.7	1314	30	T68169	H. pylori inner membr	9.07e+02
	750	12	19.7	1316	13	Q75377	Human elastase inhibi	9.07e+02
	751	12	19.7	1337	26	T28539	S. pyogenes detection	9.07e+02
c	752	12	19.7	1347	24	T41700	Brugia pahangi beta-t	9.07e+02
	753	12	19.7	1358	1	Q06208	Beta-1,3-glucanase cD	9.07e+02
	754	12	19.7	1358	2	Q13750	Plasmid pGLN17 encodi	9.07e+02
	755	12	19.7	1378	10	Q45241	Sequence encoding the	9.07e+02
	756	12	19.7	1407	1	Q03255	Sequence encoding bet	9.07e+02

	757	12	19.7	1425	2	N60144	Sequence that encodes 9.07e+02
	758	12	19.7	1425	6	Q31940	ARV-2 env gene. 9.07e+02
c	759	12	19.7	1429	40	V04237	Arabidopsis C-14 ster 9.07e+02
	760	12	19.7	1446	23	T37063	Streptococcus mutans 9.07e+02
	761	12	19.7	1446	39	W13454	Streptococcus mutans 9.07e+02
	762	12	19.7	1448	13	Q73689	HIV virus-1(JR-FL) gp 9.07e+02
	763	12	19.7	1448	13	Q73691	HIV virus-1(JR-FL) gp 9.07e+02
	764	12	19.7	1463	2	N71221	Sequence of soybean l 9.07e+02
c	765	12	19.7	1468	29	T66425	Human P28 coding sequ 9.07e+02
c	766	12	19.7	1511	40	T86645	Heterodera glycines c 9.07e+02
c	767	12	19.7	1527	30	T67337	H. pylori cytoplasmic 9.07e+02
	768	12	19.7	1538	2	Q10483	DNA methylase gene M. 9.07e+02
c	769	12	19.7	1542	3	Q20695	Mouse OSF.1. 9.07e+02
	770	12	19.7	1578	24	T30248	Cotton fibre clone SK 9.07e+02
	771	12	19.7	1578	18	T13028	Cotton fibre-specific 9.07e+02
	772	12	19.7	1578	31	T70034	Cotton E6-4A gene and 9.07e+02
c	773	12	19.7	1608	9	Q55607	Helianthus tuberosus 9.07e+02
	774	12	19.7	1618	31	T70033	Cotton E6-1A gene and 9.07e+02
	775	12	19.7	1618	18	T13027	Cotton fibre-specific 9.07e+02
	776	12	19.7	1624	23	T35310	Human SH-PTP1 cDNA cl 9.07e+02
	777	12	19.7	1647	18	T13025	Cotton fibre-specific 9.07e+02
	778	12	19.7	1647	24	T30245	Cotton fibre clone SK 9.07e+02
	779	12	19.7	1647	31	T70031	Cotton E6-2A gene and 9.07e+02
	780	12	19.7	1668	31	T70032	Cotton E6-3B gene and 9.07e+02
	781	12	19.7	1668	24	T30246	Cotton fibre clone SK 9.07e+02
c	782	12	19.7	1672	28	T62625	Cotton fibre specific 9.07e+02
c	783	12	19.7	1676	31	T73289	C. elegans origin of 9.07e+02
	784	12	19.7	1679	40	V03349	Homo sapiens interleu 9.07e+02
	785	12	19.7	1689	3	Q22205	Murine mast cell grow 9.07e+02
c	786	12	19.7	1772	19	T14045	1-Aminocyclopropane-1 9.07e+02
	787	12	19.7	1777	13	Q79355	Sequence of the exten 9.07e+02
	788	12	19.7	1783	13	Q74691	Early Ripening Tomato 9.07e+02
	789	12	19.7	1812	22	T27105	Yeast calcineurin sub 9.07e+02
	790	12	19.7	1827	40	V19700	Insecticidal toxin su 9.07e+02
c	791	12	19.7	1833	1	Q03578	Synthetic Bacillus th 9.07e+02
c	792	12	19.7	1833	17	T05738	B. thuringiensis var. 9.07e+02
c	793	12	19.7	1833	24	T40340	Synthetic Btt gene en 9.07e+02
c	794	12	19.7	1833	24	T40341	Native Btt gene encod 9.07e+02
	795	12	19.7	1841	38	V05504	DNA for BPN' substili 9.07e+02
	796	12	19.7	1868	22	T27626	Wild type subtilisin 9.07e+02
	797	12	19.7	1878	1	N90416	DNA sequence encoding 9.07e+02
c	798	12	19.7	1911	5	Q31874	Cyclin D2 gene. 9.07e+02
	799	12	19.7	1926	15	Q92366	Chick p78 cDNA. 9.07e+02
c	800	12	19.7	1932	8	Q47288	Delta endotoxin gene. 9.07e+02
c	801	12	19.7	1945	16	T02814	Spinach 2-oxoglutarat 9.07e+02
c	802	12	19.7	1949	34	T76612	Peanut allergen Ara h 9.07e+02
c	803	12	19.7	1953	24	T43223	Antiscarab pest toxin 9.07e+02
c	804	12	19.7	1977	1	N91495	Bacillus thuringiensi 9.07e+02
c	805	12	19.7	1984	18	T13030	Cotton fibre-specific 9.07e+02
c	806	12	19.7	1985	24	T30250	Cotton fibre clone SK 9.07e+02
c	807	12	19.7	2023	39	V06638	Arabidopsis thaliana 9.07e+02

c	808	12	19.7	2032	34	T76613	Peanut allergen Ara h 9.07e+02
c	809	12	19.7	2082	14	Q55135	Staphylococcus aureus 9.07e+02
c	810	12	19.7	2085	40	V04262	Restin truncated cDNA 9.07e+02
c	811	12	19.7	2108	1	N91116	Flagellin gene of E.c 9.07e+02
c	812	12	19.7	2109	2	N70985	E.coli flagellin gene 9.07e+02
c	813	12	19.7	2172	3	Q14182	Plasminogen-like grow 9.07e+02
c	814	12	19.7	2173	6	Q37308	Encodes haematopoieti 9.07e+02
c	815	12	19.7	2184	4	Q26052	Human HGF. 9.07e+02
c	816	12	19.7	2229	3	Q21883	BamHI/KpnI pSZ-4 clon 9.07e+02
c	817	12	19.7	2237	24	T38267	200 gene differential 9.07e+02
c	818	12	19.7	2246	4	Q25541	Consensus sequence fo 9.07e+02
c	819	12	19.7	2246	4	Q25537	Fragment of the S gen 9.07e+02
c	820	12	19.7	2246	9	Q52442	Feline infectious per 9.07e+02
	821	12	19.7	2256	3	Q20532	Sequence of clone lam 9.07e+02
	822	12	19.7	2256	7	Q42661	Lambda clone APCP1681 9.07e+02
	823	12	19.7	2287	17	T00785	Human afamin cDNA. 9.07e+02
c	824	12	19.7	2288	32	T72894	Plasmodium ebl-1 gene 9.07e+02
c	825	12	19.7	2334	39	V09695	M. hyopneumoniae 2.3k 9.07e+02
c	826	12	19.7	2355	7	Q43248	Sequence of AtT20 mur 9.07e+02
c	827	12	19.7	2355	12	Q71447	AtT 20 murine prohorm 9.07e+02
	828	12	19.7	2368	39	T94119	Human haematopoietic 9.07e+02
	829	12	19.7	2382	28	T58228	Thermotoga maritima m 9.07e+02
c	830	12	19.7	2383	39	V09638	Human M6 cDNA. 9.07e+02
c	831	12	19.7	2393	3	Q14038	Human hepatocyte grow 9.07e+02
	832	12	19.7	2400	1	Q06622	Asparagine synthetase 9.07e+02
	833	12	19.7	2424	1	N80154	Encodes Neisseria IgA 9.07e+02
c	834	12	19.7	2425	1	Q05679	btPGS1208 gene. 9.07e+02
c	835	12	19.7	2438	1	N90373	DNA encoding amphireg 9.07e+02
c	836	12	19.7	2458	2	Q12003	Human TR2-7 DNA bindi 9.07e+02
c	837	12	19.7	2458	30	T58943	Mouse dopamine D2 rec 9.07e+02
	838	12	19.7	2460	25	T42611	Malate permease mael 9.07e+02
c	839	12	19.7	2460	30	T67106	Murine lymphoid enhan 9.07e+02
	840	12	19.7	2469	16	Q92944	Schizosaccharomyces p 9.07e+02
	841	12	19.7	2469	12	Q70809	Protein kinase (Hhp1+ 9.07e+02
c	842	12	19.7	2478	4	Q27388	Protein disulphide is 9.07e+02
	843	12	19.7	2509	2	Q12900	Tobacco intracellular 9.07e+02
	844	12	19.7	2509	36	T89954	Nicotiana sp. intrace 9.07e+02
c	845	12	19.7	2514	13	Q81344	Mouse thymopoietin ge 9.07e+02
c	846	12	19.7	2522	34	T87082	Hyperthermostable xyl 9.07e+02
	847	12	19.7	2557	11	Q62625	Rat N-acetyl-glucosam 9.07e+02
	848	12	19.7	2595	2	N70259	Sequence of env gene 9.07e+02
c	849	12	19.7	2598	34	T72286	Breast cancer mammary 9.07e+02
c	850	12	19.7	2600	3	N82019	Sequence of amylogluc 9.07e+02
	851	12	19.7	2604	3	Q14945	Rhizopus PGK2 promote 9.07e+02
c	852	12	19.7	2615	1	N81002	Bacillus thuringiensi 9.07e+02
	853	12	19.7	2638	40	V20468	Human B-myb oncogene. 9.07e+02
	854	12	19.7	2652	9	Q51233	Sequence encoding pep 9.07e+02
	855	12	19.7	2656	5	Q30172	Bovine lung 59 kD CaM 9.07e+02
	856	12	19.7	2656	14	Q83964	Cyclic-GMP stimulated 9.07e+02
	857	12	19.7	2656	26	T51091	Coding sequence for c 9.07e+02
	858	12	19.7	2663	39	T94120	Alternatively spliced 9.07e+02

c	859	12	19.7	2745	3	Q14808	B.thuringiensis toxin 9.07e+02
	860	12	19.7	2763	36	T85305	Human methylthioadeno 9.07e+02
	861	12	19.7	2763	15	Q92813	Human MTAse. 9.07e+02
c	862	12	19.7	2781	40	V04155	Blood group antigen b 9.07e+02
c	863	12	19.7	2790	2	N60473	Sequence encoding the 9.07e+02
	864	12	19.7	2799	16	T05334	Peptide transport gen 9.07e+02
c	865	12	19.7	2821	36	T89395	Mutant EPM1 gene asso 9.07e+02
c	866	12	19.7	2821	36	T89394	Human mutant EPM1 ass 9.07e+02
	867	12	19.7	2835	20	T28772	Human calcium sensor 9.07e+02
	868	12	19.7	2835	39	V05993	Human calcium sensor 9.07e+02
c	869	12	19.7	2870	10	Q56190	pH6Z-3001 comprising 9.07e+02
	870	12	19.7	2880	18	T06304	Tomato pathogen resis 9.07e+02
	871	12	19.7	2919	30	T51124	Human homeoprotein re 9.07e+02
c	872	12	19.7	2946	34	T84563	Human SATB1 cDNA. 9.07e+02
	873	12	19.7	2949	1	Q05086	Sequence encodes NAP- 9.07e+02
	874	12	19.7	2973	9	Q53501	Sequence of bean mali 9.07e+02
	875	12	19.7	2974	13	Q79630	Partial FRAXE region 9.07e+02
c	876	12	19.7	2980	1	N82139	CryC gene. 9.07e+02
c	877	12	19.7	2980	1	N91071	Nucleotide sequence i 9.07e+02
c	878	12	19.7	2983	12	Q71026	CryIIIA gene which en 9.07e+02
	879	12	19.7	3006	1	N91049	Sequence encoding nov 9.07e+02
	880	12	19.7	3050	18	T06309	Tomato pathogen resis 9.07e+02
c	881	12	19.7	3077	39	V12857	ADE2 gene. 9.07e+02
c	882	12	19.7	3077	30	T60658	ADE2 gene. 9.07e+02
c	883	12	19.7	3077	40	V20592	Pichia methanolica AD 9.07e+02
c	884	12	19.7	3101	1	Q02047	Sequence encoding a c 9.07e+02
c	885	12	19.7	3106	38	V02323	Enterohaemorrhagic E. 9.07e+02
c	886	12	19.7	3106	38	T99905	Enterohaemorrhagic E. 9.07e+02
c	887	12	19.7	3158	9	Q53212	Human cyclin D3 promo 9.07e+02
c	888	12	19.7	3176	29	T62124	Arabidopsis thaliana 9.07e+02
	889	12	19.7	3287	34	T70386	Neisseria adhesion pr 9.07e+02
	890	12	19.7	3316	2	Q05632	Lambda gt10chl101 enco 9.07e+02
c	891	12	19.7	3340	40	V04154	Blood group antigen b 9.07e+02
c	892	12	19.7	3340	33	T84746	Alloreaction associat 9.07e+02
	893	12	19.7	3372	1	N80755	Encodes all HAV struc 9.07e+02
c	894	12	19.7	3437	17	T03437	Human matrix metallop 9.07e+02
	895	12	19.7	3463	33	T61728	Streptococcus pneumon 9.07e+02
c	896	12	19.7	3463	33	T61728	Streptococcus pneumon 9.07e+02
c	897	12	19.7	3511	22	T33246	PML-RAR-alpha RNA seq 9.07e+02
	898	12	19.7	3558	3	Q22596	Coding region of plas 9.07e+02
	899	12	19.7	3600	40	V04279	Staphylococcus epider 9.07e+02
	900	12	19.7	3603	25	T42217	Human TATA-binding pr 9.07e+02
	901	12	19.7	3692	36	V00160	Rat hexokinase III en 9.07e+02
c	902	12	19.7	3740	9	Q54682	Potato sucrose phosph 9.07e+02
	903	12	19.7	3807	3	Q14753	HIV-1 BA-L clone. 9.07e+02
	904	12	19.7	3846	34	T86170	Human Lasp-1 nucleoti 9.07e+02
c	905	12	19.7	3878	1	Q05672	Keratinocytre growth 9.07e+02
	906	12	19.7	3881	19	T13335	Blec4 gene and promot 9.07e+02
	907	12	19.7	3905	18	T06303	Tomato pathogen resis 9.07e+02
	908	12	19.7	3907	11	Q66820	AmEPV spheroidin gene 9.07e+02
c	909	12	19.7	4041	39	V16172	DNA sequence encoding 9.07e+02

c	910	12	19.7	4041	20	T13944	VIP2A(a) and VIP1A(a) 9.07e+02
	911	12	19.7	4156	22	T30166	Human endothelin-bomb 9.07e+02
c	912	12	19.7	4301	6	Q34584	ETb receptor gene. 9.07e+02
	913	12	19.7	4323	28	T36364	Lipoma preferred part 9.07e+02
c	914	12	19.7	4352	9	Q52440	Feline infectious per 9.07e+02
c	915	12	19.7	4365	4	Q25536	S gene of TS FIPV. 9.07e+02
c	916	12	19.7	4365	9	Q52446	Feline infectious per 9.07e+02
c	917	12	19.7	4415	7	Q46293	Phenylalanine ammonia 9.07e+02
c	918	12	19.7	4428	40	V15700	Human IDS gene. 9.07e+02
	919	12	19.7	4467	2	N92428	Sequence encoding amy 9.07e+02
	920	12	19.7	4483	1	N90843	Tobacco basic beta-1, 9.07e+02
	921	12	19.7	4487	3	Q20514	Encodes germ cell alk 9.07e+02
c	922	12	19.7	4673	4	Q27189	P. yoelii SSP2 antige 9.07e+02
	923	12	19.7	4698	1	N90844	Tobacco basic beta-1, 9.07e+02
c	924	12	19.7	4739	32	T78203	Physiologically activ 9.07e+02
	925	12	19.7	4818	32	T79909	Actin gene. 9.07e+02
c	926	12	19.7	4865	22	T36122	S. mutans antigen I/I 9.07e+02
c	927	12	19.7	4989	15	Q85962	Mouse ST2L gene. 9.07e+02
c	928	12	19.7	5012	24	T43175	Adzuki bean peroxidas 9.07e+02
	929	12	19.7	5149	23	T32455	Calpain large subunit 9.07e+02
	930	12	19.7	5406	13	Q81014	Flk1 receptor protein 9.07e+02
	931	12	19.7	5406	9	Q53504	Murine flk-1 cDNA. 9.07e+02
c	932	12	19.7	5460	20	T16508	Vector pAC3A1 contg. 9.07e+02
c	933	12	19.7	5613	11	Q63241	Crucifer 1-aminocyclo 9.07e+02
c	934	12	19.7	5643	38	T09945	Aspergillus oryzae ar 9.07e+02
c	935	12	19.7	5643	38	T97394	Aspergillus oryzae ar 9.07e+02
c	936	12	19.7	5648	24	T41406	Human REST protein cD 9.07e+02
c	937	12	19.7	5904	6	Q37813	Sequence encoding the 9.07e+02
c	938	12	19.7	6049	39	V16165	DNA encoding vegetati 9.07e+02
c	939	12	19.7	6049	20	T13939	B. cereus VIP2A(a) ge 9.07e+02
c	940	12	19.7	6049	38	T73993	B. cereus VIP2A(a) co 9.07e+02
c	941	12	19.7	6106	13	Q74678	Bacillus cereus vegat 9.07e+02
	942	12	19.7	6295	39	V02043	Plasmid pWRG3196 enco 9.07e+02
	943	12	19.7	6412	39	V06017	Human calcium protein 9.07e+02
	944	12	19.7	6412	20	T28796	Human calcium sensor 9.07e+02
	945	12	19.7	6560	9	Q54629	Genomic region contai 9.07e+02
c	946	12	19.7	6575	15	Q84655	Human neuronal calciu 9.07e+02
c	947	12	19.7	6611	29	T66129	Mycobacteriophage DS6 9.07e+02
c	948	12	19.7	6611	34	T70491	Mycobacteriophage DS6 9.07e+02
	949	12	19.7	6766	39	V14507	AmEPV entomopoxvirus 9.07e+02
	950	12	19.7	6768	39	V14517	AmEPV entomopoxvirus 9.07e+02
	951	12	19.7	7287	39	V02042	Plasmid pWRG3169 enco 9.07e+02
c	952	12	19.7	7400	38	T93024	Simian hepatitis A vi 9.07e+02
	953	12	19.7	7849	16	Q94109	hML genomic DNA. 9.07e+02
c	954	12	19.7	8355	24	T35524	Human interferon-beta 9.07e+02
	955	12	19.7	8450	36	T74994	Plasmid pFU1001 conta 9.07e+02
	956	12	19.7	8591	38	V04866	cDNA encoding amyloid 9.07e+02
	957	12	19.7	8591	34	T84562	Plasmid pCLL621 encod 9.07e+02
	958	12	19.7	8591	40	V05850	APP-REP 751 gene from 9.07e+02
	959	12	19.7	9060	40	V15702	Zucchini ACC synthase 9.07e+02
	960	12	19.7	9192	2	Q11943	Nucleotide sequence o 9.07e+02

c	961	12	19.7	9272	13	Q79353	Human genomic clone h 9.07e+02
	962	12	19.7	9737	5	Q31936	ARV-2 proviral DNA fr 9.07e+02
	963	12	19.7	9737	2	N60140	Sequence of ARV-2 (9B 9.07e+02
	964	12	19.7	10079	14	Q88057	Human EC-SOD gene. 9.07e+02
	965	12	19.7	10079	35	T92317	Human EC-SOD gene. 9.07e+02
	966	12	19.7	10140	22	T32232	Plasmid pBE92. 9.07e+02
c	967	12	19.7	10718	9	Q51476	DEN1-S275/90 (ECACC V 9.07e+02
c	968	12	19.7	10798	4	Q28505	Tomato vacuolar inver 9.07e+02
c	969	12	19.7	10965	4	Q28507	Tomato vacuolar inver 9.07e+02
c	970	12	19.7	11715	36	T91087	Bovine lysosomal alph 9.07e+02
c	971	12	19.7	11800	11	Q67676	Acetyl CoA carboxylas 9.07e+02
	972	12	19.7	11817	35	T74199	Murine Lystl long iso 9.07e+02
c	973	12	19.7	11907	10	Q45339	cDNA encoding htrx. 9.07e+02
c	974	12	19.7	12284	1	Q06001	Sequence encoding pro 9.07e+02
c	975	12	19.7	12537	24	T41705	Lymphocyte specific i 9.07e+02
c	976	12	19.7	12616	37	T89256	Human bg gene (long f 9.07e+02
c	977	12	19.7	13417	24	T09224	Partial sequence of v 9.07e+02
	978	12	19.7	14042	20	T28774	Human placental calci 9.07e+02
	979	12	19.7	14042	39	V05995	Human placenta calciu 9.07e+02
c	980	12	19.7	14042	39	V05995	Human placenta calciu 9.07e+02
	981	12	19.7	14044	20	T28776	Human parathyroid cal 9.07e+02
	982	12	19.7	14044	39	V05997	Human parathyroid cal 9.07e+02
c	983	12	19.7	14044	39	V05997	Human parathyroid cal 9.07e+02
c	984	12	19.7	14044	20	T28776	Human parathyroid cal 9.07e+02
	985	12	19.7	14080	39	V05996	Human kidney calcium 9.07e+02
c	986	12	19.7	14080	39	V05996	Human kidney calcium 9.07e+02
c	987	12	19.7	14080	20	T28775	Human kidney calcium 9.07e+02
	988	12	19.7	14086	39	V05994	Human calcium protein 9.07e+02
c	989	12	19.7	14086	39	V05994	Human calcium protein 9.07e+02
c	990	12	19.7	14086	20	T28773	Human calcium sensor 9.07e+02
c	991	12	19.7	14255	40	V20477	Human ALL-1/AF-4 onco 9.07e+02
c	992	12	19.7	14255	7	Q43526	ALL-1 gene cDNA. 9.07e+02
	993	12	19.7	15155	3	N50107	DNA encoding Factor-V 9.07e+02
	994	12	19.7	28994	39	V15826	Genomic DNA for inter 9.07e+02
c	995	12	19.7	28994	39	V15826	Genomic DNA for inter 9.07e+02
	996	12	19.7	29392	40	V15422	Mouse poly Ig recepto 9.07e+02
	997	12	19.7	35408	31	T59272	Recombinant adenoviru 9.07e+02
	998	12	19.7	53526	40	T94101	Human PKD1 gene. 9.07e+02
	999	12	19.7	53577	40	T94108	Human PKD1 locus betw 9.07e+02
	1000	12	19.7	53577	28	T18551	Human polycystic kidn 9.07e+02

ALIGNMENTS

```
RESULT    1
ID   T20449 standard; cDNA to mRNA; 62 BP.
AC   T20449;
DT   19-JUL-1996  (first entry)
DE   Human gene signature HUMGS01603.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
```



```
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 639; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   62 BP;    30 A;    11 C;    9 G;    11 T;

  Query Match          100.0%;  Score 61;  DB 19;  Length 62;
  Best Local Similarity 100.0%;  Pred. No. 2.64e-29;
  Matches   62;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 gatccccattcttttcaaaaacaaaaataaaacaataaagactgcaaggaagactgangga 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      1 GATCCCCATTCTTTTCAAAAACAAAATAAAACAATAAAGACTGCAAGGAAGACTGANGGA 60

Db     61 aa 62
          ||
Qy     61 AA 62


RESULT    2
ID   T61096 standard; DNA; 1116 BP.
AC   T61096;
DT   17-OCT-1997  (first entry)
DE   N-acetylglucosaminyl transferase III promoter.
KW   Human N-acetylglucosylaminyl transferase III promoter;
```



```
KW   gene therapy; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   misc_feature    300..1116
FT                   /*tag=  a
FT                   /note= "Claimed subfragment"
FT   misc_feature    595..116
FT                   /*tag=  b
FT                   /note= "Claimed subfragment"
FT   misc_feature    917..1116
FT                   /*tag=  c
FT                   /note= "Claimed subfragment"
PN   EP-753579-A2.
PD   15-JAN-1997.
PF   28-JUN-1996; 110512.
PR   17-AUG-1995; JP-233364.
PR   30-JUN-1995; JP-187753.
PA   (TAKI ) TAKARA SHUZO CO LTD.
PI   Ihara Y,  Koyama N,  Miyoshi E,  Nishikawa A,  Taniguchi N;
DR   WPI; 97-079381/08.
PT   Human N-acetyl:glucosyl:aminyl transferase III promoter - for use in
PT   vectors for expression of useful genes in animal cells
PS   Claim 1; Page 8-9; 14pp; English.
CC   This sequence represents the human N-acetylglucosylaminyl transferase
CC   III promoter.  This sequence has a higher activity than conventional
CC   promoters in animal cells.  It may be used for expression of useful,
CC   heterologous genes in animal cells.
SQ   Sequence   1116 BP;    235 A;    281 C;    316 G;    284 T;

  Query Match           27.9%;  Score 17;  DB 32;  Length 1116;
  Best Local Similarity 100.0%;  Pred. No. 2.52e+00;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    215 ttttatttttgtttttga 231
          ||||||||||||||||||
Cp     31 TTTTATTTTGTTTTTGA 15


RESULT    3
ID   T19003 standard; cDNA to mRNA; 625 BP.
AC   T19003;
DT   14-JUN-1996  (first entry)
DE   Human gene signature HUMGS00003.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
```



```
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 267; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   625 BP;    142 A;    120 C;    154 G;    199 T;

  Query Match           26.2%;  Score 16;  DB 18;  Length 625;
  Best Local Similarity 100.0%;  Pred. No. 8.79e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    155 tgttttattttgtttt 170
          ||||||||||||||||
Cp     33 TGTTTTATTTTGTTTT 18


RESULT    4
ID   T66381 standard; cDNA; 1855 BP.
AC   T66381;
DT   20-JUL-1997  (first entry)
DE   Human adrenal gland cysteine protease Incyte Clone 100877.
KW   Cysteine protease; autoimmune disease; degenerative disease;
KW   anaemia; arteriosclerosis; asthma; bronchitis; emphysema;
KW   gingivitis; inflammatory bowel disease; diabetes mellitus;
KW   leukaemia; multiple endocrine neoplasia; osteoarthritis;
KW   osteoporosis; pulmonary fibrosis; rheumatoid arthritis;
KW   septic shock; systemic lupus erythematosus; diagnosis; therapy; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             54..1349
FT                   /*tag=  a
PN   WO9715592-A1.
```

```
PD   01-MAY-1997.
PF   23-OCT-1996; U16926.
PR   23-OCT-1995; US-005799.
PA   (INCY-) INCYTE PHARM INC.
PI   Braxton SM,  Delegeane AM,  Diep D;
DR   WPI; 97-259954/23.
DR   P-PSDB; W15460.
PT   Human adrenal gland cysteine protease - used for diagnosing
PT   autoimmune and degenerative diseases
PS   Claim 3; Page 37; 62pp; English.
CC   A cDNA clone (T66381), designated Incyte Clone 100877, codes for a
CC   novel human cysteine protease (CP) (W15460) that is expressed in
CC   cells of the adrenal gland, and which is thought to be involved in
CC   systemic cleanup and defence.  The clone was identified via a
CC   computer search for local sequence alignments among the cDNAs of an
CC   adrenal gland library.  A modified XL-PCR procedure (see also
CC   T66404-05) was used to obtain the full-length 1855 bp cDNA sequence.
CC   A partial nucleotide sequence, Incyte Clone 66931 (T66383), was
CC   obtd. from a human umbilical vein endothelial cell cDNA library,
CC   and partial sequences (T66382, T66384-403) were also identified from
CC   several other libraries.  CP nucleic acids can be used as probes for
CC   diagnosis of conditions associated with altered CP expression,
CC   partic. autoimmune and degenerative diseases, to produce
CC   recombinant CP polypeptides, and to design antisense molecules
CC   useful in reducing expression of the genomic sequence.
SQ   Sequence   1855 BP;    498 A;    447 C;    452 G;    458 T;

  Query Match           26.2%;  Score 16;  DB 30;  Length 1855;
  Best Local Similarity 100.0%;  Pred. No. 8.79e+00;
  Matches    16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    571 ttttcaaaaacaaaat 586
          ||||||||||||||||
Qy     12 TTTTCAAAAACAAAAT 27


RESULT    5
ID   T48454 standard; DNA; 6060 BP.
AC   T48454;
DT   12-APR-1997  (first entry)
DE   Human beta-globin gene.
KW   Haemoglobinopathy; sickle cell anaemia; beta-thalassaemia;
KW   haemoglobin; beta-globin; HbS; HbA; gene therapy; triple helix;
KW   triplex; targetted DNA replacement; homologous recombination;
KW   mutagenesis; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             2187..2189
FT                   /*tag=  a
FT                   /codon_start= 2187..2189
FT   mutation        2206
```



```
FT                   /*tag=  b
FT                   /note= "mutation of adenine at base 2206 (base 19
FT                   of the coding sequence) to thymine causes a
FT                   glutamic acid to valine substn. leading to
FT                   sickle cell anaemia"
FT   misc_binding    781..801
FT                   /*tag=  c
FT                   /note= "preferred site for targetted DNA
FT                   replacement"
FT   misc_binding    1214..1246
FT                   /*tag=  d
FT                   /note= "preferred site for targetted DNA
FT                   replacement"
FT   misc_binding    1608..1623
FT                   /*tag=  e
FT                   /note= "preferred site for targetted DNA
FT                   replacement"
FT   misc_binding    2655..2675
FT                   /*tag=  f
FT                   /note= "preferred site for targetted DNA
FT                   replacement"
FT   misc_binding    2825..2851
FT                   /*tag=  g
FT                   /note= "preferred site for targetted DNA
FT                   replacement"
FT   misc_binding    3186..3208
FT                   /*tag=  h
FT                   /note= "preferred site for targetted DNA
FT                   replacement"
FT   misc_binding    4895..4915
FT                   /*tag=  i
FT                   /note= "preferred site for targetted DNA
FT                   replacement"
FT   misc_binding    5149..5192
FT                   /*tag=  j
FT                   /note= "preferred site for targetted DNA
FT                   replacement"
PN   WO9640271-A1.
PD   19-DEC-1996.
PF   06-JUN-1996; U09430.
PR   07-JUN-1995; US-473845.
PA   (UYYA ) UNIV YALE.
PI   Glazer PM;
DR   WPI; 97-099895/09.
PT   Repairing mutation(s) in haemoglobin by targetted mutagenesis or
PT   homologous recombination - mediated by a triplex forming
PT   oligonucleotide, opt. carrying a mutagen, partic. for treatment of
PT   sickle cell anaemia or thalassemia
PS   Claim 32; Page 41-44; 70pp; English.
CC   The human beta-globin gene (T48454) contains third-strand binding
CC   sites that can be used in targetted DNA replacement (TDR).  This
```



```
CC   method for genetic defect repair involves targetting third strands
CC   (see also T48467-70) to native DNA sequences (see also T48455-57)
CC   to induce DNA damage that stimulates homologous recombination with
CC   an introduced donor nucleic acid strand.  The method is used to
CC   correct mutations in a target region, esp. for treatment of beta-
CC   thalassaemia and sickle cell anaemia.  TDR is highly specific and,
CC   since the corrected gene is in its native chromosome, the cure is
CC   permanent.  A mutation at position 19 (thymine) of the coding
CC   sequence may also be corrected by targetted mutagenesis, in which
CC   highly specific damage to the thymine by psoralen-bound third
CC   strands (see also T48460-66) is preferentially 'misrepaired' to
CC   adenine, thereby changing HbS to normal HbA.
SQ   Sequence   6060 BP;    1895 A;    1096 C;    1169 G;    1900 T;

  Query Match           26.2%;  Score 16;  DB 28;  Length 6060;
  Best Local Similarity 100.0%;  Pred. No. 8.79e+00;
  Matches    16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    715 tttattgttttatttt 730
          ||||||||||||||||
Cp     38 TTTATTGTTTTATTTT 23


RESULT    6
ID   T73568 standard; DNA; 14507 BP.
AC   T73568;
DT   28-SEP-1997  (first entry)
DE   Expression augmenting sequence element (EASE).
KW   Expression augmenting sequence element; EASE; vector;
KW   recombinant protein; gene expression; Chinese hamster ovary; CHO;
KW   ss.
OS   Cricetulus sp.
FH   Key             Location/Qualifiers
FT   misc_feature    1..14507
FT                   /*tag=  a
FT                   /note= "EASE (Claim 1)"
FT   misc_feature    5980..14507
FT                   /*tag=  b
FT                   /note= "EASE (Claim 1)"
FT   misc_feature    8671..14507
FT                   /*tag=  c
FT                   /note= "EASE (Claim 1)"
FT   misc_feature    8671..10515
FT                   /*tag=  d
FT                   /note= "EASE (Claim 1)"
FT   misc_feature    9277..10515
FT                   /*tag=  e
FT                   /note= "EASE (Claim 1)"
FT   misc_feature    8672..12273
FT                   /*tag=  f
FT                   /note= "EASE (Claim 1)"
```



```
FT   misc_feature    10100..14923
FT                   /*tag=  g
FT                   /note= "EASE (Claim 1)"
FT   misc_feature    14290..14507
FT                   /*tag=  h
FT                   /note= "the EASE sequence is preferably ligated
FT                   to DNA comprising nucleotides 14290-14507
FT                   (Claim 3)"
PN   WO9725420-A1.
PD   17-JUL-1997.
PF   10-JAN-1997; U00483.
PR   11-JAN-1996; US-586509.
PA   (IMMV ) IMMUNEX CORP.
PI   Lee C,  Morris AE,  Thomas JN;
DR   WPI; 97-372861/34.
PT   Expression augmenting sequence elements - used in vectors for
PT   expressing recombinant proteins at high levels in shorter periods of
PT   time
PS   Claim 1; Page 23-30; 36pp; English.
CC   A 14507 bp DNA sequence (T73568) comprises a novel transcription
CC   regulatory element, expression augmenting sequence element (EASE),
CC   that facilitates high expression of recombinant proteins in
CC   mammalian host cells.  It was identified by cloning the integration
CC   site of a unique expression cassette encoding recombinant dimeric
CC   tumour necrosis factor receptor-immunoglobulin Fc fusion protein
CC   from genomic DNA of a CHO 2A5-3 cell line expressing this protein
CC   at a high level.  Expression vectors incorporating an EASE show
CC   a 2- to 8-fold improvement of recombinant protein expression levels
CC   in CHO host cells.  The EASE sequence also facilitates high
CC   expression in shorter periods of time.
SQ   Sequence   14507 BP;    4225 A;    2548 C;    3136 G;    4598 T;

  Query Match           26.2%;  Score 16;  DB 32;  Length 14507;
  Best Local Similarity 100.0%;  Pred. No. 8.79e+00;
  Matches    16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   8454 caaaaacaaaataaaa 8469
          ||||||||||||||||
Qy     16 CAAAAACAAAATAAAA 31


RESULT    7
ID   T39432 standard; RNA; 26 BP.
AC   T39432;
DT   13-NOV-1996  (first entry)
DE   Hel-N2 selected sequence, e-6.
KW   Human; neuron-specific protein; Hel-N1; 3'-UTR instability sequence;
KW   paraneoplastic sensory neuropathy; oncoprotein; lymphokine; rat; elav;
KW   RNA recognition motif; RRM; Drosophila; cellular growth; localisation;
KW   instability; translatability; neurons; autoimmune protein; PE; PCD; PSN;
KW   central nervous system; cancer; paraneoplastic cerebellar degeneration;
```

```
KW   paraneoplastic encephalomyelitis; RNP-1 octamer sequence;
KW   human elav-like neuronal protein-1; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_rna        13..21
FT                   /*tag= a
FT                   /note= "Consensus sequence"
PN   US5525495-A.
PD   11-JUN-1996.
PF   11-MAY-1992; 881075.
PR   11-MAY-1992; US-881075.
PR   15-SEP-1993; US-120827.
PA   (UYDU-) UNIV DUKE.
PI   Gao F,  Keene JD,  Levine T;
DR   WPI; 96-286398/29.
PT   Prodn of cDNA library for related proteins - by screening total cell
PT   mRNA with RNA-binding protein Hel-N1 or Hel-N2
PS   Disclosure; Fig 11; 66pp; English.
CC   The sequences given in T39427-46 represent random oligonucleotides
CC   which were isolated due to their ability to bind to the human neuron-
CC   specific protein, Hel-N2.  These sequences contain short stretches
CC   of uridylate residues interspersed with other residues.  These U-
CC   rich regions share homology with the 3'-UTR instability sequences
CC   that are found in mRNA's.  Instability sequences are target elements
CC   which reside in the 3'-non-coding regions of mRNA's which encode
CC   oncoproteins and lymphokines.  Hel-N2 is a deleted form of Hel-N1
CC   in which the residues 239-251 have been deleted.  This protein is
CC   expressed in medulloblastoma tumour cells and is not found in whole
CC   human brain.  A small amount of Hel-N2 is also found in fetal brain,
CC   which may indicate a correlation with rapid growth.  Hel-N1 cDNA was
CC   isolated by probing for rat and human elav counterparts using degenerate
CC   primers designed to simulate the RNP-1 octamer sequence present in two
CC   of the three RNA recognition motifs (RRM's) of Drosophila elav.  Hel-N1
CC   was found to contain 3 RRM's, where the third one (see also W00244) is
CC   sufficient for mRNA 3'-UTR binding activity.  Full length Hel-N1, when
CC   transfected into a cell, causes cellular growth to cease, however, if
CC   just the third binding domain is transfected into cells, the cells
CC   undergo rapid growth.  Hel-N1 binds as a multimer along the mRNA,
CC   presumably enhancing its localisation, instability and/or regulating it
CC   translatability and/or deadenylating it.  This protein may be
CC   responsible for the growth cessation of neurons. Hel-N1 is an autoimmune
CC   protein in certain patients who show central nervous system
CC   manifestations of cancer called paraneoplastic cerebellar degeneration
CC   (PCD), paraneoplastic encephalomyelitis (PE) or paraneoplastic sensory
CC   neuropathy (PSN).
SQ   Sequence   26 BP;    5 A;    1 C;    4 G;    16 U;

  Query Match           24.6%;  Score 15;  DB 23;  Length 26;
  Best Local Similarity  20.0%;  Pred. No. 2.97e+01;
  Matches    3;  Conservative   12;  Mismatches   0;  Indels   0;  Gaps   0;
```



```
Db      6 uuuauuguuuuauuu 20
          :::|::|::::|:::
Cp     38 TTTATTGTTTTATTT 24


RESULT    8
ID   T20056 standard; cDNA to mRNA; 103 BP.
AC   T20056;
DT   17-JUL-1996  (first entry)
DE   Human gene signature HUMGS01197.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 548; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   103 BP;    19 A;    16 C;    21 G;    41 T;

  Query Match           24.6%;  Score 15;  DB 19;  Length 103;
  Best Local Similarity  93.8%;  Pred. No. 2.97e+01;
  Matches   15;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db     14 tgtttttntttttgtttt 29
          |||||| |||||||||
Cp     33 TGTTTTATTTTGTTTT 18
```



```
RESULT    9
ID   T23541 standard; cDNA to mRNA; 111 BP.
AC   T23541;
DT   02-SEP-1996  (first entry)
DE   Human gene signature HUMGS05390.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 1394; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   111 BP;    38 A;    23 C;    22 G;    25 T;

  Query Match           24.6%;  Score 15;  DB 20;  Length 111;
  Best Local Similarity 100.0%;  Pred. No. 2.97e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     81 aaataaaacaataaa 95
          ||||||||||||||||
Qy     24 AAATAAAACAATAAA 38


RESULT   10
```



```
ID   T19543 standard; cDNA to mRNA; 330 BP.
AC   T19543;
DT   28-JUN-1996  (first entry)
DE   Human gene signature HUMGS00607.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 413; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   330 BP;    61 A;    80 C;    84 G;    89 T;

  Query Match           24.6%;  Score 15;  DB 18;  Length 330;
  Best Local Similarity 100.0%;  Pred. No. 2.97e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    230 ttgttttattttgtt 244
          ||||||||||||||||
Cp     34 TTGTTTTATTTTGTT 20


RESULT   11
ID   Q59994 standard; cDNA; 369 BP.
AC   Q59994;
DT   16-MAR-1994  (first entry)
```

```
DE   Human brain Expressed Sequence Tag EST01975.
KW   Gene transcription product; genetic markers; tagging; in vivo;
KW   transcription; mapping; locations; chromosomes; chromosomal; ss.
OS   Homo sapiens.
PN   WO9316178-A.
PD   19-AUG-1993.
PF   12-FEB-1993; U01294.
PR   12-FEB-1992; US-837195.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
PI   Adams MD, Moreno RF, Venter CJ;
DR   WPI; 93-272882/34.
PT   Enriched oligonucleotides and corresp. sequences - used as
PT   markers for human genes transcribed in-vivo, facilitate tagging
PT   of most human genes
PS   Example 4; Page 264; 500pp; English.
CC   The Expressed Sequence Tag was isolated from a human brain cDNA
CC   library as part of a large set of ESTs which can be used as markers
CC   for human genes transcribed in vivo. They can be used to facilitate
CC   tagging of most human genes, for mapping locations of expressed genes
CC   on chromosomes, for individual or forensic identification, for mapping
CC   locations of disease-associated genes, for identification of tissue
CC   type, and for prepn. of antisense sequences, probes and constructs.
CC   EST01975 has a "poor" coding probability as evaluated using the
CC   coding-region prediction program CRM. See also Q59041-Q61440.
SQ   Sequence   369 BP;    114 A;    72 C;    73 G;    109 T;

  Query Match           24.6%;  Score 15;  DB 8;  Length 369;
  Best Local Similarity 100.0%;  Pred. No. 2.97e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     26 ttttcaaaaacaaaa 40
          ||||||||||||||||
Qy     12 TTTTCAAAAACAAAA 26


RESULT   12
ID   Q02418 standard; DNA; 503 BP.
AC   Q02418;
DT   31-AUG-1990  (first entry)
DE   Promoter of pCP-A1 expression vector.
KW   Promoter; expression vector; ds.
OS   Candida maltosa.
PN   JO2072882-A.
PD   13-MAR-1990.
PF   8-SEP-1988; 225947.
PR   8-SEP-1988; JP-225947.
PA   (SUMO) Sumitomo Chem Ind KK.
DR   WPI; 90-121046/16.
PT   Vector for expression of foreign genes in Candida maltosa -
PT   contg. origin of replication functioning in C.maltosa, promoter
PT   region, restriction site, and terminator from Saccharomyces
```



```
PT   cerevisiae LEU2 gene.
PS   Claim 2; Fig 5; 9pp; Japanese.
CC   Promoter region combined with origin of replication from C.maltosa,
CC   and a termination site from S.cerevisiae form this expression
CC   vector.
SQ   Sequence   503 BP;   140 A;   80 C;   67 G;   216 T;

  Query Match           24.6%;  Score 15;  DB 1;  Length 503;
  Best Local Similarity 100.0%;  Pred. No. 2.97e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    478 ttttatttttgttttt 492
          ||||||||||||||||
Cp     31 TTTTATTTTGTTTTT 17


RESULT   13
ID   T91347 standard; DNA; 633 BP.
AC   T91347;
DT   20-APR-1998  (first entry)
DE   Human cystatin F polypeptide encoding DNA.
KW   Cystatin f; human; treatment; diagnosis; neoplasia; screening;
KW   HIV; tumour; immunological disorder; infection; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             67..504
FT                   /*tag= a
FT                   /note= "mature human cystatin F"
PN   WO9736915-A1.
PD   09-OCT-1997.
PF   03-APR-1996; U04572.
PR   03-APR-1996; WO-U04572.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Gentz RL, Li H, Ni J, Yu G;
DR   WPI; 97-503039/46.
DR   P-PSDB; W32323.
PT   Human cystatin F polypeptide - useful for e.g. treating patients
PT   requiring cystatin F, for diagnosis of neoplasia and screening of
PT   inhibitory compounds
PS   Claim 8; Pages 60-61; 77pp; English.
CC   This DNA encodes a mature human cystatin F polypeptide. The cystatin F
CC   polypeptide can be administered therapeutically, especially by expressing
CC   this encoding DNA in vivo, to treat patients requiring cystatin F. They
CC   are useful in the treatment or prevention of bacterial infection, viral
CC   infection and eye inflammation, and in eye remodelling and treatment of
CC   immunological disorders. They can be used to identify compounds which
CC   inhibit or activate cystatin F activity or bind cystatin F. Inhibitory
CC   compounds can be administered to patients requiring cystatin F inhibition
CC   e.g. in treatment of cerebral haemorrhages, encephalopathy and tumours
CC   and to inhibit HIV infection. The polynucleotides can be used to diagnose
CC   diseases involving cystatin F under-expression, by detecting
```



```
CC   polynucleotide mutations and as hybridisation probes to isolate genes
CC   with high sequence similarity. The polypeptides can be used in assays to
CC   diagnose polypeptide levels in samples and to generate antibodies. They
CC   are useful in detection of hereditary cystatin C amyloidosis angiopathy
CC   and neoplasia.
SQ   Sequence   633 BP;    158 A;    206 C;    148 G;    121 T;

  Query Match           24.6%;  Score 15;  DB 38;  Length 633;
  Best Local Similarity 100.0%;  Pred. No. 2.97e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    183 taaaacaataaagac 197
          ||||||||||||||||
Qy     27 TAAAACAATAAAGAC 41


RESULT   14
ID   T89171 standard; DNA; 633 BP.
AC   T89171;
DT   14-APR-1998  (first entry)
DE   Human cystatin F encoding sequence.
KW   Human cystatin F; malignant cell; inhibitor; rheumatoid arthritis;
KW   purulent bronchiestasis; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             67..504
FT                   /*tag= a
FT                   /product= cystatin_F
PN   WO9737021-A1.
PD   09-OCT-1997.
PF   03-APR-1997; U05595.
PR   03-APR-1996; WO-U04572.
PR   03-APR-1996; US-014795.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Gentz RL, Ni J, Yu G;
DR   WPI; 97-503106/46.
DR   P-PSDB; W31902.
PT   Human cystatin F and related DNA - used for the diagnosis and
PT   treatment of disease in which cystatin F activation is beneficial,
PT   e.g. cancer
PS   Claim 1; Pages 62-63; 79pp; English.
CC   This sequence encodes the protein cystatin F. The sequence contains
CC   an open reading frame of 145 codons which encodes the protein.
CC   Cystatin F polypeptides have been found to display a variety of
CC   functions including influencing the intra- and extracellular catabolism
CC   of proteins and peptides, regulating proteolytic processing of
CC   pro-hormones and pro-enzymes, protecting against penetration of normal
CC   tissues by malignant cells or microorganisms and modulating local
CC   inflammatory processes in rheumatoid arthritis and purulent
CC   brochiestasis. Cystatin F and its agonists can be used for the
CC   treatment and diagnosis of diseases related to under-expression of
```



```
CC   cystatin F such as specific cancer states. Inhibitors of cystatin F have
CC   the opposite effect and can be used to treat conditions where cystatin F
CC   inhibition is beneficial.
SQ   Sequence   633 BP;    158 A;    205 C;    148 G;    122 T;

  Query Match           24.6%;  Score 15;  DB 37;  Length 633;
  Best Local Similarity 100.0%;  Pred. No. 2.97e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    183 taaaacaataaagac 197
          ||||||||||||||||
Qy     27 TAAAACAATAAAGAC 41


RESULT   15
ID   Q53499 standard; DNA; 674 BP.
AC   Q53499;
DT   24-JUN-1994  (first entry)
DE   Sequence encoding protein required for surfactin synthesis.
KW   Surfactin; Bacillus subtilis; antibiotic; fungicide; anticoagulant;
KW   thrombosis; myocardial infarction; embolism; ss.
OS   Bacillus subtilis.
PN   EP-576050-A.
PD   29-DEC-1993.
PF   23-APR-1993; 201177.
PR   24-APR-1992; IT-MI0976.
PA   (ENIE ) ENIRICERCHE SPA.
PI   Cosmina P,  De Ferra F,  Grandi G,  Perego M,  Rodriguez F;
DR   WPI; 94-001231/01.
DR   P-PSDB; R44533.
PT   DNA from Bacillus subtilis - encodes surfactin synthesis protein
PS   Claim 1; Figure 2; 24pp; English.
CC   The DNA (sfp gene) encodes a protein required for surfactin
CC   synthesis.  It is useful for the in vitro and in vivo production of
CC   surfactin and analogues having anti-cholesterase, fungicidal and
CC   antibiotic properties. It can also be used as an anticoagulant for
CC   thrombosis and for prophylaxis of myocardial infarcts and pulmonary
CC   emboli.  This gene was isolated from a producer strain of
CC   microorganism.
SQ   Sequence   674 BP;    194 A;    167 C;    149 G;    164 T;

  Query Match           24.6%;  Score 15;  DB 9;  Length 674;
  Best Local Similarity 100.0%;  Pred. No. 2.97e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    364 ttcttttcaaaaaca 378
          ||||||||||||||||
Qy      9 TTCTTTTCAAAAACA 23


RESULT   16
```

```
ID   Q53500 standard; DNA; 675 BP.
AC   Q53500;
DT   24-JUN-1994  (first entry)
DE   Sequence encoding protein required for surfactin synthesis.
KW   Surfactin; Bacillus subtilis; antibiotic; fungicide; anticoagulant;
KW   thrombosis; myocardial infarction; embolism; sfp; non-producer; ss.
OS   Bacillus subtilis.
PN   EP-576050-A.
PD   29-DEC-1993.
PF   23-APR-1993; 201177.
PR   24-APR-1992; IT-MI0976.
PA   (ENIE ) ENIRICERCHE SPA.
PI   Cosmina P,  De Ferra F,  Grandi G,  Perego M,  Rodriguez F;
DR   WPI; 94-001231/01.
DR   P-PSDB; R44534.
PT   DNA from Bacillus subtilis - encodes surfactin synthesis protein
PS   Disclosure; Figure 3; 24pp; English.
CC   The DNA encodes a protein required for surfactin synthesis (sfp
CC   gene).  It is useful for the in vitro and in vivo production of
CC   surfactin and analogues having anti-cholesterase, fungicidal and
CC   antibiotic properties. It can also be used as an anticoagulant for
CC   thrombosis and for prophylaxis of myocardial infarcts and pulmonary
CC   emboli.  This sfp gene was isolated from a non-producer strain of
CC   microorganism (The sfp gene has an inserted adenine base causing a
CC   frameshift and resulting in the production of a truncated protein).
SQ   Sequence   675 BP;    195 A;    167 C;    148 G;    165 T;

  Query Match           24.6%;  Score 15;  DB 9;  Length 675;
  Best Local Similarity 100.0%;  Pred. No. 2.97e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    364 ttcttttcaaaaaca 378
          ||||||||||||||||
Qy      9 TTCTTTTCAAAAACA 23


RESULT   17
ID   Q48942 standard; DNA; 690 BP.
AC   Q48942;
DT   13-APR-1994  (first entry)
DE   Human HT6.
KW   T-cell; differentiation; marker; expression; allele; variant;
KW   HT6; diagnosis; predisposition; diabetes; autoimmune disease;
KW   T-lymphocytes; insulin-dependent diabetes;
KW   PCR; polymerase chain reaction; primer; amplification; ss.
OS   Homo sapiens.
PN   DE4209216-A.
PD   23-SEP-1993.
PF   21-MAR-1992; 209216.
PR   21-MAR-1992; DE-209216.
PA   (BOEF ) BOEHRINGER MANNHEIM GMBH.
```



```
PI   Haag F,  Nolte F,  Thiele H;
DR   WPI; 93-304369/39.
DR   P-PSDB; R41876.
PT   New HT6 gene DNA encoding T-cell differentiation marker - is used
PT   to produce allelic variants, derived protein etc. for diagnosing
PT   e.g. predisposition to diabetes and other auto-immune diseases
PS   Claim 1; Page 7; 12pp; German.
CC   HT6 DNA (Q48942) was obtained by PCR using the primers given in
CC   Q48945-Q48950.  Fragments (bases 1-271 and 272-690 or the entire
CC   sequence) are used as probes (claim 6-7).
CC   HT6 DNA, or its fragments, are used to detect allelic variants
CC   (which can then be cloned, identified and isolated).  HT6 may be
CC   produced by recombinant DNA methods.  HT6 DNA is used in analysis of
CC   HT6 expression defects in cases of autoimmune diseases and to identify
CC   a particular subset of T-lymphocytes.  HT6 is a T-cell differentiation
CC   marker and its expression can serve to diagnose predisposition to
CC   insulin-dependent diabetes or other autoimmune diseases.
SQ   Sequence   690 BP;    242 A;    123 C;    135 G;    190 T;

  Query Match           24.6%;  Score 15;  DB 8;  Length 690;
  Best Local Similarity 100.0%;  Pred. No. 2.97e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    167 tcaaaaacaaaataa 181
          ||||||||||||||||
Qy     15 TCAAAAACAAAATAA 29


RESULT   18
ID   Q13299 standard; cDNA; 696 BP.
AC   Q13299;
DT   30-OCT-1991  (first entry)
DE   Growth hormone-like glycoprotein gene.
KW   Pituitary; ss.
OS   Pleuronectes flesus.
FH   Key             Location/Qualifiers
FT   signal_peptide  1..24
FT                   /*tag=  a
FT   mat_peptide     25..207
FT                   /*tag=  b
PN   EP-443790-A.
PD   28-AUG-1991.
PF   18-FEB-1991; 301252.
PR   19-FEB-1990; JP-036090.
PA   (NIOC ) NIPPON OIL KK.
PI   Sakata S, Kawauchi H, Ono M;.
DR   WPI; 91-254370/35.
DR   P-PSDB; R13583.
PT   Growth hormone-like glyco-protein - isolated from pituitary gland
PT   of fish of order Pleuronectina, family Paralichthydae, partic.
PT   flounder.
```



```
PS   Claim 9; Fig 4; 29pp; English.
CC   The gene encodes a growth hormone-like protein which is found in
CC   the pituitary gland of flounders. The cDNA sequence was determined
CC   by forming deletion mutants with restriction enzymes and if no
CC   restriction enzyme site is found, by synthesising primers. Using
CC   transformante large amts. of the growth hormone like protein can
CC   be obtd. The protein can be used for growth promotion in animals,
CC   esp. fish. See also Q13300.
SQ   Sequence   696 BP;    180 A;    190 C;    172 G;    154 T;

  Query Match           24.6%;  Score 15;  DB 2;  Length 696;
  Best Local Similarity 100.0%;  Pred. No. 2.97e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     81 agactgcaaggaaga 95
          ||||||||||||||||
Qy     38 AGACTGCAAGGAAGA 52


RESULT   19
ID   T75004 standard; DNA; 1163 BP.
AC   T75004;
DT   06-OCT-1997  (first entry)
DE   Human endogenous retroviral sequence 11-29.
KW   Breast cancer; tumour; B18Ag1; prognosis; diagnosis; vaccine; ss.
OS   Human retrovirus.
PN   WO9725431-A1.
PD   17-JUL-1997.
PF   10-JAN-1997; U00398.
PR   10-JAN-1996; US-587329.
PA   (CORI-) CORIXA CORP.
PI   Frudakis TN,  Smith JM;
DR   WPI; 97-384982/35.
PT   Endogenous human tumour-associated retroviral element, B18Ag1 - used
PT   for the prognosis, diagnosis and monitoring of human cancers,
PT   especially breast cancer
PS   Claim 10; Page 28-29; 74pp; English.
CC   Human endogenous retroviral sequences 10, 11-29, 3, 6, 12, 13, 14
CC   and 11-22 (T75003-10) were obtd. by screening human genomic
CC   libraries using human breast tumour-associated retroviral element
CC   B18Ag1 (see also T75002) as probe.  These non-contiguous sequences
CC   lie in order 11-22, 14, B18Ag-1, 13, 12, 10, 3, 11-29, 6 in the
CC   retrovirus genome (see also T75001).  B18Ag1 and the other
CC   retroviral sequences can be used in genetic vaccines and for the
CC   prognosis, diagnosis and monitoring of human breast cancer.
SQ   Sequence   1163 BP;    257 A;    289 C;    176 G;    377 T;

  Query Match           24.6%;  Score 15;  DB 32;  Length 1163;
  Best Local Similarity 100.0%;  Pred. No. 2.97e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```



```
Db    185 tgttttatttttgttt 199
          ||||||||||||||||
Cp     33 TGTTTTATTTTGTTT 19


RESULT   20
ID   Q31826 standard; DNA; 1227 BP.
AC   Q31826;
DT   16-APR-1993  (first entry)
DE   Recombinant TCS sequence (pQ21D).
KW   TCS; alpha-trichosanthin; Radix-trichosanthin; primer; inhibition;
KW   viral expression; HIV; T-cell; macrophage; ss.
OS   Trichosanthes kirilowii.
FH   Key             Location/Qualifiers
FT   cds             340..1209
FT                   /*tag=  a
FT                   /label= TCS
FT   signal_peptide  342..410
FT                   /*tag=  b
FT                   /note= "putative N-terminal extension of the
FT                   mature TCS"
FT   mat_peptide     411..1151
FT                   /*tag=  c
FT                   /note= "Claim 3; page 36"
FT   misc_rna        1152..1208
FT                   /*tag=  d
FT                   /note= "putative C-terminal extension of the
FT                   mature TCS"
PN   US5166056-A.
PD   24-NOV-1992.
PF   04-APR-1989; 333184.
PR   04-APR-1989; US-333184.
PR   07-SEP-1989; US-404326.
PR   09-DEC-1991; US-804293.
PA   (GENE-) GENELABS INC.
PI   Chow TP, Piatak M;
DR   WPI; 92-414954/50.
DR   P-PSDB; R29272.
PT   Recombinant Tricosanthin protein prodn. in E. coli - for use in
PT   the selective inhibition of viral expression in HIV infected
PT   cells
PS   Disclosure; Fig 4; 37pp; English.
CC   The TCS coding sequence was amplified using the primers of Q31828-30.
CC   The amplified prod. has the sequence of Q31827, which was used as a
CC   probe.  One isolate, pQ21D, comprises the sequence of Q31826.
CC   The recombinant TCS sequence may be used in the recombinant prodn.
CC   of TCS.  TCS can be used for the selective inhibition of viral
CC   expression in HIV-infected human T-cells or macrophages.
SQ   Sequence   1227 BP;    402 A;    250 C;    223 G;    352 T;

  Query Match           24.6%;  Score 15;  DB 5;  Length 1227;
```

```
  Best Local Similarity 100.0%;  Pred. No. 2.97e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     67 ttttcaaaaacaaaa 81
          |||||||||||||||
Qy     12 TTTTCAAAAACAAAA 26


RESULT   21
ID   Q06343 standard; DNA; 1227 BP.
AC   Q06343;
DT   06-FEB-1991  (first entry)
DE   Encodes Trichosanthin from Trichosanthes kirilowii.
KW   trichosanthin (TCS); Human Immunodeficiency Virus; inhibition; ss.
OS   Trichosanthes kirilowii.
FH   Key             Location/Qualifiers
FT   cds             340..1209
FT                   /*tag=  a
FT   mat_peptide     409..1209
FT                   /*tag=  b
FT                   /product=TCS
PN   WO9012097-A.
PD   18-OCT-1990.
PF   04-APR-1990; U01816.
PR   04-APR-1989; US-333184.
PA   (GENE-) Genelabs Inc.
PI   Piatek M, Chow T, Fry K;
DR   WPI; 90-334847/44.
DR   P-PSDB; R07514.
PT   Recombinant trichosanthin protein - with selective inhibitory effect
PT   on viral expression in HIV infected T-cells or monocytes/macrophage.
PS   Claim 2; Page 59; 102pp; English.
CC   Genomic DNA was isolated from T.kirilowii leaves from Korea and a
CC   library was constructed. Clone pQ21D was identified as likely to
CC   contain a TCS-encoding sequence in its 4kb insert. The deduced
CC   amino acid sequence is identical to that of TCS purified from
CC   Cantonese T.kirilowii roots, except for 2 conservative substitu-
CC   tions. The isolated coding sequence can be used to produce TCS in
CC   pure form for use as a therapeutic agent.
CC   See also Q06344-Q06351.
SQ   Sequence   1227 BP;    402 A;    250 C;    223 G;    352 T;

  Query Match           24.6%;  Score 15;  DB 1;  Length 1227;
  Best Local Similarity 100.0%;  Pred. No. 2.97e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     67 ttttcaaaaacaaaa 81
          |||||||||||||||
Qy     12 TTTTCAAAAACAAAA 26
```



```
RESULT   22
ID   Q26499 standard; DNA; 1228 BP.
AC   Q26499;
DT   13-JAN-1993  (first entry)
DE   Trichosanthin coding sequence from T.kirilowii.
KW   TCS; alpha-trichosanthin; Radix trichosanthin; abortifacient;
KW   ribosome inactivating protein; RIP; HIV-infected human T cells;
KW   human immunodeficiency virus; ss.
OS   Trichosanthes kirilowii.
FH   Key             Location/Qualifiers
FT   cds             341..1210
FT                   /*tag=  a
FT   mat_peptide     410..1150
FT                   /*tag=  b
FT                   /product= trichosanthin
PN   US5128460-A.
PD   07-JUL-1992.
PF   04-APR-1989; 333184.
PR   04-APR-1989; US-333184.
PA   (GENE-) GENELABS INC.
PI   Chow TP, Fry K, Piatak M;
DR   WPI; 92-249485/30.
DR   P-PSDB; R25572.
PT   Nucleic acid encoding trichosanthin protein - which can be used
PT   to inactivate ribosome(s) to inhibit protein synthesis or to
PT   inhibit HIV expression
PS   Claim 1; Column 3; 53pp; English.
CC   Trichosanthin protein was isolated from T.kirilowii root tuber and
CC   sequenced. The amino acid sequence was used to design sets of
CC   degenerate primers (see Q26506-8) for PCR amplification of the TCS
CC   coding sequence. The amplified product was used as a probe to
CC   isolate TCS coding sequence from T.kirilowii genomic libraries. One
CC   clone (pQ21D) contained a 4kb insert. The amino acid sequence deduced
CC   from the pQ21D insert differed from the purified TCS sequence at
CC   amino acids 211 and 224. The substitutions at each position were
CC   conservative and are postulated to indicate minor variations between
CC   strains; the purified TCS was obtained from the Canton region of
CC   China and the genomic DNA was obtained from T.kirilowii leaves from
CC   Korea. The claimed nucleic acid encodes at least the mature TCS and
CC   may further include the regions coding for the N- and C-terminal
CC   amino acid extensions or encodes TCS fused to a ligand peptide
CC   which confers cell-surface recognition properties on the fused
CC   protein. See also Q26500-Q26505.
SQ   Sequence   1228 BP;    403 A;    250 C;    223 G;    352 T;

  Query Match           24.6%;  Score 15;  DB 4;  Length 1228;
  Best Local Similarity 100.0%;  Pred. No. 2.97e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     67 ttttcaaaaacaaaa 81
          |||||||||||||||
```



```
Qy     12 TTTTCAAAAACAAAA 26


RESULT   23
ID   T16869 standard; DNA; 1233 BP.
AC   T16869;
DT   14-MAY-1996  (first entry)
DE   Pepper Group 2 constitutive promoter coding sequence.
KW   Pepper Group 2 promoter; transgenic plant; disease-resistance;
KW   crop improvement; mutant acetolactate-synthase; electroporation;
KW   microinjection; particle bombardment; Agrobacterium tumefaciens;
KW   chlorsulfuron-resistance selectable marker; ss.
OS   Capsicum annuum.
FH   Key             Location/Qualifiers
FT   misc_feature    1..1233
FT                   /*tag=  a
FT                   /note= "pepper Group 2 promoter sequence (see
FT                   T168721)"
FT   caat_signal     1100..1103
FT                   /*tag=  b
FT   tata_signal     1139..1146
FT                   /*tag=  c
FT   misc_feature    1169
FT                   /*tag=  d
FT                   /note= "transcriptional start site"
FT   misc_feature    1172
FT                   /*tag=  e
FT                   /note= "pGP50 5' end"
PN   WO9604781-A1.
PD   22-FEB-1996.
PF   10-AUG-1995; U10188.
PR   12-AUG-1994; US-289458.
PA   (DNAP ) DNA PLANT TECHNOLOGY CORP.
PI   Baden CS,  Dunsmuir P,  Lee KY;
DR   WPI; 96-139348/14.
PT   Plant Group 2 promoter sequence isolated from pepper plants  - used
PT   to prepare expression cassettes for transformation of e.g. pepper,
PT   tomato and tobacco plants.
PS   Claim 1; Page 57-58; 76pp; English.
CC   This sequence represents the constitutive pepper Group 2 (Gp2)
CC   promoter sequence up to the ATG start codon.  It is useful in the
CC   production of transgenic plants, preferably pepper, tomato and
CC   tobacco.  Desired phenotypes can be conferred to plants by
CC   transformation of the plant cells with a DNA construct containing
CC   heterologous DNA operably linked to a Gp2 promoter.  The
CC   constitutive promoter may be used for expression of the mutant
CC   acetolactate-synthase selectable marker gene, which confers
CC   chlorsulfuron-resistance.  When introduced into plant cells and
CC   constitutively expressed, it confers a phenotype that is useful in
CC   the production of transgenic plants.  A variety of genes capable of
CC   altering a plant phenotype can also be constitutively expressed,
```



```
CC   e.g. genes for herbicide resistance, fungal disease-resistance (e.g
CC   chitinases and glucanases), bacterial disease-resistance (e.g.
CC   cecropins), and insect resistance (e.g. Bacillus thuringiensis
CC   toxin).  DNA constructs containing a Gp2 promoter linked to
CC   heterologous DNA can be introduced directly into the DNA of the
CC   plant cell using techniques such as electroporation and
CC   microinjection of plant cell protoplasts.  Alternatively, the DNA
CC   constructs can be introduced directly to plant tissues using
CC   biolistic methods, such as DNA particle bombardment.  In addition,
CC   the DNA constructs may be combined with suitable T-DNA flanking
CC   regions and introduced into a conventional Agrobacterium
CC   tumefaciens host vector.
SQ   Sequence   1233 BP;    458 A;    173 C;    116 G;    486 T;

  Query Match           24.6%;  Score 15;  DB 10;  Length 1233;
  Best Local Similarity 100.0%;  Pred. No. 2.97e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1020 aaaaacaaaataaaa 1034
          |||||||||||||||
Qy     17 AAAAACAAAATAAAA 31


RESULT   24
ID   T45418 standard; DNA; 1400 BP.
AC   T45418;
DT   05-FEB-1997  (first entry)
DE   B. subtilis enterobactin type siderophore, pENTA22 spf gene.
KW   pENTA22 spf; enterobactin; siderophore; ferric ion; carrier;
KW   molecule; identification; antibiotic; Gram-positive;
KW   microorganism; ss.
OS   Bacillus subtilis.
FH   Key             Location/Qualifiers
FT   cds             198..695
FT                   /*tag=  a
PN   US5554507-A.
PD   10-SEP-1996.
PF   30-JUL-1992; 922602.
PR   30-JUL-1992; US-922602.
PR   22-APR-1993; US-051142.
PA   (AMCY ) AMERICAN CYANAMID CO.
PI   Grossman TH,  Osburne MS;
DR   WPI; 96-424652/42.
DR   P-PSDB; W07644.
PT   Identifying cpds. antibiotic against microorganisms with
PT   enterobactin-type siderophore - by screening using host cells
PT   transformed with a siderophore gene for growth in a low iron medium
PS   Claim 1; Columns 9-12; 19pp; English.
CC   The present sequence is the Bacillus subtilis derived pENTA22 spf
CC   gene, which encodes an enterobactin type siderophore (ETS) protein
CC   prod., i.e. a ferric ion specific carrier mol.. A host cell
```

```
CC   transformed with the gene can be used to identify cpds. having
CC   potential antibiotic activity against microorganisms with an ETS
CC   mechanism, partic. Gram-positive microorganisms.
CC   The host cell is contacted with the cpd. to be tested, and the
CC   presence or absence of an effect on the host cell's ability to
CC   grow in low iron medium observed, where cpds. that inhibit the
CC   cell's ability to grow in low iron medium are identified as
CC   potential antibiotics.
CC   The pENTA22 spf gene was obtd. by screening B. subtilis chromosomal
CC   DNA for the ability to complement E. coli ent(-) mutants deficient
CC   in different enterobactin biosynthetic enzymes.
SQ   Sequence   1400 BP;    373 A;    334 C;    315 G;    378 T;

  Query Match           24.6%;  Score 15;  DB 24;  Length 1400;
  Best Local Similarity 100.0%;  Pred. No. 2.97e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    561 ttcttttcaaaaaca 575
          ||||||||||||||||
Qy      9 TTCTTTTCAAAAACA 23


RESULT   25
ID   Q62460 standard; DNA; 1400 BP.
AC   Q62460;
DT   09-OCT-1994  (first entry)
DE   Bacillus subtilis siderophore sfp gene.
KW   Siderophore, plasmid pENTA22; surfactin; sfp gene; ss.
OS   Bacillus subtilis.
FH   Key             Location/Qualifiers
FT   cds             198..695
FT                   /*tag=  a
PN   CA2101493-A.
PD   31-JAN-1994.
PF   28-JUL-1993; 101493.
PR   30-JUL-1992; US-922602.
PA   (AMCY ) AMERICAN CYANAMID CO.
PI   Grossman TH,  Osburne MS;
DR   WPI; 94-126966/16.
DR   P-PSDB; R51898.
PT   Nucleic acid encoding Gram-positive bacterial siderophore genes -
PT   used for transforming host cells to screen for cpds. having
PT   potential antibiotic activity
PS   Disclosure; Fig.7; 34pp; English.
CC   This siderophore biosynthetic gene may be expressed in a host, e.g.
CC   E. coli, and the resulting transformant used for identifying
CC   compounds with potential antibiotic activity against microorganisms
CC   with an enterobactin-type siderophore by testing the effect of the
CC   compound on the host's ability to grow in low iron medium.
SQ   Sequence   1400 BP;    373 A;    336 C;    315 G;    376 T;
```



```
  Query Match           24.6%;  Score 15;  DB 10;  Length 1400;
  Best Local Similarity 100.0%;  Pred. No. 2.97e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    561 ttcttttcaaaaaca 575
          ||||||||||||||||
Qy      9 TTCTTTTCAAAAACA 23


RESULT   26
ID   T15760 standard; cDNA; 1510 BP.
AC   T15760;
DT   24-JUL-1997  (revised)
DT   25-JUN-1996  (first entry)
DE   Human interleukin-15 receptor cDNA from clone A133.
KW   interleukin-15 receptor; IL-15R; murine; human; graft rejection;
KW   treatment; prevention; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             3..359
FT                   /*tag=  a
FT                   /note= "partial interleukin-15 receptor"
PN   WO9530695-A.
PD   16-NOV-1995.
PF   04-MAY-1995; U05585.
PR   06-MAY-1994; US-236919.
PR   06-SEP-1994; US-300903.
PA   (IMMV ) IMMUNEX CORP.
PI   Anderson DM,  Giri JG;
DR   WPI; 96-077255/08.
DR   P-PSDB; R90845.
PT   DNA encoding an interleukin-15 receptor - used to ameliorate undesired
PT   effects of interleukin-15, e.g. to prevent or treat graft rejection
PS   Example 3; Page 41-42; 53pp; English.
CC   The present sequence is the cDNA insert of a second clone isolated from
CC   an A172 library. The clone exhibits an incomplet 5' region which begins
CC   at the equivalent of nucleotide 355 of T15995, and an in-frame insertion
CC   downstream of the transmembrane region that results in a different
CC   cytoplasmic tail coding domain.  A hybrid construct encoding the 5' half
CC   of W5 fused to A133 to give the alternate cytoplasmic tail (R90846) was
CC   prepared and expressed.  Cell surface binding experiments using radio-
CC   labeled human IL-15 showed that this hybrid construct encoded a
CC   polypeptide that bound IL-15. The IL-15R DNA is a multi-exon gene. IL-15R
CC   variants can be attributed to different mRNA splicing events following
CC   transcription or from proteolytic cleavage of the IL-15R protein, where
CC   the IL-15R binding property is retained. The IL-15R protein is useful in
CC   treatment of undesired effects of interleukin-15, e.g. to prevent or
CC   treat graft rejection.
CC   (Revised entry submitted to correct sequence analysis breakdown.)
SQ   Sequence   1510 BP;    391 A;    422 C;    350 G;    347 T;
```



```
  Query Match           24.6%;  Score 15;  DB 18;  Length 1510;
  Best Local Similarity 100.0%;  Pred. No. 2.97e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1208 aaacaaaataaaaca 1222
          ||||||||||||||||
Qy     19 AAACAAAATAAAACA 33


RESULT   27
ID   T15995 standard; cDNA; 1534 BP.
AC   T15995;
DT   24-JUL-1997  (revised)
DT   08-JUN-1996  (first entry)
DE   Human interleukin-15 receptor cDNA.
KW   interleukin-15 receptor; IL-15R; murine; human; graft rejection;
KW   treatment; prevention; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             1..756
FT                   /*tag=  a
FT                   /product= immature_interleukin-15_receptor
FT   5'utr           757..1534
FT                   /*tag=  b
FT   signal_peptide  1..33
FT                   /*tag=  c
FT   mat_peptide     34..753
FT                   /*tag=  d
PN   WO9530695-A.
PD   16-NOV-1995.
PF   04-MAY-1995; U05585.
PR   06-MAY-1994; US-236919.
PR   06-SEP-1994; US-300903.
PA   (IMMV ) IMMUNEX CORP.
PI   Anderson DM,  Giri JG;
DR   WPI; 96-077255/08.
DR   P-PSDB; R91594.
PT   DNA encoding an interleukin-15 receptor - used to ameliorate undesired
PT   effects of interleukin-15, e.g. to prevent or treat graft rejection
PS   Claim 1; Page 23-34; 53pp; English.
CC   The cDNA shows a human open reading frame (ORF) encoding interleukin-15
CC   receptor (IL-15R) (see R91594), isolated from clone W5. It was isolated
CC   by cross-species hybridisation with the murine IL-15R cDNA (T15757). The
CC   mammalian IL-15R is encoded by a multi-exon gene. IL-15R variants can be
CC   attributed to different mRNA splicing events following transcription or
CC   from proteolytic cleavage of the IL-15R protein, where the IL-15R binding
CC   property is retained. The IL-15R protein is useful in treatment of
CC   undesired effects of interleukin-15, e.g. to prevent or treat graft
CC   rejection.
CC   (Revised entry submitted to correct sequence analysis breakdown.)
SQ   Sequence   1534 BP;    425 A;    445 C;    354 G;    310 T;
```



```
  Query Match           24.6%;  Score 15;  DB 18;  Length 1534;
  Best Local Similarity 100.0%;  Pred. No. 2.97e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1440 aaacaaaataaaaca 1454
          ||||||||||||||||
Qy     19 AAACAAAATAAAACA 33


RESULT   28
ID   T15758 standard; cDNA; 1641 BP.
AC   T15758;
DT   24-JUL-1997  (revised)
DT   25-JUN-1996  (first entry)
DE   Human interleukin-15 receptor cDNA from clone P1.
KW   interleukin-15 receptor; IL-15R; murine; human; graft rejection;
KW   treatment; prevention; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             3..842
FT                   /*tag=  a
FT                   /product= interleukin-15_receptor
FT                   /note= "clone P1"
PN   WO9530695-A.
PD   16-NOV-1995.
PF   04-MAY-1995; U05585.
PR   06-MAY-1994; US-236919.
PR   06-SEP-1994; US-300903.
PA   (IMMV ) IMMUNEX CORP.
PI   Anderson DM,  Giri JG;
DR   WPI; 96-077255/08.
DR   P-PSDB; R90843.
PT   DNA encoding an interleukin-15 receptor - used to ameliorate undesired
PT   effects of interleukin-15, e.g. to prevent or treat graft rejection
PS   Example 3; Page 35-37; 53pp; English.
CC   A lambda-gt10 library from human peripheral blood lymphocytes, was
CC   screened for the presence of a full-length clone encoding human
CC   interleukin-15 receptor (IL-15R) using a random prime labeled human
CC   IL-15R cDNA probe consisting of the entier W5 cDNA clone (T99999)
CC   without the polyA tail.  The resulting cDNA insert from this clone
CC   T15758 encodes R90843, and exhibited an in-frame insertion of 153 bp
CC   at the mature N-terminus (amino acids 24 through 74), an in-frame
CC   deletion of 99 bp downstream of the mature N-terminus that deleted
CC   nucleotides 236 to 334 of T15995, and also contained additional 5'
CC   sequence as compared to clone W5 (amino acids 1 through 10), but still
CC   did not contain an initiator methionine. The IL-15R DNA is a multi-exon
CC   gene. IL-15R variants can be attributed to different mRNA splicing events
CC   following transcription or from proteolytic cleavage of the IL-15R
CC   protein, where the IL-15R binding property is retained. The IL-15R
CC   protein is useful in treatment of undesired effects of interleukin-15,
```

```
CC   e.g. to prevent or treat graft rejection.
CC   (Revised entry submitted to correct sequence analysis breakdown.)
SQ   Sequence   1641 BP;    478 A;    463 C;    385 G;    315 T;

  Query Match          24.6%;  Score 15;  DB 18;  Length 1641;
  Best Local Similarity 100.0%;  Pred. No. 2.97e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1526 aaacaaaataaaaca 1540
          ||||||||||||||||
Qy     19 AAACAAAATAAAACA 33


RESULT   29
ID   T16871 standard; DNA; 1727 BP.
AC   T16871;
DT   14-MAY-1996  (first entry)
DE   Pepper Group 2 gene fragment coding sequence.
KW   Pepper Group 2 gene fragment; Gp2 constitutive promoter;
KW   Gp2 protein; multigene family; Bowman Birk protease-inhibitor;
KW   thionin; alpha-amylase-inhibitor; transgenic plant; ss.
OS   Capsicum annuum.
FH   Key             Location/Qualifiers
FT   misc_feature    1..1727
FT                   /*tag= a
FT                   /note= "Group 2 genomic DNA clone"
FT   misc_feature    1..1233
FT                   /*tag= b
FT                   /note= "Group 2 promoter sequence (see T16869)"
FT   caat_signal     1100..1103
FT                   /*tag= c
FT   tata_signal     1139..1146
FT                   /*tag= d
FT   misc_feature    1169
FT                   /*tag= e
FT                   /note= "transcriptional start site"
FT   misc_feature    1172
FT                   /*tag= f
FT                   /note= "pGP50 5' end"
FT   misc_feature    1234..1236
FT                   /*tag= g
FT                   /note= "translational start codon"
FT   misc_feature    1298
FT                   /*tag= h
FT                   /note= "intron start site"
PN   WO9604781-A1.
PD   22-FEB-1996.
PF   10-AUG-1995; U10188.
PR   12-AUG-1994; US-289458.
PA   (DNAP ) DNA PLANT TECHNOLOGY CORP.
PI   Baden CS,  Dunsmuir P,  Lee KY;
```



```
DR   WPI; 96-139348/14.
PT   Plant Group 2 promoter sequence isolated from pepper plants  - used
PT   to prepare expression cassettes for transformation of e.g. pepper,
PT   tomato and tobacco plants.
PS   Claim 30; Page 56-57; 76pp; English.
CC   Pepper Group 2 gene fragment, consisting of a short form of a Gp2;
CC   constitutive promoter (see T16869), including the first exon and a;
CC   part of the first intron, operatively linked to a nucleic acid;
CC   sequence encoding a Gp2 protein (see R92752).  The Gp2 protein is;
CC   highly enriched in sulphur-containing amino acids and shows some;
CC   homology with Bowman Birk protease-inhibitors, thionins and alpha-;
CC   amylase-inhibitors.  Gp2 proteins appear to form a multigene family;
CC   in pepper plants.  Full length or oligonucleotide DNA probes may be;
CC   generated by comparison of this sequence and T16870.  Such probes;
CC   can be used directly in hybridization assays to isolate DNA encoding.
SQ   Sequence   1727 BP;    626 A;    243 C;    202 G;    656 T;

  Query Match          24.6%;  Score 15;  DB 18;  Length 1727;
  Best Local Similarity 100.0%;  Pred. No. 2.97e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1020 aaaaacaaaataaaa 1034
          ||||||||||||||||
Qy     17 AAAAACAAAATAAAA 31


RESULT   30
ID   Q58742 standard; cDNA; 2161 BP.
AC   Q58742;
DT   22-SEP-1994  (first entry)
DE   dMEF2.
KW   MEF2; myocyte-specific enhancer factor; isoform; MADS domain;
KW   myocyte-specific transcipton enhancer factor; muscle; atrophy; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             661..2058
FT                   /*tag= a
FT                   /label= MEF2
FT   misc_rna        667..840
FT                   /*tag= b
FT                   /label= putative_MADS_domain
FT   misc_rna        841..918
FT                   /*tag= c
FT                   /note= "region necessary for sequence
FT                   specificity of the MEF2
FT                   related factors"
FT   misc_rna        1678..1773
FT                   /*tag= d
FT                   /note= "alternatively spliced
FT                   (96nt) region"
PN   WO9405776-A.
```



```
PD   17-MAR-1994.
PF   07-SEP-1993; U08386.
PR   04-SEP-1992; US-939898.
PA   (CHIL-) CHILDRENS MEDICAL CENT.
PI   Nadal-Ginard B;
DR   WPI; 94-101181/12.
DR   P-PSDB; R50041.
PT   New nucleic acid encoding MEF2-protein family member - i.e.
PT   myocyte-specific transcription enhancing factor, useful to
PT   increase muscle mass and to counteract muscle atrophy, e.g. as in
PT   muscular dystrophy
PS   Disclosure; Page 66-69; 129pp; English.
CC   MEF2 transcription factors can be used to produce transgenic animals
CC   with increased muscle mass, to prevent or counteract muscle atrophy
CC   in humans or animals suffering a pathological muscular condition, or
CC   to develop pharmacological agents which regulate the expression of
CC   muscle-tissue genes.  MEF2 isoforms are given in Q58740-44,
CC   R50038-45 and R57772.  Nucleic acid encoding a member of the MEF2
CC   protein family pref. comprises a sequence encoding at least eleven
CC   consecutive glutamine residues or a sequence encoding the amino acid
CC   sequence SEEEELEL (R57773).
SQ   Sequence   2161 BP;    678 A;    486 C;    472 G;    523 T;

  Query Match          24.6%;  Score 15;  DB 10;  Length 2161;
  Best Local Similarity 100.0%;  Pred. No. 2.97e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     58 aaaaacaaaataaaa 72
          ||||||||||||||||
Qy     17 AAAAACAAAATAAAA 31


RESULT   31
ID   T51715 standard; cDNA to mRNA; 2623 BP.
AC   T51715;
DT   30-SEP-1997  (first entry)
DE   cDNA for adenovirus E1A associated protein BS69.
KW   Adenovirus; early region 1A; E1A; transactivation inhibitor; 243R;
KW   289R; viral replication inhibitor; immunoassay; gene therapy; ss.
OS   Adenovirus.
FH   Key             Location/Qualifiers
FT   cds             150..1838
FT                   /*tag= a
PN   WO9700323-A1.
PD   03-JAN-1997.
PF   14-JUN-1996; G01413.
PR   14-JUN-1995; GB-012092.
PA   (VEHE-) VER HET NEDERLANDS KANKER INST.
PI   Bernards R,  Hateboer G;
DR   WPI; 97-077529/07.
DR   P-PSDB; W12096.
```



```
PT   New adenovirus E1A associated protein BS69 - useful for inhibiting
PT   trans:activation, e.g. as antiviral agent and for gene therapy of
PT   tumours.
PS   Claim 5; Page 33-36; 49pp; English.
CC   T51715 encodes the adenovirus E1A-associated protein BS69. BS69 can
CC   bind to early region 1A (E1A) proteins 243R and 289R to inhibit the
CC   transactivation activity of 289R. In doing so the peptide also inhibits
CC   viral replication and is therefore useful as an anti-adenoviral agent.
CC   Vectors encoding the BS69 protein are also used in vivo for gene
CC   therapy (e.g. to treat uncontrolled proliferation of tumour cells).
SQ   Sequence   2623 BP;    849 A;    499 C;    604 G;    671 T;

  Query Match          24.6%;  Score 15;  DB 32;  Length 2623;
  Best Local Similarity 100.0%;  Pred. No. 2.97e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2599 aacaaaataaaacaa 2613
          ||||||||||||||||
Qy     20 AACAAAATAAAACAA 34


RESULT   32
ID   Q32864 standard; DNA; 2655 BP.
AC   Q32864;
DT   05-MAY-1993  (first entry)
DE   Tcr gene derived from Bacteroides conjugal element TcrEmr-DOT donor.
KW   Prevotella ruminicola; genetic manipulation; genetic engineering;
KW   recombinant DNA technology; shuttle vectors; DNA transfer;
KW   transformation; conjugal transfer; ss.
OS   Bacteroides sp.
FH   Key             Location/Qualifiers
FT   misc_difference 1880..1882
FT                   /*tag= a
FT                   /codon= seq:"TAA", aa:Ile
FT                   /note= "normally TAA is the ochre codon."
FT   misc_difference 1934..1936
FT                   /*tag= b
FT                   /codon= seq:"GAT", aa:Glu
FT   misc_difference 2120..2122
FT                   /*tag= c
FT                   /codon= seq:"ACA", aa:Tyr
FT   misc_difference 1855..1857
FT                   /*tag= d
FT                   /codon= seq:"GGA", aa:Arg
FT   cds             584..2506
FT                   /*tag= e
PN   WO9221761-A.
PD   10-DEC-1992.
PF   01-JUN-1992; U04546.
PR   05-JUN-1991; US-718535.
PA   (BIOT-) BIOTECHNOLOGY RES & DEV CORP.
```

```
PA   (UNII ) UNIV ILLINOIS FOUND.
PI   Nikolich MP, Salyers AA, Shoemaker NB;
DR   WPI; 92-433659/52.
DR   P-PSDB; R30089.
PT   Introducing heterologous DNA into Prevotella ruminicola - by
PT   conjugal transfer of a shuttle vector comprising the heterologous
PT   DNA operatively linked to a promoter functional in P. ruminicola
PS   Claim 19; Page 55; 85pp; English.
CC   This sequence represents the Bacteroides Tcr gene, originally
CC   derived from the Bacteroides conjugal element TcrEmr-DOT.
CC   [Shoemaker et al., J. Bacteriol., 171:1294-1302 (1989)]. This was
CC   subcloned on a 2.7kbp fragment, and the whole of this fragment then
CC   sequenced. The gene codes for a protein of the ribosome protection
CC   type of tetracycline resistance. However, the amino acid sequence
CC   coded for by the cloned gene was found to be only about 40%
CC   identical to sequences coded for by the TetM and TetO genes. It was
CC   therefore concluded that the bacteroides Tcr gene was clearly in a
CC   separate DNA-DNA hybridisation class from TetM and TetO, and the
CC   new class of tetracycline resistance genes designated TetQ.
SQ   Sequence   2655 BP;    867 A;    485 C;    533 G;    770 T;

  Query Match           24.6%;  Score 15;  DB 6;  Length 2655;
  Best Local Similarity 100.0%;  Pred. No. 2.97e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1996 ccattcttttcaaaa 2010
            |||||||||||||||
Qy      6 CCATTCTTTTCAAAA 20


RESULT   33
ID   N71250 standard; DNA; 2844 BP.
AC   N71250;
DT   17-APR-1991  (first entry)
DE   Sequence of Vicia fabia type B gene.
KW   Plant seed protein; nutrition; ss.
OS   Vicia fabia.
FH   Key             Location/Qualifiers
FT   exon            746..1294
FT                   /*tag=  a
FT   exon            1390..1917
FT                   /*tag=  b
FT   exon            2018..2395
FT                   /*tag=  c
PN   DD-240911-A.
PD   19-NOV-1986.
PF   31-DEC-1983; 259075.
PR   31-DEC-1983; DD-259075.
PA   (DEAK ) AKAD WISSENSCHAFT DDR.
PI   Muntz K, Bassuner R, Saalbach G, Jung R, Wobus U, Weschke W,
PI   Baumlein H;
```



```
DR   WPI; 87-080534/12.
DR   P-PSDB; P71081.
PT   Prodn. of plant seed protein genes - by incorporation of codon(s)
PT   for relevant limiting amino acids, e.g. for methionine in
PT   leguminous protein legumin
PS   Example; pp6-9; 12pp; German.
CC   The patent claims a method for the prodn. of nutrition-
CC   physiologically improved plant-seed proteins. Suitable plant seed
CC   protein genes are the genes for legumin or vicilin from field beans,
CC   or corresponding suitable genes for seed proteins of cereals.
SQ   Sequence   2844 BP;    945 A;    584 C;    534 G;    781 T;

  Query Match           24.6%;  Score 15;  DB 2;  Length 2844;
  Best Local Similarity 100.0%;  Pred. No. 2.97e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2486 aataaaacaataaag 2500
            |||||||||||||||
Qy     25 AATAAAACAATAAAG 39


RESULT   34
ID   T41616 standard; DNA; 2960 BP.
AC   T41616;
DT   11-JUN-1997  (first entry)
DE   Adzuki bean endo-xyloglucan transferase 2 promoter gene.
KW   Adzuki; bean; promoter; induction; expression; plant; cell; wall;
KW   xyloglucan; reconstitution; enzyme; endo-xyloglucan; transferase;
KW   transgenic; increase; yield; alter; change; function;
KW   morphology; EXT2; ds.
OS   Phaseolus angularis.
OS   Vigna angularis.
PN   WO9630509-A1.
PD   03-OCT-1996.
PF   26-MAR-1996; J00777.
PR   30-MAR-1995; JP-073043.
PA   (TAKI ) TAKARA SHUZO CO LTD.
PI   Anma Y,  Asada K,  Kato I,  Oba T,  Takahashi S;
DR   WPI; 96-455356/45.
PT   Plant endo-xylo:glucan transferase promoters - allowing expression
PT   of plant genes at appropriate times to influence plant morphology
PS   Claim 9; Pages 129-131; 206pp; Japanese.
CC   The present sequence is an adzuki bean (Phaseolus angularis,
CC   given as Vigna angularis in the specification) promoter gene,
CC   which induces the expression of a plant cell wall xyloglucan
CC   reconstitution enzyme, specifically endo-xyloglucan
CC   transferase 2 (EXT2). The promoter gene can be incorporated into
CC   transgenic plants to alter function, e.g. to express particular
CC   proteins or increase plant yield, or change the plant morphology.
SQ   Sequence   2960 BP;    1205 A;    397 C;    270 G;    1088 T;
```



```
  Query Match           24.6%;  Score 15;  DB 29;  Length 2960;
  Best Local Similarity 100.0%;  Pred. No. 2.97e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    415 tttttgaaaagaatg 429
          |||||||||||||||
Cp     21 TTTTTGAAAAGAATG 7


RESULT   35
ID   Q21188 standard; DNA; 3012 BP.
AC   Q21188;
DT   16-MAY-1992  (first entry)
DE   Sequence encoding a 302 amino acid endochitinase which lacks a 3'
DE   region, from plasmid pCH3.5.
KW   Pest-resistant plant; transformed plant; fungi; insect; bacteria;
KW   nematode; ss.
OS   Tomato.
FH   Key             Location/Qualifiers
FT   exon            1942..2383
FT                   /*tag=  a
FT   intron          2384..2443
FT                   /*tag=  b
FT   exon            2444..2597
FT                   /*tag=  c
FT   intron          2598..2677
FT                   /*tag=  d
FT   exon            2678..3012
FT                   /*tag=  e
PN   WO9201792-A.
PD   06-FEB-1992.
PF   24-JUL-1991; F00607.
PR   24-JUL-1990; FR-009460.
PA   (SNFI ) SANOFI SA.
PA   (ERAP ) SOC NAT ELF AQUITAINE.
PI   Dubois M, Grison R, Leguay JJ, Pignard A, Toppan A;
DR   WPI; 92-064950/08.
DR   P-PSDB; R20820.
PT   New recombinant protein with endo chitinase activity - also DNA
PT   encoding it and plants which express it, resistant to fungi,
PT   insects, bacteria and nematodes.
PS   Example; Fig 2; 82pp; French.
CC   The inventors claim a new recombinant gene which codes for a protein
CC   with encochitinase activity, or its precursor (see Q21007).  The
CC   coding part of the gene contains at least the 5' part of DNA for
CC   tomato endochitinase and at least the 3' part of DNA for the tobacco
CC   enzyme (DNA can be genomic or cDNA, but at least one intron is pref.
CC   present).  The recombinant gene includes the 35S promoter of
CC   cauliflower mosaic virus plus the terminator from the nopaline
CC   synthase gene of Agrobacterium tumefaciens.
SQ   Sequence   3012 BP;    965 A;    634 C;    436 G;    977 T;
```



```
  Query Match           24.6%;  Score 15;  DB 3;  Length 3012;
  Best Local Similarity 100.0%;  Pred. No. 2.97e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    231 tttgtttttgaaaag 245
          |||||||||||||||
Cp     25 TTTGTTTTTGAAAAG 11


RESULT   36
ID   Q49956 standard; cDNA; 3017 BP.
AC   Q49956;
DT   29-APR-1994  (first entry)
DE   Impatiens Necrotic Spot Virus S RNA cDNA clone.
KW   INSV; tospovirus; resistance; pan-handle secondary structure;
KW   plant RNA virus; ss.
OS   Impatiens Necrotic Spot Virus.
FH   Key             Location/Qualifiers
FT   cds             87..1436
FT                   /*tag=  a
FT                   /product= Non-Structural_protein
FT                   /note= "predicted mol.wt. 51.2kD"
FT   cds             complement (2080..2868)
FT                   /*tag=  b
FT                   /product= Nucleocapsid_protein
FT                   /note= "predicted mol.wt. 28.7kd"
FT   repeat_unit     1..36
FT                   /*tag=  c
FT                   /rpt_type= TERMINAL
FT                   /note= "forms pan-handle structure with
FT                   complementary terminal repeat at 3' end;
FT                   not known whether there are 16 or 17 N's
FT                   but the exact number is not critical for
FT                   pan-handle formation"
FT   repeat_unit     2982..3017
FT                   /*tag=  d
FT                   /rpt_type= TERMINAL
FT                   /note= "forms pan-handle structure with
FT                   complementary terminal repeat at 5' end"
FT   misc_rna        1437..2079
FT                   /*tag=  e
FT                   /note= "intergenic region can be folded into sec.
FT                   structure"
PN   EP-566525-A.
PD   20-OCT-1993.
PF   16-MAR-1993; 810190.
PR   19-MAR-1992; GB-006016.
PA   (SANO ) SANDOZ LTD.
PA   (VGRI/) VAN GRINSVEN M Q J M.
PA   (SANO ) SANDOZ PATENT GMBH.
```

```
PA   (SANO ) SANDOZ-ERFINDUNGEN VERW GES MBH.
PA   (SANO ) SANDOZ AG.
PI   De Haan PT,  Gielen JJL,  Goldbach RW,  Peters D,  Van Grinsven MQJ;
PI   Van Grinsven MQJM;
DR   WPI; 93-329814/42.
DR   P-PSDB; R42078, R42079.
PT   Recombinant INSV DNA constructs comprising INSV DNA coding for
PT   transcription into INSV RNA sequences - are used to transform
PT   plants to reduce susceptibility to INSV infection
PS   Claim 2; Page 16-17; 64pp; English.
CC   INSV RNA was purified from systemically infected Nicotiana rustica
CC   leaves. The short (S) transcript comprises two genes, one coding for
CC   a non-structural protein and the other (on the complementary strand)
CC   coding for the nucleocapsid protein. The 5'- and 3'-terminal
CC   sequences of the S RNA are capable of hybridising to each other; the
CC   double-stranded structure obtained by such hybridisation is referred
CC   to as a "pan-handle".
SQ   Sequence   3017 BP;   976 A;   528 C;   456 G;   1041 T;

  Query Match          24.6%;  Score 15;  DB 8;  Length 3017;
  Best Local Similarity 100.0%;  Pred. No. 2.97e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1764 ttttattttgttttt 1778
          |||||||||||||||
Cp     31 TTTTATTTTGTTTTT 17


RESULT   37
ID   Q14263 standard; cDNA; 3207 BP.
AC   Q14263;
DT   06-FEB-1992  (first entry)
DE   p-meta-1 metastasis-specific variant of CD44 glycoprotein.
KW   Tumour therapy; CD44 surface protein; ss.
OS   Rattus rattus.
FH   Key             Location/Qualifiers
FT   cds             113..1625
FT                   /*tag=  a
FT   signal_peptide  113..175
FT                   /*tag=  b
FT   mat_peptide     176..1625
FT                   /*tag=  c
FT   misc_feature    782..1249
FT                   /*tag=  d
FT                   /note= "r-meta-1 "
FT   polya_signal    3143..3148
FT                   /*tag=  e
FT   polya_site      3164..3207
FT                   /*tag=  f
PN   DE4014510-A.
PD   14-NOV-1991.
```



```
PF   07-MAY-1990; 014510.
PR   07-MAY-1990; DE-014510.
PA   (GESL ) KERNFORSCHUNGS KARLSRUHE.
PA   (UYKA-) UNIV KARLSRUHE.
PA   (DEKR-) DEUT KREBSFORSCHUNGSZENT.
PI   Herrlich P, Ponta H, Gunthert U, Matzku S, Wenzel A;
DR   WPI; 91-340908/47.
DR   P-PSDB; R14768.
PT   Variant CD44-surface proteins - and cDNA sequences and antibodies
PT   for proteins, used in diagnosis and therapy of tumour metastases
PS   Disclosure; Fig 2; 12pp; German.
CC   The sequence is that of the cDNA clone p-meta-1 which represents a
CC   metastasis-specific variant of CD44 surface glycoprotein. The
CC   protein is useful in tumour and metastasis therapy, by its
CC   action as a ligand blocker. Antibodies can also be raised to the
CC   protein and these can be similarly useful when linked to cytotoxic
CC   agents. The indicated r-meta-1 clone corresponds to the
CC   extracellular region.See also Q14264.
SQ   Sequence   3207 BP;   939 A;   763 C;   681 G;   824 T;

  Query Match          24.6%;  Score 15;  DB 3;  Length 3207;
  Best Local Similarity 100.0%;  Pred. No. 2.97e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2213 ttttattttgttttt 2227
          |||||||||||||||
Cp     31 TTTTATTTTGTTTTT 17


RESULT   38
ID   Q04476 standard; cDNA; 3649 BP.
AC   Q04476;
DT   05-OCT-1990  (first entry)
DE   Sequence derived by reverse transcription of RNA1 of AMV.
KW   Alfalfa mosaic virus; AMV; cryptic infection; probe; ss.
OS   Nicotiniana tabacum L 'Xanthi-nc'
FH   Key             Location/Qualifiers
FT   cds             100..3486
FT                   /*tag=  a
FT                   /note="protein encoded by RNA1"
PN   US4921802-A.
PD   01-MAY-1990.
PF   13-JAN-1988; 144692.
PR   05-MAY-1983; US-492582.
PR   13-JAN-1988; US-144692.
PA   (PIOE) Pioneer Hi-Bred International Ltd.
PI   Hall TC, Loesch-Fries SL, Jarvis NP, Barker RF;
DR   WPI; 90-171297/22.
DR   P-PSDB; R05107.
PT   Recombinant cDNA plasmids derived from alfalfa mosaic virus RNA -
PT   for detection of cryptic infection by alfalfa mosaic virus.
```



```
PS   Disclosure; p; English.
CC   The sequence is that of the bottom (B)- or "RNA1" genome fragment
CC   of AMV which is necessary to initiate infection. The sequence can
CC   be used as a probe to detect cryptic infection of plants by AMV.
CC   See also Q04477 and 78.
SQ   Sequence   3649 BP;   1039 A;   744 C;   808 G;   1058 T;

  Query Match          24.6%;  Score 15;  DB 1;  Length 3649;
  Best Local Similarity 100.0%;  Pred. No. 2.97e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1776 ccattcttttcaaaa 1790
          |||||||||||||||
Qy      6 CCATTCTTTTCAAAA 20


RESULT   39
ID   T32143 standard; DNA; 4495 BP.
AC   T32143;
DT   30-JAN-1997  (first entry)
DE   Helicobacter-specific ATPase gene 948.
KW   Helicobacter pylori; ATPase; model; screening; metabolism; ss.
OS   Helicobacter pylori.
FH   Key             Location/Qualifiers
FT   cds             10..939
FT                   /*tag=  a
FT                   /label= ORF-1
FT   cds             1162..1875
FT                   /*tag=  b
FT                   /label= ORF-2
FT   cds             1710..1856
FT                   /*tag=  c
FT                   /label= ORF-3
FT   cds             1872..4097
FT                   /*tag=  d
FT                   /label= ORF-4
FT   cds             4098..4298
FT                   /*tag=  e
FT                   /label= ORF-5
FT   cds             4401..4481
FT                   /*tag=  f
FT                   /label= ORF-6
FT                   /note= "incomplete at end"
PN   WO9617066-A1.
PD   06-JUN-1996.
PF   30-NOV-1995; E04711.
PR   02-DEC-1994; DE-442970.
PR   18-FEB-1995; DE-005645.
PA   (BYKG ) BYK GULDEN LOMBERG CHEM FAB.
PI   Melchers K,  Schafer KP,  Steinhilber W,  Weitzenegger T;
DR   WPI; 96-286831/29.
```



```
DR   P-PSDB; R97282, R83570, R83571, R83572, R83573, R83574.
PT   Model for screening cpds. for inhibition of Helicobacter - is
PT   recombinant cell transformed with gene for Helicobacter P-type
PT   ATPase under control of inducible promoter
PS   Claim 13; Fig 4; 52pp; German.
CC   Helicobacter-specific ATPase genes 439, 514 and 948 are given
CC   in T28950, T28951 and T32143.  These genes can be expressed in
CC   usual heterologous hosts.  When inserted into E. coli they cause
CC   a significant change in metabolism.
SQ   Sequence   4495 BP;   1368 A;   745 C;   1052 G;   1330 T;

  Query Match          24.6%;  Score 15;  DB 24;  Length 4495;
  Best Local Similarity 100.0%;  Pred. No. 2.97e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1644 ctttattgttttatt 1658
          |||||||||||||||
Cp     39 CTTTATTGTTTTATT 25


RESULT   40
ID   T78867 standard; DNA; 6152 BP.
AC   T78867;
DT   08-OCT-1997  (first entry)
DE   P. falciparum liver stage antigen-3 genomic sequence.
KW   Plasmodium falciparum; pre-erythrocyte; liver stage antigen; serum;
KW   prophylaxis; Thai strain; gene organisation; exon; intron; hydrophobic;
KW   glycosyl-phosphatidylinositol membrane anchoring sequence; antibody;
KW   vaccine; immunotherapy; malaria; ds.
OS   Plasmodium falciparum.
FH   Key             Location/Qualifiers
FT   cds             77..5605
FT                   /*tag=  a
FT                   /product= liver stage antigen-3
FT   exon            77..274
FT                   /*tag=  b
FT                   /number= 1
FT   intron          275..442
FT                   /*tag=  c
FT                   /number= 1
FT   exon            443..5602
FT                   /*tag=  d
FT                   /number= 2
PN   WO9641877-A2.
PD   27-DEC-1996.
PF   12-JUN-1996; F00894.
PR   13-JUN-1995; FR-007007.
PA   (INSP ) INST PASTEUR.
PI   Daubersies P,  Druilhe P;
DR   WPI; 97-065464/06.
PT   Plasmodium falciparum poly:peptide(s) and related nucleic acids -
```

```
PT   derived from the liver stage antigen-3, useful for malaria vaccine
PT   prodn. and diagnosis
PS   Claim 20; Fig 1A-E; 69pp; French.
CC   This sequence corresponds to the genomic sequence encoding a Plasmodium
CC   falciparum strain K1 pre-erythrocytic liver stage antigen-3 (LSA-3)
CC   protein (W24790).  The gene sequence was isolated by screening a
CC   P. falciparum strain T9/96 library with the serum from a missionary
CC   treated by prophylaxis (for strain T6/96 see FR9101286).  Of 20 clones
CC   isolated, clone 729S was used to screen a library generated from Thai
CC   strain K1.  One clone contained a 6.85 kb insert including the sequence
CC   presented here.  The gene organisation comprises a first exon, a short
CC   168 bp intron and a 5 kb second exon containing a 1.8 kb region encoding
CC   7 blocks of 4 amino acid repeats and a 3' hydrophobic region
CC   corresponding to a glycosyl-phosphatidylinositol membrane anchoring
CC   sequence.  The invention relates to new polypeptides of at least 10
CC   amino acids derived from the LSA-3 polypeptide with the exception of
CC   peptides W24791-4. The LSA-3 peptides can be used to raise antibodies
CC   and as vaccines for immunotherapy of malaria.
SQ   Sequence   6152 BP;    2725 A;    463 C;    1215 G;    1749 T;

  Query Match           24.6%;  Score 15;  DB 32;  Length 6152;
  Best Local Similarity 100.0%;  Pred. No. 2.97e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     12 tttattgttttattt 26
          |||||||||||||||
Cp     38 TTTATTGTTTTATTT 24


RESULT   41
ID   V05856 standard; DNA; 7451 BP.
AC   V05856;
DT   01-JUN-1998  (first entry)
DE   Complete sequence of CpG island 3 at locus Xq12-q13.
KW   Human; anhidrotic ectodermal dysplasia; X chromosome; genetic linkage;
KW   translocation; CpG island; foetal development; hair; sweat gland;
KW   tooth; primer; PCR; amplification; probe; hybridisation; ss.
OS   Homo sapiens.
PN   US5700926-A.
PD   23-DEC-1997.
PF   22-JUL-1996; 684672.
PR   22-JUL-1996; US-684672.
PR   27-APR-1993; US-052997.
PA   (UNIW ) UNIV WASHINGTON.
PI   De La Chapelle A, Kere J, Schlessinger D, Srivastava AK;
DR   WPI; 98-062436/06.
PT   Human anhidrotic ectodermal dysplasia gene - useful for research
PT   into hair growth
PS   Disclosure; Fig 3; 37pp; English.
CC   This sequence represents the complete CpG island 3 region found at locus
CC   Xq12-q13 in humans.  The sequence has been shown to contain the sequence
```



```
CC   encoding the human anhidrotic ectodermal dysplasia (EDA) gene (V05851).
CC   Deficiencies in the EDA gene are observed by translocations with a
CC   breakpoint in the CpG island 3.  The EDA gene can be used to study the
CC   dynamics of EDA gene expression during foetal development, and processes
CC   affecting normal hair growth in adults.  The EDA gene can also be used to
CC   study hair, sweat gland and tooth formation and growth, and ectodermal
CC   dysplasias.
SQ   Sequence   7451 BP;    1964 A;    1653 C;    1758 G;    2075 T;

  Query Match           24.6%;  Score 15;  DB 40;  Length 7451;
  Best Local Similarity 100.0%;  Pred. No. 2.97e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    453 caaaataaaacaata 467
          |||||||||||||||
Qy     22 CAAAATAAAACAATA 36


RESULT   42
ID   T03852 standard; DNA; 8878 BP.
AC   T03852;
DT   26-OCT-1996  (first entry)
DE   Human fibrinogen B-beta chain coding sequence.
KW   Human fibrinogen B-beta chain; transgenic animal milk; treatment;
KW   sheep beta-lactoglobulin signal peptide fusion protein;
KW   surgical adhesive; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   misc_rna        1..469
FT                   /*tag= a
FT   exon            470..583
FT                   /*tag= b
FT                   /note= "1"
FT   intron          584..3257
FT                   /*tag= c
FT                   /note= "1"
FT   exon            3258..3449
FT                   /*tag= d
FT                   /note= "2"
FT   intron          3450..3938
FT                   /*tag= e
FT                   /note= "2"
FT   exon            3939..4122
FT                   /*tag= f
FT                   /note= "3"
FT   intron          4123..5042
FT                   /*tag= g
FT                   /note= "3"
FT   exon            5043..5270
FT                   /*tag= h
FT                   /note= "4"
```



```
FT   intron          5271..5830
FT                   /*tag= i
FT                   /note= "4"
FT   exon            5831..5944
FT                   /*tag= j
FT                   /note= "5"
FT   intron          5945..6632
FT                   /*tag= k
FT                   /note= "5"
FT   exon            6633..6758
FT                   /*tag= l
FT                   /note= "6"
FT   intron          6759..6966
FT                   /*tag= m
FT                   /note= "6"
FT   exon            6967..7252
FT                   /*tag= n
FT                   /note= "7"
FT   intron          7253..7870
FT                   /*tag= o
FT                   /note= "7"
FT   exon            7871..8102
FT                   /*tag= p
FT                   /note= "8"
FT   3'utr           8103..8537
FT                   /*tag= q
FT   misc_rna        8538..8878
FT                   /*tag= r
FT   cds             470..8102
FT                   /*tag= s
FT                   /note= "exons 1-8"
PN   WO9523868-A1.
PD   08-SEP-1995.
PF   01-MAR-1995; U02648.
PR   03-MAR-1994; US-206176.
PA   (PHAR-) PHARM PROTEINS LTD.
PA   (ZYMO ) ZYMOGENETICS INC.
PI   Dalrymple MA,  Foster DC,  Garner I,  Prunkard DE;
DR   WPI; 95-320582/41.
DR   P-PSDB; RB2243.
PT   Production of fibrinogen in transgenic mammals - by introducing DNA
PT   segments into the germ line of a non-human mammal and collecting
PT   milk from female progeny.
PS   Claim 9; Page 47-55; 99pp; English.
CC   This sequence encodes the human fibrinogen B-beta chain.  Preferably
CC   nucleotides 470-8100, and more preferably nucleotides 512-8100, are
CC   operably linked to the signal peptide, preferably of the sheep beta-
CC   lactoglobulin gene (see T03855) and, together with the fibrinogen A-
CC   alpha chain sequence (see T03853) and the gamma chain sequence (see
CC   T03854), is introduced into a fertilised egg or the germ line of a
CC   non-human animal, preferably a sheep, pig, goat or cattle at a molar
```



```
CC   ratio of 0.5-1.0:0.5-1.0:0.5-1.0, respectively, for fusion protein
CC   gene expression in transgenic animal milk.  The recombinantly
CC   produced active fibrinogen is useful in human and veterinary
CC   medicine, e.g. in the formulation of surgical adhesives, which also
CC   consist of Factor-XIII, and as a coating surface for polymeric
CC   articles, such as synthetic vascular grafts.
SQ   Sequence   8878 BP;    2912 A;    1477 C;    1664 G;    2825 T;

  Query Match           24.6%;  Score 15;  DB 22;  Length 8878;
  Best Local Similarity 100.0%;  Pred. No. 2.97e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    6945 ttttattttgttttt 6959
          |||||||||||||||
Cp      31 TTTTATTTTGTTTTT 17


RESULT   43
ID   T62125 standard; DNA; 9295 BP.
AC   T62125;
DT   10-JUN-1997  (first entry)
DE   Arabidopsis thaliana plant morphogenesis regulatory protein DNA.
KW   Plant; morphogenesis; regulation; short; stem; alteration;
KW   inflorescence; extraneous; gene; expression; transformation;
KW   increase; control; form; length; ds.
OS   Arabidopsis thaliana.
FH   Key             Location/Qualifiers
FT   exon            1803..1881
FT                   /*tag= a
FT   intron          1882..2227
FT                   /*tag= b
FT   exon            2228..2366
FT                   /*tag= c
FT   intron          2367..2467
FT                   /*tag= d
FT   intron          2540..2643
FT                   /*tag= e
FT   exon            2468..2539
FT                   /*tag= f
FT   exon            2644..2715
FT                   /*tag= g
FT   intron          2716..2809
FT                   /*tag= h
FT   exon            2810..2878
FT                   /*tag= i
FT   intron          2879..2968
FT                   /*tag= j
FT   exon            2969..3040
FT                   /*tag= k
FT   intron          3041..3118
FT                   /*tag= l
```

```
FT   exon            3119..3190
FT                   /*tag=  m
FT   intron          3191..3266
FT                   /*tag=  n
FT   exon            3267..3338
FT                   /*tag=  o
FT   intron          3339..3421
FT                   /*tag=  p
FT   exon            3422..3493
FT                   /*tag=  q
FT   intron          3494..3586
FT                   /*tag=  r
FT   exon            3587..3655
FT                   /*tag=  s
FT   intron          3656..3740
FT                   /*tag=  t
FT   exon            3741..3812
FT                   /*tag=  u
FT   intron          3813..3888
FT                   /*tag=  v
FT   exon            3889..3960
FT                   /*tag=  w
FT   intron          3961..4048
FT                   /*tag=  x
FT   exon            4049..4120
FT                   /*tag=  y
FT   intron          4121..4209
FT                   /*tag=  z
FT   exon            4210..4281
FT                   /*tag=  aa
FT   intron          4282..4349
FT                   /*tag=  ab
FT   exon            4350..4421
FT                   /*tag=  ac
FT   intron          4422..4508
FT                   /*tag=  ad
FT   exon            4509..4580
FT                   /*tag=  ae
FT   intron          4581..4706
FT                   /*tag=  af
FT   exon            4707..4778
FT                   /*tag=  ag
FT   intron          4779..4860
FT                   /*tag=  ah
FT   exon            4861..4932
FT                   /*tag=  ai
FT   intron          4933..5018
FT                   /*tag=  aj
FT   exon            5019..5090
FT                   /*tag=  ak
FT   intron          5091..5176
```



```
FT                   /*tag=  al
FT   exon            5177..5248
FT                   /*tag=  am
FT   intron          5249..5412
FT                   /*tag=  an
FT   exon            5413..5481
FT                   /*tag=  ao
FT   intron          5482..5576
FT                   /*tag=  ap
FT   exon            5577..5648
FT                   /*tag=  aq
FT   intron          5649..5726
FT                   /*tag=  ar
FT   exon            5727..5800
FT                   /*tag=  as
FT   intron          5801..5882
FT                   /*tag=  at
FT   exon            5883..6011
FT                   /*tag=  au
FT   exon            6096..6443
FT                   /*tag=  av
FT   intron          6012..6095
FT                   /*tag=  aw
FT   intron          6444..6519
FT                   /*tag=  ax
FT   exon            6520..5412
FT                   /*tag=  ay
FT   intron          6891..6974
FT                   /*tag=  az
FT   exon            6975..7328
FT                   /*tag=  ba
PN   J09056382-A.
PD   04-MAR-1997.
PF   24-AUG-1995; 216187.
PR   24-AUG-1995; JP-216187.
PA   (MITS-) MITSUI GYOSAI SHOKUBUTSU BIO KENKYUSHO.
PA   (CHIK-) ZH CHIKYU KANKYO SANGYO GIJITSU KENKYU.
DR   WPI; 97-206629/19.
PT   DNA encoding plant morphogenesis regulatory protein - useful to
PT   yield plants with short stems or altered inflorescence
PS   Claim 6; Pages 12-15; 17pp; Japanese.
CC   The present sequence encodes an Arabidopsis thaliana plant
CC   morphogenesis regulatory protein (MRP), which can be used to yield
CC   a plant with, e.g. short stems or altered inflorescence. The MRP
CC   acts on a plant at a specific site for a specific period, and can
CC   therefore be used to regulate extraneous gene expression in a
CC   plant. The MRP's cDNA or genomic DNA can be used to transform a
CC   plant to increase its MRP expression, and therefore control the
CC   form (particularly stem length) of the plant.
SQ   Sequence   9295 BP;    2873 A;    1581 C;    1602 G;    3239 T;
```



```
  Query Match           24.6%;  Score 15;  DB 29;  Length 9295;
  Best Local Similarity 100.0%;  Pred. No. 2.97e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2114 gttttattttgtttt 2128
          |||||||||||||||
Cp     32 GTTTTATTTTGTTTT 18


RESULT   44
ID   Q28505 standard; DNA; 10798 BP.
AC   Q28505;
DT   11-JAN-1993  (first entry)
DE   Tomato vacuolar invertase gene (genomic).
KW   RNA polymerase II; promoter; cauliflower mosaic virus; CaMV; ss.
OS   Lycopersicon esculentum.
FH   Key             Location/Qualifiers
FT   promoter        1..3519
FT                   /*tag=  a
FT                   /note= "claim 35, page 80"
FT   promoter        2772..3519
FT                   /*tag=  b
FT                   /note= "claim 36, page 80"
FT   promoter        2440..3519
FT                   /*tag=  c
FT                   /note= "claim 37, page 80"
FT   tata_signal     3471..3474
FT                   /*tag=  d
FT   repeat_unit     1846..1853
FT                   /*tag=  e
FT                   /note= "sequence matches known plant nuclear
FT                   binding sites"
FT   repeat_unit     2012..2019
FT                   /*tag=  f
FT                   /note= "sequence matches known plant nuclear
FT                   binding sites"
FT   repeat_unit     2178..2185
FT                   /*tag=  g
FT                   /note= "sequence matches known plant nuclear
FT                   binding sites"
FT   repeat_unit     2344..2351
FT                   /*tag=  h
FT                   /note= "sequence matches known plant nuclear
FT                   binding sites"
FT   repeat_unit     2510..2517
FT                   /*tag=  i
FT                   /note= "sequence matches known plant nuclear
FT                   binding sites"
FT   repeat_unit     2676..2683
FT                   /*tag=  j
FT                   /note= "sequence matches known plant nuclear
```



```
FT                   binding sites"
FT   repeat_unit     2842..2849
FT                   /*tag=  k
FT                   /note= "sequence matches known plant nuclear
FT                   binding sites"
FT   misc_rna        3502
FT                   /*tag=  l
FT                   /label= transcription_start_site
FT   precursor_rna   3520..7445
FT                   /*tag=  m
PN   WO9214831-A.
PD   03-SEP-1992.
PF   21-FEB-1992; U01385.
PR   22-FEB-1991; US-660344.
PR   04-OCT-1991; US-771331.
PA   (SALK ) SALK INST BIOTECHNOLOGY IND ASSOC.
PI   Butler WO, Dickinson CD, Elliott KJ, Fitzmaurice LC;
PI   Konno Y, Mirkov TE;
DR   WPI; 92-316183/38.
PT   Tomato vacuolar invertase gene - useful for producing transgenic
PT   plants with higher soluble solid content in fruit, giving greater
PT   commercial value
PS   Disclosure; Page 64-68; 87pp; English.
CC   In order to increase the soluble solids content of tomato fruit
CC   produced by a tomato plant, a DNA construct contg. DNA encoding an
CC   invertase is introduced.  The DNA encoding the invertase is pref.
CC   operatively linked to a promoter recognised by the plant RNA
CC   polymerase II.  The promoter may be a constitutive promoter,
CC   such as the cauliflower mosaic virus (CaMV) 35S promoter, or a
CC   developmentally regulated promoter that confers fruit specificity
CC   and appropriate temporal control on the expression of the DNA
CC   encoding invertase, e.g. a native Lycopersicon invertase promoter
CC   or the histidine decarboxylase (HDC) promoter region.
CC   If the DNA encodes an invertase that is not a vacuolar invertase,
CC   DNA encoding invertase is operatively linked to DNA that encodes
CC   vacuolar targeting sequences, and, if necessary, DNA encoding
CC   signal sequences.
SQ   Sequence   10798 BP;    3619 A;    1709 C;    1771 G;    3699 T;

  Query Match           24.6%;  Score 15;  DB 4;  Length 10798;
  Best Local Similarity 100.0%;  Pred. No. 2.97e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   9215 taaaacaataaagac 9229
          |||||||||||||||
Qy     27 TAAAACAATAAAGAC 41


RESULT   45
ID   Q28507 standard; DNA; 10965 BP.
AC   Q28507;
```

```
DT   11-JAN-1993  (first entry)
DE   Tomato vacuolar invertase gene (genomic).
KW   RNA polymerase II; promoter; histidine decarboxylase; HDC;
KW   cauliflower mosaic virus; CaMV; ss.
OS   Lycopersicon pimpinellifolium.
FH   Key             Location/Qualifiers
FT   promoter        1..3679
FT                   /*tag= a
FT                   /note= "claim 34, page 80"
FT   tata_signal     3637..3640
FT                   /*tag= b
FT   repeat_unit     1846..1853
FT                   /*tag= c
FT                   /note= "sequence matches known plant nuclear
FT                   binding sites"
FT   repeat_unit     2178..2185
FT                   /*tag= d
FT                   /note= "sequence matches known plant nuclear
FT                   binding sites"
FT   repeat_unit     2344..2351
FT                   /*tag= e
FT                   /note= "sequence matches known plant nuclear
FT                   binding sites"
FT   repeat_unit     2510..2517
FT                   /*tag= f
FT                   /note= "sequence matches known plant nuclear
FT                   binding sites"
FT   repeat_unit     2676..2683
FT                   /*tag= g
FT                   /note= "sequence matches known plant nuclear
FT                   binding sites"
FT   repeat_unit     2842..2849
FT                   /*tag= h
FT                   /note= "sequence matches known plant nuclear
FT                   binding sites"
FT   repeat_unit     3008..3015
FT                   /*tag= i
FT                   /note= "sequence matches known plant nuclear
FT                   binding sites"
FT   misc_rna        3668
FT                   /*tag= j
FT                   /label= transcription_start_site
FT   precursor_rna   3686..7612
FT                   /*tag= k
PN   WO9214831-A.
PD   03-SEP-1992.
PF   21-FEB-1992; U01385.
PR   22-FEB-1991; US-660344.
PR   04-OCT-1991; US-771331.
PA   (SALK ) SALK INST BIOTECHNOLOGY IND ASSOC.
PI   Butler WO, Dickinson CD, Elliott KJ, Fitzmaurice LC;
```



```
PI   Konno Y, Mirkov TE;
DR   WPI; 92-316183/38.
PT   Tomato vacuolar invertase gene - useful for producing transgenic
PT   plants with higher soluble solid content in fruit, giving greater
PT   commercial value
PS   Disclosure; Page 71-75; 87pp; English.
CC   In order to increase the soluble solids content of tomato fruit
CC   produced by a tomato plant, a DNA constructcontg. DNA encoding an
CC   invertase is introduced.  The DNA encoding the invertase is pref.
CC   operatively linked to a promoter recognised by the plant RNA
CC   polymerase II.  The promoter may be a constitutive promoter,
CC   such as the cauliflower mosaic virus (CaMV) 35S promoter, or a
CC   developmentally regulated promoter that confers fruit specificity
CC   and appropriate temporal control on the expression of the DNA
CC   encoding invertase, e.g. a native Lycopersicon invertase promoter
CC   or the histidine decarboxylase (HDC) promoter region.
CC   If the DNA encodes an invertase that is not a vacuolar invertase,
CC   DNA encoding invertase is operatively linked to DNA that encodes
CC   vacuolar targeting sequences and, if necessary, DNA encoding signal
CC   sequences.
SQ   Sequence   10965 BP;    3668 A;    1743 C;    1784 G;    3770 T;

  Query Match          24.6%;  Score 15;  DB 4;  Length 10965;
  Best Local Similarity 100.0%;  Pred. No. 2.97e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   9381 taaaacaataaagac 9395
           |||||||||||||||
Qy     27 TAAAACAATAAAGAC 41


RESULT   46
ID   V02032 standard; DNA; 40352 BP.
AC   V02032;
DT   12-JUN-1998  (first entry)
DE   MAGE-B cluster DNA sequence.
KW   MAGE-B cluster DNA; testicular seminoma; cancer; screening; leukaemia;
KW   carcinoma; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   prim_transcript 3266..7979
FT                   /*tag= a
FT                   /label= B2 gene
FT   exon            3266..3364
FT                   /*tag= b
FT                   /note= "exon 1 of B2 gene"
FT   intron          3365..6277
FT                   /*tag= c
FT                   /note= "intron 1 of B2 gene"
FT   exon            6278..7979
FT                   /*tag= d
```



```
FT                   /note= "exon 2 of B2 gene"
FT   CDS             6283..7224
FT                   /*tag= e
FT                   /note= "coding sequence for B2 gene"
FT   polyA_signal    7961..7966
FT                   /*tag= f
FT                   /note= "poly-A signal for B2 gene"
FT   exon            23546..25194
FT                   /*tag= g
FT                   /note= "exon 1 of B3 gene"
FT   CDS             23607..24647
FT                   /*tag= h
FT                   /note= "coding sequence for B3 gene"
FT   polyA_signal    25152..25157
FT                   /*tag= i
FT                   /note= "poly-A signal for B3 gene"
FT   prim_transcript 29748..31827
FT                   /*tag= j
FT                   /label= B4 gene
FT   CDS             29808..30848
FT                   /*tag= k
FT                   /note= "coding sequence for B4 gene"
FT   polyA_signal    31822..31827
FT                   /*tag= l
FT                   /note= "poly-A signal for B4 gene"
FT   prim_transcript 31403..39691
FT                   /*tag= m
FT                   /label= B1 gene
FT   exon            31403..31474
FT                   /*tag= n
FT                   /note= "exon 1 of B1 gene"
FT   intron          31475..33957
FT                   /*tag= o
FT                   /note= "intron 1 of B1 gene"
FT   exon            33958..34062
FT                   /*tag= p
FT                   /note= "exon 2 of B1 gene"
FT   intron          34063..35056
FT                   /*tag= q
FT                   /note= "intron 2 of B1 gene"
FT   exon            35057..35139
FT                   /*tag= r
FT                   /note= "exon 3 of B1 gene"
FT   intron          35140..38087
FT                   /*tag= s
FT                   /note= "intron 3 of B1 gene"
FT   exon            38088..39691
FT                   /*tag= t
FT                   /note= "exon 4 of B1 gene"
FT   CDS             38148..39191
FT                   /*tag= u
```



```
FT                   /note= "coding sequence for B1 gene"
FT   polyA_signal    39674..39679
FT                   /*tag= v
FT                   /note= "poly-A signal for B1 gene"
PN   WO9746710-A1.
PD   11-DEC-1997.
PF   05-JUN-1997; U09774.
PR   25-APR-1997; US-846111.
PR   05-JUN-1996; US-658578.
PA   (LUDW-) LUDWIG INST CANCER RES.
PI   Boon T, Brasseur F, Lurquin C;
DR   WPI; 98-042216/04.
PT   Screening for the possibility of testicular seminoma and other
PT   cancers - using MAGE-Xp gene primers
PS   Claim 7; Pages 22-34; 40pp; English.
CC   This is the MAGE-B cluster DNA sequence. The invention provides a novel
CC   method for screening for the possibility of a testicular seminoma,
CC   non-small cell lung carcinoma, melanoma, breast cancer, sarcoma, or
CC   leukaemia in a sample. The method comprises contacting the sample with
CC   at least one nucleic acid molecule which hybridises to mRNA corresponding
CC   to an MAGE-Xp gene, and determining hybridisation as a determination of
CC   possible presence of testicular seminoma, non-small cell lung carcinoma
CC   melanoma, breast cancer, sarcoma, or leukaemia in the sample. By assaying
CC   for the MAGE-B1 gene, or MAGE-B2 gene, both contained within this 40352
CC   base pair MAGE-B cluster DNA sequence, the presence of auch cancers in a
CC   sample can be determined. The genomic DNA that encodes the MAGE-B2 gene
CC   consists of nucleotides 3266-7979 of this MAGE-B cluster DNA sequence.
CC   The nucleic acid molecule that encodes a MAGE-B1 gene, consists of, in
CC   5' to 3' order, nucleotides 31403-31474, 33958-34062, 35057-35139, and
CC   38088-39691, nucleotides 31403-31474, 33958-34062, and 38088-39691,
CC   nucleotides 35057-35139 and 38088-39691, and nucleotides 33958-34062 and
CC   38088-39691 of this sequence. The primers specific for the MAGE-B
CC   gene can be used in a kit to amplify a MAGE-B gene.
SQ   Sequence   40352 BP;    11599 A;    8654 C;    9772 G;    10327 T;

  Query Match          24.6%;  Score 15;  DB 40;  Length 40352;
  Best Local Similarity 100.0%;  Pred. No. 2.97e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  31058 attttgtttttgaaa 31072
          |||||||||||||||
Cp     27 ATTTTGTTTTTGAAA 13


RESULT   47
ID   T13635 standard; DNA; 133894 BP.
AC   T13635;
DT   03-SEP-1996  (first entry)
DE   AcNPV genomic DNA clone 6.
KW   Autographa californica nuclear polyhedrosis virus clone 6;
KW   disruption; non-essential gene; heterologous protein production;
```

```
KW   expression vector; baculovirus; ss.
OS   Autographa californica nuclear polyhedrosis virus clone 6.
PN   WO9601320-A2.
PD   18-JAN-1996.
PF   30-JUN-1995; IB0578.
PR   04-JUL-1994; GB-013420.
PA   (NATU-) NATURAL ENVIRONMENT RES COUNCIL.
PI   Ayres M, Bishop D, Possee R;
DR   WPI; 96-087670/09.
DR   GENBANK; L22858.
PT   Autographa californica nuclear polyhedrosis virus complete genome
PT   sequence - useful in the prodn. of vectors for enhanced
PT   heterologous protein expression, such as interleukin(s),
PT   interferon(s) and neurotoxin(s)
PS   Disclosure; Page 90-186; 122pp; English.
CC   The complete nucleotide sequence of the genome of clone 6 of the
CC   baculovirus Autographa californica nuclear polyhedrosis virus (AcNPV)
CC   has been determined. The sequence is taken from the Genbank record
CC   L22858. The patent specification claims a polynucleotide selected from
CC   open reading frames (ORFs 13, 20, 22-26, 28-30, 32, 38, 41-46, 50-60,
CC   62-63, 66, 68-79, 81-87, 91-92, 96-98, 101-103, 106-126, 129-130,
CC   140-146, 148-150, 152 and 154 from a total of 154 ORFs identified by
CC   the patentees. See T13636-731. Expression vectors contg. the complete
CC   genomic sequence of AcNPV, with the exception that at least one non-
CC   essential ORF is disrupted or replaced are useful for the synthesis of
CC   heterologous proteins.
SQ   Sequence   133894 BP;    39195 A;    27151 C;    27347 G;    40201 T;

  Query Match           24.6%;  Score 15;  DB 17;  Length 133894;
  Best Local Similarity 100.0%;  Pred. No. 2.97e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  15818 ttattgttttatttt 15832
          |||||||||||||||
Cp     37 TTATTGTTTTATTTT 23


RESULT   48
ID   T58840 standard; DNA; 580073 BP.
AC   T58840;
DT   27-MAR-1997  (first entry)
DE   Mycoplasma genitalium genome.
KW   M. genitalium; DNAA; DNA gyrase; origin of replication;
KW   megabase shotgun sequencing method; open reading frame; ORF; ss.
OS   Mycoplasma genitalium.
FH   Key             Location/Qualifiers
FT   cds             8552..9184
FT                   /*tag= a
FT                   /label= MG006
FT                   /note= "Previously identified as MORF-20076, the
FT                   encoded protein shows 27.59 percentage
```



```
FT                   identity to thymidylate kinase (CDC8)
FT                   from Saccharomyces cerevisiae"
FT   cds             11252..12040
FT                   /*tag= b
FT                   /label= MG009
FT                   /note= "Previously identified as MORF-20078, the
FT                   encoded protein shows 35.43 percentage
FT                   identity to the Bacillus subtilis hypothetical
FT                   protein covered in accession number
FT                   GB:D26185_102"
FT   cds             12069..12725
FT                   /*tag= c
FT                   /label= MG010
FT                   /note= "Previously identified as MORF-20079, the
FT                   encoded protein shows 25.73 percentage
FT                   identity to DNA primase (dnaE) from
FT                   Clostridium acetobutylicum"
FT   cds             complement (13570..14247)
FT                   /*tag= d
FT                   /label= MG012
FT                   /note= "Previously identified as MORF-20080, the
FT                   encoded protein shows 31.50 percentage
FT                   identity to the ribosomal protein S6
FT                   modification protein (rimK) from Escherichia
FT                   coli"
FT   cds             complement (14396..15217)
FT                   /*tag= e
FT                   /label= MG013
FT                   /note= "Previously identified as MORF-19823, MORF-20080
FT                   and MORF-20081, the encoded protein shows 33.04
FT                   percentage identity to 5,10-methylene-tetra-
FT                   hydrofolate dehydrogenase (folD) from E. coli"
FT   cds             17474..19243
FT                   /*tag= f
FT                   /label= MG015
FT                   /note= "Previously identified as MORF-20084, the
FT                   encoded protein shows 32.23 percentage
FT                   identity to transport ATP-binding protein
FT                   (msbA) from E. coli"
FT   cds             26478..27344
FT                   /*tag= g
FT                   /label= MG023
FT                   /note= "Previously identified as MORF-20092, the
FT                   encoded protein shows 45.96 percentage
FT                   identity to fructose-bisphosphate aldolase
FT                   (tsr) from B. subtilis"
FT   cds             27345..28448
FT                   /*tag= h
FT                   /label= MG024
FT                   /note= "Previously identified as MORF-19826 and
FT                   MORF-20093, the encoded protein shows 46.84
```



```
FT                   percentage identity to GTP-binding protein
FT                   from E. coli"
FT   cds             36987..38978
FT                   /*tag= i
FT                   /label= MG032
FT                   /note= "Previously identified as MORF-20099, the
FT                   encoded protein shows 26.62 percentage
FT                   identity to ATP-dependent nuclease (addA)
FT                   from B. subtilis"
FT   cds             39242..39904
FT                   /*tag= j
FT                   /label= MG033
FT                   /note= "Previously identified as MORF-20100, the
FT                   encoded protein shows 35.90 percentage
FT                   identity to glycerol uptake facilitator
FT                   (glpF) from B. subtilis"
FT   cds             complement (39873..40514)
FT                   /*tag= k
FT                   /label= MG034
FT                   /note= "Previously identified as MORF-20101, the
FT                   encoded protein shows 48.13 percentage
FT                   identity to thymidylate kinase (tdk)
FT                   from B. subtilis"
FT   cds             40543..41787
FT                   /*tag= l
FT                   /label= MG035
FT                   /note= "Previously identified as MORF-20102, the
FT                   encoded protein shows 30.71 percentage
FT                   identity to histidyl-tRNA synthetase (hisS)
FT                   from Mycobacterium leprae"
FT   cds             complement (44751..46277)
FT                   /*tag= m
FT                   /label= MG038
FT                   /note= "Previously identified as MORF-20105, the
FT                   encoded protein shows 46.83 percentage
FT                   identity to glycerol kinase (glpK)
FT                   from E. coli"
FT   cds             complement (46268..47422)
FT                   /*tag= n
FT                   /label= MG039
FT                   /note= "Previously identified as MORF-19831 and
FT                   MORF-20106, the encoded protein shows 43.20
FT                   percentage identity to glycerol-3-phosphate
FT                   dehydrogenase (GUT2) from S. cerevisiae"
FT   cds             49377..49643
FT                   /*tag= o
FT                   /label= MG041
FT                   /note= "The encoded protein shows 48.86 percentage
FT                   identity to phosphohistidinoprotein-hexose
FT                   phosphotransferase (ptsH) from Mycoplasma
FT                   capricolum"
```



```
FT   cds             50060..51520
FT                   /*tag= p
FT                   /label= MG042
FT                   /note= "Previously identified as MORF-19832 and
FT                   MORF-20108, the encoded protein shows 41.92
FT                   percentage identity to spermidine/
FT                   putrescine transport ATP-binding protein
FT                   (potA) from E. coli"
FT   cds             51525..52382
FT                   /*tag= q
FT                   /label= MG043
FT                   /note= "Previously identified as MORF-20110, the
FT                   encoded protein shows 26.51 percentage
FT                   identity to spermidine/putrescine transport
FT                   system permease protein (potB) from E. coli"
FT   cds             52366..53220
FT                   /*tag= r
FT                   /label= MG044
FT                   /note= "Previously identified as MORF-20111, the
FT                   encoded protein shows 29.45 percentage
FT                   identity to spermidine/putrescine transport
FT                   system permease protein C (potC) from E. coli"
FT   cds             54658..55605
FT                   /*tag= s
FT                   /label= MG046
FT                   /note= "Previously identified as MORF-20112, the
FT                   encoded protein shows 36.60 percentage
FT                   identity to sialoglycoprotease (gcp)
FT                   from Pasteurella haemolytica"
FT   cds             complement (56970..58310)
FT                   /*tag= t
FT                   /label= MG048
FT                   /note= "Previously identified as MORF-19834,
FT                   MORF-20114 and MORF-20115, the encoded protein
FT                   shows 43.02 percentage identity to signal
FT                   recognition particle protein (ffh) from B.
FT                   subtilis"
FT   cds             58117..59079
FT                   /*tag= u
FT                   /label= MG049
FT                   /note= "Previously identified as MORF-20114 and
FT                   MORF-20115, the encoded protein shows 44.78
FT                   percentage identity to purine-nucleoside
FT                   phosphorylase (deoD) from E. coli"
FT   cds             59083..59754
FT                   /*tag= v
FT                   /label= MG050
FT                   /note= "Previously identified as MORF-20117, the
FT                   encoded protein shows 83.03 percentage
FT                   identity to deoxyribose-phosphate aldolase
FT                   (deoC) from Mycoplasma pneumoniae"
```

```
FT   cds             complement (64898..65731)
FT                   /*tag= w
FT                   /label= MG056
FT                   /note= "Previously identified as MORF-20122, the
FT                   encoded protein shows 30.25 percent
FT                   identity to the protein disclosed in
FT                   GB:D26185_99 from B. subtilis"
FT   cds             complement (65713..66249)
FT                   /*tag= x
FT                   /label= MG057
FT                   /note= "Previously identified as MORF-20123, the
FT                   encoded protein shows 38.90 percentage
FT                   identity to the protein disclosed in
FT                   GB:D26185_104 from B. subtilis"
FT   cds             81047..82597
FT                   /*tag= y
FT                   /label= MG067
FT                   /note= "Previously identified as MORF-19845, the
FT                   encoded protein shows 28.84 percentage
FT                   identity to glutamic acid specific protease
FT                   (SPase) from Staphylococcus aureus"
FT   cds             91065..91919
FT                   /*tag= z
FT                   /label= MG070
FT                   /note= "Previously identified as MORF-20136, the
FT                   encoded protein shows 34.8 percentage
FT                   identity to ribosomal protein S2 (rpS2)
FT                   from Spirulina plantensis"
FT   cds             103104..104324
FT                   /*tag= aa
FT                   /label= MG077
FT                   /note= "Previously identified as MORF-20140, the
FT                   encoded protein shows 28.05 percentage
FT                   identity to oligopeptide transport system
...
Note: remainder of annotations omitted.

  Query Match          24.6%;  Score 15;  DB 27;  Length 580073;
  Best Local Similarity 100.0%;  Pred. No. 2.97e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db 273732 aaaaacaaaataaaa 273746
          ||||||||||||||||
Qy     17 AAAAACAAAATAAAA 31


RESULT   49
ID   T58840 standard; DNA; 580073 BP.
AC   T58840;
DT   27-MAR-1997  (first entry)
DE   Mycoplasma genitalium genome.
```



```
KW   M. genitalium; DNAA; DNA gyrase; origin of replication;
KW   megabase shotgun sequencing method; open reading frame; ORF; ss.
OS   Mycoplasma genitalium.
FH   Key             Location/Qualifiers
FT   cds             8552..9184
FT                   /*tag= a
FT                   /label= MG006
FT                   /note= "Previously identified as MORF-20076, the
FT                   encoded protein shows 27.59 percentage
FT                   identity to thymidylate kinase (CDC8)
FT                   from Saccharomyces cerevisiae"
FT   cds             11252..12040
FT                   /*tag= b
FT                   /label= MG009
FT                   /note= "Previously identified as MORF-20078, the
FT                   encoded protein shows 35.43 percentage
FT                   identity to the Bacillus subtilis hypothetical
FT                   protein covered in accession number
FT                   GB:D26185_102"
FT   cds             12069..12725
FT                   /*tag= c
FT                   /label= MG010
FT                   /note= "Previously identified as MORF-20079, the
FT                   encoded protein shows 25.73 percentage
FT                   identity to DNA primase (dnaE) from
FT                   Clostridium acetobutylicum"
FT   cds             complement (13570..14247)
FT                   /*tag= d
FT                   /label= MG012
FT                   /note= "Previously identified as MORF-20080, the
FT                   encoded protein shows 31.50 percentage
FT                   identity to the ribosomal protein S6
FT                   modification protein (rimK) from Escherichia
FT                   coli"
FT   cds             complement (14396..15217)
FT                   /*tag= e
FT                   /label= MG013
FT                   /note= "Previously identified as MORF-19823, MORF-20080
FT                   and MORF-20081, the encoded protein shows 33.04
FT                   percentage identity to 5,10-methylene-tetra-
FT                   hydrofolate dehydrogenase (folD) from E. coli"
FT   cds             17474..19243
FT                   /*tag= f
FT                   /label= MG015
FT                   /note= "Previously identified as MORF-20084, the
FT                   encoded protein shows 32.23 percentage
FT                   identity to transport ATP-binding protein
FT                   (msbA) from E. coli"
FT   cds             26478..27344
FT                   /*tag= g
FT                   /label= MG023
```



```
FT                   /note= "Previously identified as MORF-20092, the
FT                   encoded protein shows 45.96 percentage
FT                   identity to fructose-bisphosphate aldolase
FT                   (tsr) from B. subtilis"
FT   cds             27345..28448
FT                   /*tag= h
FT                   /label= MG024
FT                   /note= "Previously identified as MORF-19826 and
FT                   MORF-20093, the encoded protein shows 46.84
FT                   percentage identity to GTP-binding protein
FT                   from E. coli"
FT   cds             36987..38978
FT                   /*tag= i
FT                   /label= MG032
FT                   /note= "Previously identified as MORF-20099, the
FT                   encoded protein shows 26.82 percentage
FT                   identity to ATP-dependent nuclease (addA)
FT                   from B. subtilis"
FT   cds             39242..39904
FT                   /*tag= j
FT                   /label= MG033
FT                   /note= "Previously identified as MORF-20100, the
FT                   encoded protein shows 35.90 percentage
FT                   identity to glycerol uptake facilitator
FT                   (glpF) from B. subtilis"
FT   cds             complement (39873..40514)
FT                   /*tag= k
FT                   /label= MG034
FT                   /note= "Previously identified as MORF-20101, the
FT                   encoded protein shows 48.13 percentage
FT                   identity to thymidylate kinase (tdk)
FT                   from B. subtilis"
FT   cds             40543..41787
FT                   /*tag= l
FT                   /label= MG035
FT                   /note= "Previously identified as MORF-20102, the
FT                   encoded protein shows 30.71 percentage
FT                   identity to histidyl-tRNA synthetase (hisS)
FT                   from Mycobacterium leprae"
FT   cds             complement (44751..46277)
FT                   /*tag= m
FT                   /label= MG038
FT                   /note= "Previously identified as MORF-20105, the
FT                   encoded protein shows 46.83 percentage
FT                   identity to glycerol kinase (glpK)
FT                   from E. coli"
FT   cds             complement (46268..47422)
FT                   /*tag= n
FT                   /label= MG039
FT                   /note= "Previously identified as MORF-19831 and
FT                   MORF-20106, the encoded protein shows 43.20
```



```
FT                   percentage identity to glycerol-3-phosphate
FT                   dehydrogenase (GUT2) from S. cerevisiae"
FT   cds             49377..49643
FT                   /*tag= o
FT                   /label= MG041
FT                   /note= "The encoded protein shows 48.86 percentage
FT                   identity to phosphohistidinoprotein-hexose
FT                   phosphotransferase (ptsH) from Mycoplasma
FT                   capricolum"
FT   cds             50060..51520
FT                   /*tag= p
FT                   /label= MG042
FT                   /note= "Previously identified as MORF-19832 and
FT                   MORF-20108, the encoded protein shows 41.92
FT                   percentage identity to spermidine/
FT                   putrescine transport ATP-binding protein
FT                   (potA) from E. coli"
FT   cds             51525..52382
FT                   /*tag= q
FT                   /label= MG043
FT                   /note= "Previously identified as MORF-20110, the
FT                   encoded protein shows 26.51 percentage
FT                   identity to spermidine/putrescine transport
FT                   system permease protein (potB) from E. coli"
FT   cds             52366..53220
FT                   /*tag= r
FT                   /label= MG044
FT                   /note= "Previously identified as MORF-20111, the
FT                   encoded protein shows 29.45 percentage
FT                   identity to spermidine/putrescine transport
FT                   system permease protein C (potC) from E. coli"
FT   cds             54658..55605
FT                   /*tag= s
FT                   /label= MG046
FT                   /note= "Previously identified as MORF-20112, the
FT                   encoded protein shows 36.60 percentage
FT                   identity to sialoglycoprotease (gcp)
FT                   from Pasteurella haemolytica"
FT   cds             complement (56970..58310)
FT                   /*tag= t
FT                   /label= MG048
FT                   /note= "Previously identified as MORF-19834,
FT                   MORF-20114 and MORF-20115, the encoded protein
FT                   shows 43.02 percentage identity to signal
FT                   recognition particle protein (ffh) from B.
FT                   subtilis"
FT   cds             58117..59079
FT                   /*tag= u
FT                   /label= MG049
FT                   /note= "Previously identified as MORF-20114 and
FT                   MORF-20115, the encoded protein shows 44.78
```

```
FT                   percentage identity to purine-nucleoside
FT                   phosphorylase (deoD) from E. coli"
FT   cds             59083..59754
FT                   /*tag= v
FT                   /label= MG050
FT                   /note= "Previously identified as MORF-20117, the
FT                   encoded protein shows 83.03 percentage
FT                   identity to deoxyribose-phosphate aldolase
FT                   (deoC) from Mycoplasma pneumoniae"
FT   cds             complement (64898..65731)
FT                   /*tag= w
FT                   /label= MG056
FT                   /note= "Previously identified as MORF-20122, the
FT                   encoded protein shows 30.25 percent
FT                   identity to the protein disclosed in
FT                   GB:D26185_99 from B. subtilis"
FT   cds             complement (65713..66249)
FT                   /*tag= x
FT                   /label= MG057
FT                   /note= "Previously identified as MORF-20123, the
FT                   encoded protein shows 38.90 percentage
FT                   identity to the protein disclosed in
FT                   GB:D26185_104 from B. subtilis"
FT   cds             81047..82597
FT                   /*tag= y
FT                   /label= MG067
FT                   /note= "Previously identified as MORF-19845, the
FT                   encoded protein shows 28.84 percentage
FT                   identity to glutamic acid specific protease
FT                   (SPase) from Staphylococcus aureus"
FT   cds             91065..91919
FT                   /*tag= z
FT                   /label= MG070
FT                   /note= "Previously identified as MORF-20136, the
FT                   encoded protein shows 34.8 percentage
FT                   identity to ribosomal protein S2 (rpS2)
FT                   from Spirulina plantensis"
FT   cds             103104..104324
FT                   /*tag= aa
FT                   /label= MG077
FT                   /note= "Previously identified as MORF-20140, the
FT                   encoded protein shows 28.05 percentage
FT                   identity to oligopeptide transport system
...
Note: remainder of annotations omitted.

  Query Match           24.6%;  Score 15;  DB 27;  Length 580073;
  Best Local Similarity 100.0%;  Pred. No. 2.97e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db 238721 ttgtttttgaaaaga 238735
```



```
                  |||||||||||||||
Cp     24 TTGTTTTTGAAAAGA 10


RESULT   50
ID   Q41022 standard; DNA; 87 BP.
AC   Q41022;
DT   29-JUL-1993  (first entry)
DE   P11 promoter mutant P11m.
KW   Fowl pox virus; FPV; P11; promoter; vaccina virus; mutation; apoAI;
KW   multiple cloning site; plasmid; thymidine kinase; tk; Marck's disease;
KW   intergenic region; post-translational modification; factor II; apoAII;
KW   factor V; factor VII; factor VIII; factor IX; tick-borne encephalitis;
KW   factor X; factor XI; factor XII; factor XIII; protein C; protein S; VV;
KW   von Willebrand factor; hepatitis B; hepatitis C; hepatitis E; malaria;
KW   plasminogen; apolipo protein; viral antigen; infectious bronchitis;
KW   TBE; tetanus; HIV; HSV; pertussis; Newcastle disease; vaccine; sa.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   promoter        60..65
FT                   /*tag= a
FT                   /note= "Vaccinia virus late promoter consensus
FT                   sequence"
PN   AU9219580-A.
PD   04-MAR-1993.
PF   10-JUL-1992; 019580.
PR   26-AUG-1991; EP-114300.
PA   (IMMO ) IMMUNO AG.
PI   Dorner F,  Falkner FG,  Scheiflinger F;
DR   WPI; 93-126452/16.
PT   Prepn. of recombinant fowl pox virus for e.g. expression of
PT   proteins - by inserting foreign DNA into enlarged inter-genic
PT   region so FPV TK-gene remains intact and codes for entire
PT   thymidine kinase
PS   Disclosure; Fig 18; 92pp; English.
CC   The sequences given in Q41021-25 represent mutations of the vaccinia
CC   virus (VV) P11 promoter sequence which were included in the expression
CC   plasmids of the invention.  The sequences were unclear when given in
CC   the specification and so represent the "best guess" of sequence. These
CC   plasmids further comprised multiple cloning sites (see Q41015-16), the
CC   FPV thymidine kinase (tk) gene and the tk gene intergenic region (see
CC   also Q41011-12).  These plasmids may be used to express foreign
CC   proteins which require post-translational modification eg. Factors II,
CC   V, VII, VIII, IX, X, XI, XII, XIII, proteins C and S, von Willebrand
CC   factor, plasminogen and derivatives, apolipo proteins, eg. apoAI and
CC   apoAII, and viral antigens, eg. hepatitis B- antigens and the antigens
CC   of hepatitis C and E viruses, of tick-borne encephalitis (TBE) virus,
CC   antigens of HIV, HSV and whole or partial sequences of antigens which
CC   cause pertussis, infectious bronchitis, tetanus, malaria, Marck's
CC   disease and Newcastle disease, these antigens being useful as vaccines.
SQ   Sequence   87 BP;    31 A;    5 C;    11 G;    37 T;
```



```
  Query Match           23.0%;  Score 14;  DB 6;  Length 87;
  Best Local Similarity  93.3%;  Pred. No. 9.69e+01;
  Matches   14;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db     35 tttnatttttgttttt 49
          ||| ||||||||||||
Cp     31 TTTTATTTTGTTTTT 17


Search completed: Tue Dec 22 21:21:27 1998
Job time : 135 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Tue Dec 22 21:07:31 1998; MasPar time 131.52 Seconds
1038.887 Million cell updates/sec

Tabular output not generated.

Title: >US-08-530-112A-1449
Description: (1-62) from US08530112A.seq
Perfect Score: 61
N.A. Sequence: 1 GATCCCCATTCTTTTCAAAA.....GCAAGGAAGACTGANGGAAA 62
Comp: CTAGGGGTAAGAAAAGTTTT.....CGTTCCTTCTGACTNCCTTT

Scoring table: TABLE jmetric
Gap 60

Nmatch STD : Dbase 0; Query 0

Searched: 567134 seqs, 1101898692 bases x 2

Post-processing: Minimum Match 0%
Listing first 1000 summaries

Database: embl55
1:em_ba 2:em_fun 3:em_htg 4:em_huml 5:em_hum2 6:em_in
7:em_om 8:em_or 9:em_ov 10:em_pat 11:em_ph 12:em_pl
13:em_ro 14:em_vi
Database: genbank107
15:gb_ba1 16:gb_ba2 17:gb_htg 18:gb_in 19:gb_om 20:gb_ov
21:gb_pat 22:gb_ph 23:gb_pl1 24:gb_pl2 25:gb_prl
26:gb_pr2 27:gb_pr3 28:gb_ro 29:gb_st 30:gb_sts 31:gb_sy
32:gb_un 33:gb_vi

Statistics: Mean 7.268; Variance 3.596; scale 2.021

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
c	1	44	72.1	113241	26	HUAC002425	Homo sapiens Chromosom	4.05e-17
	2	44	72.1	156641	26	HUAC002544	Homo sapiens Chromosom	4.05e-17
	3	22	36.1	116271	17	HS39417	Human DNA sequence ***	7.55e-02
	4	21	34.4	268	30	HSPA32H10	H.sapiens flow-sorted	3.07e-01
	5	20	32.8	41351	18	CELF42C5	Caenorhabditis elegans	1.22e+00
	6	20	32.8	43514	27	AC005545	Homo sapiens chromosom	1.22e+00
c	7	20	32.8	78065	26	AC005355	Homo sapiens chromosom	1.22e+00
	8	20	32.8	85132	26	U82668	Homo sapiens shox gene	1.22e+00
c	9	20	32.8	100000	26	AP000049	Homo sapiens genomic D	1.22e+00
	10	20	32.8	105984	17	AC005087	*** SEQUENCING IN PROG	1.22e+00
c	11	20	32.8	129577	26	AC004659	Homo sapiens chromosom	1.22e+00
c	12	20	32.8	193869	17	AC005043	*** SEQUENCING IN PROG	1.22e+00
c	13	20	32.8	208396	26	AC003101	Homo sapiens chromosom	1.22e+00
	14	19	31.1	882	23	SAU25696	Sinapis alba transcrip	4.66e+00
	15	19	31.1	1256	15	MCAAX	M.capricolum DNA for C	4.66e+00
c	16	19	31.1	2746	27	HSA010069	Homo sapiens mRNA for	4.66e+00
	17	19	31.1	3365	18	TTU46561	Tetrahymena thermophil	4.66e+00
	18	19	31.1	29101	18	CEZK1058	Caenorhabditis elegans	4.66e+00
c	19	19	31.1	37558	18	CELD2092	Caenorhabditis elegans	4.66e+00
c	20	19	31.1	79635	26	AC004650	Homo sapiens chromosom	4.66e+00
	21	19	31.1	82697	23	ATT9A21	Arabidopsis thaliana D	4.66e+00
c	22	19	31.1	86574	27	HS833B7	Human DNA sequence fro	4.66e+00
c	23	19	31.1	102484	26	HS510H16	Homo sapiens DNA seque	4.66e+00
c	24	19	31.1	114411	26	HUAC002551	Human Chromosome 16 BA	4.66e+00
c	25	19	31.1	121024	23	CHMPXX	Liverwort Marchantia p	4.66e+00
c	26	19	31.1	133191	26	AC004027	Human PAC clone DJ0791	4.66e+00
c	27	19	31.1	136299	26	HS359N14	Human DNA sequence fro	4.66e+00
c	28	19	31.1	144136	17	AC004987	*** SEQUENCING IN PROG	4.66e+00
	29	19	31.1	153687	17	HS441J1	Human DNA sequence ***	4.66e+00
c	30	19	31.1	160210	17	HS500L14	Human DNA sequence ***	4.66e+00
c	31	19	31.1	161146	17	HS323A16	Human DNA sequence ***	4.66e+00
	32	19	31.1	162701	26	AC004129	Homo sapiens BAC clone	4.66e+00
	33	19	31.1	177137	17	AC005089	*** SEQUENCING IN PROG	4.66e+00
c	34	19	31.1	183039	17	AC005639	*** SEQUENCING IN PROG	4.66e+00
c	35	19	31.1	185402	17	AC003066	*** SEQUENCING IN PROG	4.66e+00
c	36	19	31.1	192078	17	AC004885	*** SEQUENCING IN PROG	4.66e+00
	37	19	31.1	196643	17	CEY1A5	Caenorhabditis elegans	4.66e+00
	38	18	29.5	309	30	G08858	human STS CHLC.GATA22D	1.72e+01
	39	18	29.5	351	30	G35487	STS h14a397 5, seque	1.72e+01
	40	18	29.5	367	26	HSU3B2	Homo sapiens U3 small	1.72e+01
	41	18	29.5	394	30	G01658	chicken STS ADL238.	1.72e+01
	42	18	29.5	435	25	HSGLNRS13	H.sapiens gene for glu	1.72e+01

	43	18	29.5	441	28	RATANTREL	Rat carcinoembryonic a	1.72e+01
c	44	18	29.5	491	30	G41237	B689016/SP6 Human Chro	1.72e+01
c	45	18	29.5	1548	26	AB005216	Homo sapiens mRNA for	1.72e+01
c	46	18	29.5	1628	25	HSU42389	human neuropeptide y/p	1.72e+01
	47	18	29.5	2012	18	DROORFB	D.virilis homeobox pro	1.72e+01
c	48	18	29.5	2180	16	AF056933	Bacillus thuringiensis	1.72e+01
c	49	18	29.5	2280	25	HSU17077	Human BENE mRNA, parti	1.72e+01
	50	18	29.5	3192	18	D88027	Dictyostelium discoide	1.72e+01
c	51	18	29.5	3444	27	HSNPYY2S2	Human type 2 neuropept	1.72e+01
	52	18	29.5	7064	28	AF013262	Mus musculus lumican (	1.72e+01
	53	18	29.5	7395	32	KTTBMAXVR	T.brucei kinetoplast m	1.72e+01
c	54	18	29.5	37290	27	HSB11B7	Human DNA sequence fro	1.72e+01
c	55	18	29.5	38017	26	AC000396	Genomic sequence from	1.72e+01
c	56	18	29.5	38640	18	CELF44A2	Caenorhabditis elegans	1.72e+01
c	57	18	29.5	39338	26	AC004644	Homo sapiens chromosom	1.72e+01
c	58	18	29.5	39416	18	CEB0001	Caenorhabditis elegans	1.72e+01
	59	18	29.5	40154	18	CELT06F4	Caenorhabditis elegans	1.72e+01
c	60	18	29.5	41413	26	AC003050	Human cosmid g5129g127	1.72e+01
c	61	18	29.5	66004	18	CEC07A9	Caenorhabditis elegans	1.72e+01
c	62	18	29.5	75723	17	AC005442	*** SEQUENCING IN PROG	1.72e+01
	63	18	29.5	76169	26	AC003014	Human PAC clone DJ290B	1.72e+01
c	64	18	29.5	84485	18	AC004351	Drosophila melanogaste	1.72e+01
	65	18	29.5	90271	17	AC004899	*** SEQUENCING IN PROG	1.72e+01
c	66	18	29.5	94386	26	HS452M16	Human DNA sequence ***	1.72e+01
c	67	18	29.5	100000	26	AP000056	Homo sapiens genomic D	1.72e+01
	68	18	29.5	102717	26	AC004079	Homo sapiens PAC clone	1.72e+01
	69	18	29.5	105576	25	HS106I20B	Human DNA sequence fro	1.72e+01
	70	18	29.5	109013	17	HS598P2	Human DNA sequence ***	1.72e+01
	71	18	29.5	117132	17	AC004954	*** SEQUENCING IN PROG	1.72e+01
c	72	18	29.5	117435	26	HUAC003964	Homo sapiens Chromosom	1.72e+01
	73	18	29.5	122459	17	HSAC000384	*** SEQUENCING IN PROG	1.72e+01
	74	18	29.5	123416	17	AC004847	*** SEQUENCING IN PROG	1.72e+01
c	75	18	29.5	125590	18	CEY49E10	Caenorhabditis elegans	1.72e+01
c	76	18	29.5	128148	28	AC003063	Mus musculus Chromosom	1.72e+01
c	77	18	29.5	132513	26	AC004520	Homo sapiens BAC clone	1.72e+01
	78	18	29.5	134149	17	AC004944	*** SEQUENCING IN PROG	1.72e+01
	79	18	29.5	135365	17	HS191L6	Human DNA sequence ***	1.72e+01
	80	18	29.5	141066	17	AC005534	*** SEQUENCING IN PROG	1.72e+01
	81	18	29.5	143747	26	HS370M22	Human DNA sequence fro	1.72e+01
c	82	18	29.5	144387	17	AC004898	*** SEQUENCING IN PROG	1.72e+01
c	83	18	29.5	144763	17	CEY40D7	Caenorhabditis elegans	1.72e+01
c	84	18	29.5	146322	17	HS108K11	Human DNA sequence ***	1.72e+01
	85	18	29.5	146810	26	HS179I15A	Human DNA sequence fro	1.72e+01
c	86	18	29.5	148280	26	AC004474	Homo sapiens chromosom	1.72e+01
	87	18	29.5	159245	17	AC004819	*** SEQUENCING IN PROG	1.72e+01
c	88	18	29.5	161076	27	HS272J12	Human DNA sequence fro	1.72e+01
c	89	18	29.5	162268	26	AC000097	Homo sapiens Chromosom	1.72e+01
c	90	18	29.5	174424	27	HSU52112	Homo sapiens Xq28 geno	1.72e+01
	91	18	29.5	174952	26	AC002460	Human BAC clone RG191D	1.72e+01
	92	18	29.5	175405	17	AC005589	*** SEQUENCING IN PROG	1.72e+01
	93	18	29.5	176433	26	HS71L16	Human DNA sequence fro	1.72e+01

	94	18	29.5	176511	27	AC003035	Homo sapiens Xp22 bins	1.72e+01
	95	18	29.5	181331	17	AC005230	*** SEQUENCING IN PROG	1.72e+01
c	96	18	29.5	192609	17	AC005345	*** SEQUENCING IN PROG	1.72e+01
	97	18	29.5	198499	17	AC004870	*** SEQUENCING IN PROG	1.72e+01
	98	18	29.5	199101	28	MMAE000665	Mus musculus TCR beta	1.72e+01
c	99	18	29.5	200349	27	HSUB5195	Homo sapiens BAC129, c	1.72e+01
	100	18	29.5	210000	17	AC005674	*** SEQUENCING IN PROG	1.72e+01
	101	18	29.5	210672	17	HS179I15	Human DNA sequence ***	1.72e+01
c	102	18	29.5	244649	17	AC005308	*** SEQUENCING IN PROG	1.72e+01
c	103	18	29.5	250529	26	HUAE000658	Homo sapiens T-cell re	1.72e+01
	104	18	29.5	251364	28	MMAE000664	Mus musculus TCR beta	1.72e+01
	105	18	29.5	259894	26	HUAC002302	Homo sapiens Chromosom	1.72e+01
c	106	18	29.5	281376	17	CEY39C12	Caenorhabditis elegans	1.72e+01
c	107	18	29.5	300197	17	CEY54G9	Caenorhabditis elegans	1.72e+01
c	108	18	29.5	345081	17	HS404L14	Human DNA sequence ***	1.72e+01
	109	17	27.9	155	26	HSE6APE10	H.sapiens E6-AP gene e	6.14e+01
c	110	17	27.9	194	24	AB014345	Oryza sativa RNA, endo	6.14e+01
	111	17	27.9	277	27	HSUBE3A16	Homo sapiens UBE3A gen	6.14e+01
c	112	17	27.9	303	30	G20995	human STS WI-30550.	6.14e+01
c	113	17	27.9	564	30	G31273	sy414g1-L Human (A.Gni	6.14e+01
c	114	17	27.9	1116	26	HUMNAIIIC	Human clone H204 N-ace	6.14e+01
c	115	17	27.9	1116	21	E12964	DNA fragment which hav	6.14e+01
c	116	17	27.9	1256	23	ATHPRPHA	Arabidopsis thaliana p	6.14e+01
c	117	17	27.9	1274	23	SCYKL156W	S.cerevisiae chromosom	6.14e+01
c	118	17	27.9	1483	28	S80989	Nkx-5.2-NK-related hom	6.14e+01
c	119	17	27.9	1650	22	BH25CDNAMT	Bacillus amyloliquefac	6.14e+01
	120	17	27.9	1726	26	HSUBE3AL6	Homo sapiens E6-AP ubi	6.14e+01
c	121	17	27.9	1744	25	HUMNAGT	Homo sapiens N-acetylg	6.14e+01
	122	17	27.9	1746	23	SCYBR042C	S.cerevisiae chromosom	6.14e+01
	123	17	27.9	1782	28	S79461	Lcn=lumican [mice, cor	6.14e+01
	124	17	27.9	2009	27	HSUBE6AP3	Homo sapiens E6-AP ubi	6.14e+01
c	125	17	27.9	2046	23	AF001505	Oryza sativa putative	6.14e+01
	126	17	27.9	2084	23	SDU47419	Solanum dulcamara NADH	6.14e+01
	127	17	27.9	2092	23	AVU08914	Anthocercis viscosa ch	6.14e+01
c	128	17	27.9	2187	22	RHO11SMT	B.subtilis phage RHO11	6.14e+01
c	129	17	27.9	2540	23	NEUMTRGS2	N.crassa cyt-6 mutant	6.14e+01
c	130	17	27.9	2755	28	AF029107	Rattus norvegicus Mint	6.14e+01
c	131	17	27.9	2805	16	AF042825	Pyrococcus furiosus DN	6.14e+01
	132	17	27.9	3056	24	AF033210	Pneumocystis carinii f	6.14e+01
	133	17	27.9	3083	24	AF033209	Pneumocystis carinii f	6.14e+01
	134	17	27.9	3089	24	AF033208	Pneumocystis carinii f	6.14e+01
	135	17	27.9	3186	26	HSU84404	Human E6-associated pr	6.14e+01
c	136	17	27.9	3320	23	SCYKL157W	S.cerevisiae chromosom	6.14e+01
c	137	17	27.9	3326	23	SCYNL082W	S.cerevisiae chromosom	6.14e+01
	138	17	27.9	3371	28	MMU010605	Mus musculus mRNA for	6.14e+01
c	139	17	27.9	3410	23	YSCPMS1A	S.cerevisiae PMS1 gene	6.14e+01
c	140	17	27.9	3502	23	SCAPE2G	S.cerevisiae APE2 gene	6.14e+01
c	141	17	27.9	3634	23	SCYNL081C	S.cerevisiae chromosom	6.14e+01
c	142	17	27.9	4068	18	PFU37225	Plasmodium falciparum	6.14e+01
c	143	17	27.9	4819	20	MGU49431	Meleagris gallopavo er	6.14e+01
c	144	17	27.9	5865	18	CEY16B4A	Caenorhabditis elegans	6.14e+01

	145	17	27.9	6036	26	HSBCL7AIG	H.sapiens DNA for BCL7 6.14e+01
	146	17	27.9	8142	16	AF053765	Bacillus megaterium Ar 6.14e+01
c	147	17	27.9	8973	18	CEB0331	Caenorhabditis elegans 6.14e+01
c	148	17	27.9	10277	15	AE001176	Borrelia burgdorferi (6.14e+01
c	149	17	27.9	12391	15	HPAE000654	Helicobacter pylori se 6.14e+01
c	150	17	27.9	13936	24	AB014344	Oryza rufipogon RNA, e 6.14e+01
c	151	17	27.9	13952	23	RICPPA	Rice endogenous double 6.14e+01
c	152	17	27.9	15433	28	RATRSKG5	R.norvegicus tonin (RS 6.14e+01
c	153	17	27.9	16390	28	MUSSEQAA	Mus musculus (clone R2 6.14e+01
	154	17	27.9	16407	18	CEY51B9A	Caenorhabditis elegans 6.14e+01
c	155	17	27.9	17684	19	RABAS1CG	Oryctolagus cuniculus 6.14e+01
	156	17	27.9	19743	18	CEF48C5	Caenorhabditis elegans 6.14e+01
c	157	17	27.9	22566	25	CEBGLOBIN	Cebus albifrons epsilo 6.14e+01
c	158	17	27.9	22887	18	CER01H5	Caenorhabditis elegans 6.14e+01
	159	17	27.9	26151	18	CELW05H9	Caenorhabditis elegans 6.14e+01
c	160	17	27.9	26600	18	CEPJ7B1	Caenorhabditis elegans 6.14e+01
	161	17	27.9	27013	18	CEW01B2	Caenorhabditis elegans 6.14e+01
	162	17	27.9	30027	18	CELT24D11	Caenorhabditis elegans 6.14e+01
	163	17	27.9	30101	18	PFSC03015	Plasmodium falciparum 6.14e+01
c	164	17	27.9	30323	18	CEP58E6	Caenorhabditis elegans 6.14e+01
	165	17	27.9	31162	18	PFSC03013	Plasmodium falciparum 6.14e+01
c	166	17	27.9	31190	18	CEZK892	Caenorhabditis elegans 6.14e+01
c	167	17	27.9	31431	23	SC23CDS	S.cerevisiae PMS1, TPM 6.14e+01
c	168	17	27.9	33490	18	CEZK20	Caenorhabditis elegans 6.14e+01
c	169	17	27.9	33819	18	CET28C6	Caenorhabditis elegans 6.14e+01
	170	17	27.9	33917	26	AC004410	Homo sapiens chromosom 6.14e+01
	171	17	27.9	34073	18	CEF11D11	Caenorhabditis elegans 6.14e+01
c	172	17	27.9	34305	18	CEW02B12	Caenorhabditis elegans 6.14e+01
	173	17	27.9	34864	26	U73649	Human Chromosome 11 Co 6.14e+01
	174	17	27.9	35807	18	CET13H5	Caenorhabditis elegans 6.14e+01
c	175	17	27.9	36075	18	CELT27A1	Caenorhabditis elegans 6.14e+01
c	176	17	27.9	36272	23	SCDCHR11	S.cerevisiae (S288C) 3 6.14e+01
	177	17	27.9	36649	25	HSU19D8	Human DNA sequence fro 6.14e+01
	178	17	27.9	36789	18	CELK07C6	Caenorhabditis elegans 6.14e+01
	179	17	27.9	38010	23	SPAC3F10	S.pombe chromosome I c 6.14e+01
c	180	17	27.9	38324	26	HSBAT2	H.sapiens Bat2 gene. 6.14e+01
	181	17	27.9	38608	18	CEF36H1	Caenorhabditis elegans 6.14e+01
	182	17	27.9	39526	26	AC003961	Human Cosmid g5129z180 6.14e+01
c	183	17	27.9	40852	18	CEF11C1	Caenorhabditis elegans 6.14e+01
c	184	17	27.9	40895	18	CELC01B7	Caenorhabditis elegans 6.14e+01
c	185	17	27.9	43104	26	AC004257	Homo sapiens chromosom 6.14e+01
	186	17	27.9	43593	26	AC003960	Human Cosmid g5129g129 6.14e+01
	187	17	27.9	43632	26	AC004144	Homo sapiens chromosom 6.14e+01
	188	17	27.9	44935	26	AC005394	Homo sapiens chromosom 6.14e+01
	189	17	27.9	45302	17	HSAJ9613	Homo sapiens *** SEQUE 6.14e+01
c	190	17	27.9	48469	26	AC002484	Human cosmid clone LL1 6.14e+01
c	191	17	27.9	48974	28	AF007560	Mus musculus presenili 6.14e+01
c	192	17	27.9	53419	28	MMU35323	Mus musculus H2-M alph 6.14e+01
c	193	17	27.9	54118	18	CELT12A2	Caenorhabditis elegans 6.14e+01
c	194	17	27.9	54724	17	AC002555	*** SEQUENCING IN PROG 6.14e+01
c	195	17	27.9	61971	18	AC002501	Drosophila melanogaste 6.14e+01

c	196	17	27.9	70069	26	AC005206	Homo sapiens chromosom 6.14e+01
c	197	17	27.9	73362	26	AC005144	Homo sapiens chromosom 6.14e+01
c	198	17	27.9	78344	23	TM017A05	Arabidopsis thaliana B 6.14e+01
	199	17	27.9	79109	24	AB016871	Arabidopsis thaliana g 6.14e+01
	200	17	27.9	79931	26	HSAC002122	Human unknown clone GS 6.14e+01
c	201	17	27.9	81357	27	AC005792	Homo sapiens chromosom 6.14e+01
c	202	17	27.9	81421	17	CEC53C7	Caenorhabditis elegans 6.14e+01
	203	17	27.9	82976	27	AC005031	Homo sapiens P1 clone 6.14e+01
	204	17	27.9	84118	17	CEY5P2	Caenorhabditis elegans 6.14e+01
c	205	17	27.9	84124	17	HSAJ9614	Homo sapiens *** SEQUE 6.14e+01
c	206	17	27.9	85020	17	AB018113	*** SEQUENCING IN PROG 6.14e+01
c	207	17	27.9	85444	26	AP000023	Homo sapiens genomic D 6.14e+01
c	208	17	27.9	87114	26	AC005353	Homo sapiens chromosom 6.14e+01
c	209	17	27.9	88528	26	AC003681	Human PAC clone DJ0394 6.14e+01
	210	17	27.9	89743	28	AC003062	Mouse Chromosome 16 Re 6.14e+01
c	211	17	27.9	94106	27	AC005667	Homo sapiens chromosom 6.14e+01
	212	17	27.9	94643	17	HS796I11	Human DNA sequence *** 6.14e+01
c	213	17	27.9	98543	26	AC005352	Homo sapiens chromosom 6.14e+01
	214	17	27.9	99083	18	CEY37A1B	Caenorhabditis elegans 6.14e+01
	215	17	27.9	100000	26	AP000037	Homo sapiens genomic D 6.14e+01
c	216	17	27.9	100000	26	AP000011	Homo sapiens genomic D 6.14e+01
	217	17	27.9	100427	26	AC004048	Homo sapiens chromosom 6.14e+01
	218	17	27.9	103632	24	AF074021	Arabidopsis thaliana B 6.14e+01
c	219	17	27.9	105580	25	HS36J3	Human DNA sequence fro 6.14e+01
c	220	17	27.9	106049	17	AC005049	*** SEQUENCING IN PROG 6.14e+01
	221	17	27.9	106148	26	HS73A14	Human DNA sequence fro 6.14e+01
	222	17	27.9	107910	26	HUAC002394	Human Chromosome 16 BA 6.14e+01
c	223	17	27.9	109389	26	AC002530	Human BAC clone RG341D 6.14e+01
	224	17	27.9	109821	27	HS10C16	Human DNA sequence fro 6.14e+01
c	225	17	27.9	110858	25	HUM2G3A	Human PAC clone 2G3A f 6.14e+01
	226	17	27.9	112400	17	HS30A23	Human DNA sequence *** 6.14e+01
	227	17	27.9	112575	17	HS715N11	Human DNA sequence *** 6.14e+01
c	228	17	27.9	113168	17	HS223H9	Human DNA sequence *** 6.14e+01
c	229	17	27.9	113872	26	HS57G9	Human DNA sequence fro 6.14e+01
c	230	17	27.9	113891	27	AC004797	Homo sapiens chromosom 6.14e+01
	231	17	27.9	114330	27	AC004772	Homo sapiens chromosom 6.14e+01
	232	17	27.9	114736	18	PFMAL3P3	Plasmodium falciparum 6.14e+01
	233	17	27.9	114977	17	AC005532	*** SEQUENCING IN PROG 6.14e+01
	234	17	27.9	115045	26	AC004414	Homo sapiens PAC clone 6.14e+01
	235	17	27.9	117193	17	HS449O17	Human DNA sequence *** 6.14e+01
	236	17	27.9	117939	27	HS211A9	Human DNA sequence fro 6.14e+01
c	237	17	27.9	118791	17	HS1177I5	Human DNA sequence *** 6.14e+01
	238	17	27.9	120703	26	AC002364	Homo sapiens Xp22 Cosm 6.14e+01
c	239	17	27.9	121464	17	AC004842	*** SEQUENCING IN PROG 6.14e+01
	240	17	27.9	121811	17	AC000096	*** SEQUENCING IN PROG 6.14e+01
	241	17	27.9	122176	28	AC003997	Mouse BAC mbac20 from 6.14e+01
c	242	17	27.9	123004	26	HS196E23	Human DNA sequence fro 6.14e+01
c	243	17	27.9	123386	24	F12F1	Arabidopsis thaliana c 6.14e+01
c	244	17	27.9	123556	26	HS507I15	Human DNA sequence fro 6.14e+01
	245	17	27.9	124990	26	HS434P1	Human DNA sequence fro 6.14e+01
	246	17	27.9	125301	26	HS388N15	Human DNA sequence fro 6.14e+01

c	247	17	27.9	125971	26	HS363L9	Human DNA sequence fro 6.14e+01
	248	17	27.9	128978	26	HSAC002073	Human PAC clone DJ515N 6.14e+01
	249	17	27.9	129281	17	AC004999	*** SEQUENCING IN PROG 6.14e+01
c	250	17	27.9	129957	17	HS415G2	Human DNA sequence *** 6.14e+01
	251	17	27.9	130755	26	AC004253	Homo sapiens chromosom 6.14e+01
	252	17	27.9	131078	27	HSU80017	Homo sapiens basic tra 6.14e+01
c	253	17	27.9	131100	17	HS466N1	Human DNA sequence *** 6.14e+01
c	254	17	27.9	131223	17	AC004853	*** SEQUENCING IN PROG 6.14e+01
	255	17	27.9	132090	28	AC005402	Mus musculus clone UWG 6.14e+01
c	256	17	27.9	132790	26	HS167A14	Human DNA sequence fro 6.14e+01
	257	17	27.9	132826	17	AC005627	*** SEQUENCING IN PROG 6.14e+01
	258	17	27.9	133618	17	HS984G1	Human DNA sequence *** 6.14e+01
c	259	17	27.9	134474	26	HS33F8	Human DNA sequence fro 6.14e+01
c	260	17	27.9	134743	26	HUAC002300	Homo sapiens Chromosom 6.14e+01
c	261	17	27.9	135434	17	HS465N24	Human DNA sequence *** 6.14e+01
c	262	17	27.9	136389	17	AC005074	*** SEQUENCING IN PROG 6.14e+01
c	263	17	27.9	136755	26	AC002529	Human BAC clone RG161N 6.14e+01
c	264	17	27.9	139042	17	AC004600	*** SEQUENCING IN PROG 6.14e+01
c	265	17	27.9	139919	26	AC002433	Human BAC clone RG317M 6.14e+01
	266	17	27.9	140823	17	HS1104E15	Human DNA sequence *** 6.14e+01
c	267	17	27.9	142396	25	HSAC000120	Human BAC clone RG161K 6.14e+01
c	268	17	27.9	142640	17	AC004936	*** SEQUENCING IN PROG 6.14e+01
c	269	17	27.9	144820	17	AC005280	*** SEQUENCING IN PROG 6.14e+01
	270	17	27.9	146828	17	AC004968	*** SEQUENCING IN PROG 6.14e+01
	271	17	27.9	148583	17	CEY17D7	Caenorhabditis elegans 6.14e+01
c	272	17	27.9	148600	26	AC002367	Homo sapiens Xp22 PAC 6.14e+01
	273	17	27.9	148883	25	HS358H7	Human DNA sequence fro 6.14e+01
	274	17	27.9	149298	26	AP000031	Homo sapiens genomic D 6.14e+01
	275	17	27.9	150140	17	AC005478	*** SEQUENCING IN PROG 6.14e+01
c	276	17	27.9	151609	17	HS410I8	Human DNA sequence *** 6.14e+01
	277	17	27.9	151798	26	HS435C23	Human DNA sequence fro 6.14e+01
	278	17	27.9	153051	17	HSAJ9612	Homo sapiens *** SEQUE 6.14e+01
	279	17	27.9	155176	26	HS453P22	Human DNA sequence fro 6.14e+01
c	280	17	27.9	155176	26	HS453P22	Human DNA sequence fro 6.14e+01
c	281	17	27.9	155398	17	AC003020	*** SEQUENCING IN PROG 6.14e+01
	282	17	27.9	157412	17	HS376D21	Human DNA sequence *** 6.14e+01
	283	17	27.9	160899	17	AC004869	*** SEQUENCING IN PROG 6.14e+01
	284	17	27.9	161475	26	HS164C20	Homo sapiens DNA seque 6.14e+01
c	285	17	27.9	162067	17	AC005483	*** SEQUENCING IN PROG 6.14e+01
	286	17	27.9	163023	17	AC004815	*** SEQUENCING IN PROG 6.14e+01
	287	17	27.9	163196	26	AC005220	Homo sapiens chromosom 6.14e+01
	288	17	27.9	164049	17	AC005232	*** SEQUENCING IN PROG 6.14e+01
	289	17	27.9	164135	17	AC004990	*** SEQUENCING IN PROG 6.14e+01
c	290	17	27.9	165363	17	HS459L4	Human DNA sequence *** 6.14e+01
	291	17	27.9	169693	26	HS57A13	Human DNA sequence fro 6.14e+01
	292	17	27.9	170245	26	HS109F14	Human DNA sequence *** 6.14e+01
c	293	17	27.9	171952	28	AC002324	Mus musculus chromosom 6.14e+01
	294	17	27.9	172084	26	AC005245	Homo sapiens chromosom 6.14e+01
	295	17	27.9	172241	26	AC004014	Human BAC clone RG022K 6.14e+01
c	296	17	27.9	172984	17	HS862K6	Human DNA sequence *** 6.14e+01
	297	17	27.9	172995	17	AC004926	*** SEQUENCING IN PROG 6.14e+01

c	298	17	27.9	176968	27	AC005669	Homo sapiens chromosom 6.14e+01
c	299	17	27.9	177380	26	AC002457	Human BAC clone RG060P 6.14e+01
	300	17	27.9	177455	17	AC004980	*** SEQUENCING IN PROG 6.14e+01
	301	17	27.9	178660	26	HSG1150	Homo sapiens chromosom 6.14e+01
	302	17	27.9	179246	17	AC004854	*** SEQUENCING IN PROG 6.14e+01
	303	17	27.9	181525	17	HS996D20	Human DNA sequence *** 6.14e+01
c	304	17	27.9	1824435	26	AC003078	Human BAC clone GS1170 6.14e+01
c	305	17	27.9	187051	17	AC004688	*** SEQUENCING IN PROG 6.14e+01
c	306	17	27.9	187647	17	CEY16B4	Caenorhabditis elegans 6.14e+01
	307	17	27.9	188251	17	AC005041	*** SEQUENCING IN PROG 6.14e+01
	308	17	27.9	189999	17	AC004617	*** SEQUENCING IN PROG 6.14e+01
c	309	17	27.9	190943	17	AC005037	*** SEQUENCING IN PROG 6.14e+01
	310	17	27.9	191908	17	CEY51B9	Caenorhabditis elegans 6.14e+01
	311	17	27.9	193161	27	AC005500	, complete sequence. 6.14e+01
	312	17	27.9	198605	17	AC005003	*** SEQUENCING IN PROG 6.14e+01
c	313	17	27.9	200167	26	AC002480	Human BAC clone GS055K 6.14e+01
c	314	17	27.9	202096	17	AC005036	*** SEQUENCING IN PROG 6.14e+01
c	315	17	27.9	202763	26	AC005181	Homo sapiens chromosom 6.14e+01
c	316	17	27.9	206817	23	ATFCA7	Arabidopsis thaliana D 6.14e+01
	317	17	27.9	208609	26	HS389A20	Human DNA sequence fro 6.14e+01
c	318	17	27.9	210000	17	AC005690	*** SEQUENCING IN PROG 6.14e+01
	319	17	27.9	214521	17	HS172B20	Human DNA sequence *** 6.14e+01
c	320	17	27.9	218074	27	HUAC002044	Human Chromosome 16 BA 6.14e+01
c	321	17	27.9	219062	17	HS407F17	Human DNA sequence *** 6.14e+01
	322	17	27.9	224645	17	AC004153	*** SEQUENCING IN PROG 6.14e+01
	323	17	27.9	227968	26	AF053356	Homo sapiens chromosom 6.14e+01
c	324	17	27.9	228500	17	HS212A2	Human DNA sequence *** 6.14e+01
c	325	17	27.9	234053	26	AC002429	Human BAC clone GS200K 6.14e+01
	326	17	27.9	243041	26	AC004388	Homo Sapiens Chromosom 6.14e+01
	327	17	27.9	243820	17	CEY75B12	Caenorhabditis elegans 6.14e+01
	328	17	27.9	250611	28	MMAE000563	Mus musculus TCR beta 6.14e+01
c	329	17	27.9	259894	26	HUAC002302	Homo sapiens Chromosom 6.14e+01
	330	17	27.9	291821	17	CEY48E1	Caenorhabditis elegans 6.14e+01
	331	17	27.9	316170	17	CEY37A1	Caenorhabditis elegans 6.14e+01
	332	17	27.9	317311	17	AC004085	*** SEQUENCING IN PROG 6.14e+01
	333	16	26.2	144	30	G38850	E87A Plasmodium falcip 2.10e+02
	334	16	26.2	184	25	HUMUG4PSA	Human U4 RNA pseudogen 2.10e+02
	335	16	26.2	203	25	HSU44P	Human U4 snRNA pseudog 2.10e+02
	336	16	26.2	228	20	PRIHPAPS12	Prosopium spilonotus D 2.10e+02
	337	16	26.2	280	19	BTU06764	Bos taurus clone bm864 2.10e+02
c	338	16	26.2	295	19	BTDR18	B.taurus DNA dinucleot 2.10e+02
	339	16	26.2	322	30	G16531	human STS SHGC-3285 cl 2.10e+02
	340	16	26.2	344	18	AF015471	Plasmodium falciparum 2.10e+02
	341	16	26.2	360	30	G35452	STS h14a334 5, seque 2.10e+02
	342	16	26.2	366	30	G36175	STS h14a2021 5, sequ 2.10e+02
	343	16	26.2	379	30	G32041	STS MCW282, sequence 2.10e+02
c	344	16	26.2	380	30	HSA183WG9	H.sapiens DNA segment 2.10e+02
c	345	16	26.2	466	25	HSRE07	Human repetitive seque 2.10e+02
	346	16	26.2	556	33	XCU70901	Xestia c-nigrum granul 2.10e+02
	347	16	26.2	603	28	RATNDPK21	Rat nucleoside diphosp 2.10e+02
	348	16	26.2	603	28	S39901	nucleoside diphosphate 2.10e+02

	349	16	26.2	612	28	MMNDIPKB	M.musculus mRNA for nu 2.10e+02
	350	16	26.2	631	28	RATRBLNDP	Sprague-Dawley (p18-12 2.10e+02
	351	16	26.2	652	33	HIV1U36005	Human immunodeficiency 2.10e+02
	352	16	26.2	653	33	HIV1U36003	Human immunodeficiency 2.10e+02
	353	16	26.2	730	20	HTU22067	Heteixalus tricolor 12 2.10e+02
c	354	16	26.2	761	18	LEIKPHURF4	Leishmania tarentolae 2.10e+02
c	355	16	26.2	764	18	LEIKPCAA	L.tarentolae kinetopla 2.10e+02
c	356	16	26.2	764	18	LEINTATPAS	Leishmania tarentolae 2.10e+02
c	357	16	26.2	973	18	PFPCNAA	P.falciparum gene for 2.10e+02
c	358	16	26.2	1193	28	RNU56863	Rattus norvegicus Ly-4 2.10e+02
	359	16	26.2	1241	33	HIVU95444	HIV-1 clone 8 isolate 2.10e+02
	360	16	26.2	1250	23	DDIM3L	D.discoideum M3L gene, 2.10e+02
	361	16	26.2	1255	33	HIVU84851	HIV-1 strain C14 from 2.10e+02
	362	16	26.2	1256	33	HIVU84852	HIV-1 strain C14 from 2.10e+02
	363	16	26.2	1256	33	HIVU84849	HIV-1 strain C14 from 2.10e+02
c	364	16	26.2	1305	18	CBTBCB1	Caenorhabditis briggsa 2.10e+02
	365	16	26.2	1320	33	YFU52413	Yellow fever virus Sen 2.10e+02
	366	16	26.2	1320	33	YFU52403	Yellow fever virus Nig 2.10e+02
	367	16	26.2	1320	33	YFU52389	Yellow fever virus Bra 2.10e+02
	368	16	26.2	1320	33	YFU52392	Yellow fever virus Car 2.10e+02
	369	16	26.2	1320	33	YFU52422	Yellow fever virus Uga 2.10e+02
	370	16	26.2	1320	33	YFU52419	Yellow fever virus Tri 2.10e+02
	371	16	26.2	1320	33	YFU52395	Yellow fever virus Car 2.10e+02
c	372	16	26.2	1329	23	NTL17	N.tabacum Ntl1-NT7 mRN 2.10e+02
c	373	16	26.2	1351	18	CBTCB110	Caenorhabditis briggsa 2.10e+02
c	374	16	26.2	1512	25	HUMREPA	Human replication prot 2.10e+02
c	375	16	26.2	1590	28	RATMHCRT	Rattus norvegicus MHC 2.10e+02
c	376	16	26.2	1708	25	HUMIL1RN	Human interleukin 1 re 2.10e+02
	377	16	26.2	1745	23	ATU59508	Arabidopsis thaliana o 2.10e+02
	378	16	26.2	1833	23	D83025	Arabidopsis thaliana m 2.10e+02
c	379	16	26.2	1846	28	RATCASB5	Rat beta-casein gene, 2.10e+02
c	380	16	26.2	1848	28	MUSOM	Mouse mRNA for mouse o 2.10e+02
	381	16	26.2	1855	23	ATPOX	A.thaliana mRNA for pr 2.10e+02
	382	16	26.2	1906	28	MMU15562	Mus musculus CDC25 (Cd 2.10e+02
c	383	16	26.2	2030	28	S6045892	TSP2=thrombospondin 2 2.10e+02
c	384	16	26.2	2030	28	MUSUTRTBP	Murine thrombospondin 2 2.10e+02
	385	16	26.2	2089	23	CHDCRRBCL	D.crassifolium chlorop 2.10e+02
	386	16	26.2	2119	15	STYCITA	S. typhimurium citA ge 2.10e+02
	387	16	26.2	2124	28	MUSCDC25A	Mus musculus cdc25 pho 2.10e+02
c	388	16	26.2	2143	23	SCSWA2	S.occidentalis SWA2 ge 2.10e+02
	389	16	26.2	2297	23	AF040270	Schizosaccharomyces po 2.10e+02
	390	16	26.2	2367	33	YFVCAPSIDM	Yellow fever virus cap 2.10e+02
c	391	16	26.2	2406	19	OCU82751	Oryctolagus cuniculus 2.10e+02
	392	16	26.2	2452	33	AF052440	Yellow fever virus clo 2.10e+02
	393	16	26.2	2452	33	AF052442	Yellow fever virus clo 2.10e+02
	394	16	26.2	2452	33	AF052447	Yellow fever virus clo 2.10e+02
	395	16	26.2	2452	33	YFPOLP1	Yellow fever virus pol 2.10e+02
c	396	16	26.2	2467	19	OCU82750	Oryctolagus cuniculus 2.10e+02
c	397	16	26.2	2497	23	SCYNL191W	S.cerevisiae chromosom 2.10e+02
	398	16	26.2	2583	18	PFSC03097	Plasmodium falciparum 2.10e+02
c	399	16	26.2	2754	23	SCYNL190W	S.cerevisiae chromosom 2.10e+02

	400	16	26.2	2781	18	AF025690	Musca domestica clone 2.10e+02
	401	16	26.2	2800	23	AF026395	Saccharomyces servazzi 2.10e+02
	402	16	26.2	2854	33	HIVU84850	HIV-1 strain C14 from 2.10e+02
c	403	16	26.2	2938	18	TTU12971	Tetrahymena thermophil 2.10e+02
	404	16	26.2	3041	26	HUMRSC192	Human mRNA for KIAA000 2.10e+02
c	405	16	26.2	3050	23	SCYLL032C	S.cerevisiae chromosom 2.10e+02
c	406	16	26.2	3216	18	CELRAF	Caenorhabditis elegans 2.10e+02
	407	16	26.2	3390	27	AF042782	Homo sapiens galanin r 2.10e+02
	408	16	26.2	3463	23	SPRRNAISP	S.pombe DNA for rRNA g 2.10e+02
c	409	16	26.2	3483	25	HSU82705	Human interferon alpha 2.10e+02
c	410	16	26.2	3486	23	SCYOR378W	S.cerevisiae chromosom 2.10e+02
	411	16	26.2	3566	20	AB006906	Xenopus laevis mRNA fo 2.10e+02
	412	16	26.2	3594	33	YFVCMENS	Yellow fever virus gen 2.10e+02
c	413	16	26.2	3608	23	SCYLL031C	S.cerevisiae chromosom 2.10e+02
c	414	16	26.2	3703	25	HUMHBB221	Human beta-globin gene 2.10e+02
c	415	16	26.2	3840	23	CHEGRPL	E.gracilis DNA for rib 2.10e+02
c	416	16	26.2	4390	24	AF053747	Arabidopsis thaliana d 2.10e+02
	417	16	26.2	4590	18	DROGGBCS	D.melanogaster (clones 2.10e+02
c	418	16	26.2	5026	28	AF055919	Mus musculus putative 2.10e+02
	419	16	26.2	5042	18	PFSC03071	Plasmodium falciparum 2.10e+02
c	420	16	26.2	5544	19	RABMHCDRA	Rabbit MHC class II RL 2.10e+02
	421	16	26.2	5806	32	MMUSF1EXN	M.musculus USF1 (exons 2.10e+02
c	422	16	26.2	5869	15	MYCRPOBC	Mycoplasma galliseptic 2.10e+02
c	423	16	26.2	5951	23	SCYOR377W	S.cerevisiae chromosom 2.10e+02
c	424	16	26.2	6084	15	STRINTE	Streptococcus pneumoni 2.10e+02
	425	16	26.2	6104	18	PFSC04016	Plasmodium falciparum 2.10e+02
c	426	16	26.2	6790	33	AF034622	Tobacco rattle virus R 2.10e+02
c	427	16	26.2	6791	33	TOTRRNA1	Tobacco Rattle virus (2.10e+02
c	428	16	26.2	6791	33	MTRRNA1	Tobacco rattle virus R 2.10e+02
	429	16	26.2	8227	18	AF019362	Drosophila melanogaste 2.10e+02
c	430	16	26.2	9133	23	EGRCPPSAAB	E.gracilis chloroplast 2.10e+02
c	431	16	26.2	9208	23	ATT3T7POL	Arabidopsis thaliana r 2.10e+02
	432	16	26.2	9448	18	CEU22327	Caenorhabditis elegans 2.10e+02
	433	16	26.2	9541	18	CECOLA2IV	C.elegans a2(IV) colla 2.10e+02
c	434	16	26.2	9764	23	ATDSDDRPL	A.thaliana genes encod 2.10e+02
	435	16	26.2	9906	19	SSU37521	Sus scrofa E-selectin 2.10e+02
c	436	16	26.2	10076	15	U32759	Haemophilus influenzae 2.10e+02
	437	16	26.2	10509	18	AF033037	Plasmodium falciparum 2.10e+02
	438	16	26.2	10690	28	MMU41741	Mus musculus transcrip 2.10e+02
	439	16	26.2	10817	15	U67521	Methanococcus jannasch 2.10e+02
c	440	16	26.2	10833	28	MMZP3G	Mus musculus mRNA for 2.10e+02
	441	16	26.2	10862	33	AF052438	Yellow fever virus clo 2.10e+02
	442	16	26.2	10862	33	YFU17067	Yellow fever virus vac 2.10e+02
	443	16	26.2	10862	33	YFU17066	Yellow fever virus vac 2.10e+02
	444	16	26.2	10862	33	AF052445	Yellow fever virus clo 2.10e+02
	445	16	26.2	10862	33	FLYF17DG	Yellow fever virus gen 2.10e+02
	446	16	26.2	10862	33	AF052444	Yellow fever virus clo 2.10e+02
	447	16	26.2	10862	33	AF052439	Yellow fever virus clo 2.10e+02
	448	16	26.2	10862	33	AF052446	Yellow fever virus clo 2.10e+02
	449	16	26.2	10862	33	AF052437	Yellow fever virus clo 2.10e+02
	450	16	26.2	10862	33	YFU21055	Yellow fever virus Fre 2.10e+02

	451	16	26.2	10862	33	YFU54798	Yellow fever virus str 2.10e+02
	452	16	26.2	10862	33	YFVCG	Yellow fever virus, co 2.10e+02
	453	16	26.2	10862	33	YFU21056	Yellow fever virus Fre 2.10e+02
	454	16	26.2	10862	33	YFVRNA	Yellow fever virus gen 2.10e+02
	455	16	26.2	10923	15	HPAE000585	Helicobacter pylori se 2.10e+02
	456	16	26.2	12481	15	U32804	Haemophilus influenzae 2.10e+02
	457	16	26.2	13205	15	U67588	Methanococcus jannasch 2.10e+02
c	458	16	26.2	14616	15	BSU09819	Bacillus subtilis spaB 2.10e+02
c	459	16	26.2	14653	18	CEF28C12	Caenorhabditis elegans 2.10e+02
	460	16	26.2	15298	17	AC005450	*** SEQUENCING IN PROG 2.10e+02
c	461	16	26.2	15865	18	MICOCNNR	Crithidia oncopelti mi 2.10e+02
	462	16	26.2	15921	18	AF034856	Drosophila melanogaste 2.10e+02
	463	16	26.2	15955	17	CEZK883	Caenorhabditis elegans 2.10e+02
c	464	16	26.2	15990	18	CET26H10	Caenorhabditis elegans 2.10e+02
	465	16	26.2	16913	28	MMSEMCLOT	M.musculus DNA coding 2.10e+02
c	466	16	26.2	18158	16	AF009224	Acinetobacter sp. ADP1 2.10e+02
c	467	16	26.2	19185	28	RATCTTG	Rat troponin T cardiac 2.10e+02
	468	16	26.2	20780	18	CEC15C8	Caenorhabditis elegans 2.10e+02
c	469	16	26.2	20992	18	LEIKPMAX	Leishmania tarentolae 2.10e+02
c	470	16	26.2	23326	23	LJGLN1	L.japonicus gln1, pgl 2.10e+02
c	471	16	26.2	23533	15	AE001039	Archaeoglobus fulgidus 2.10e+02
c	472	16	26.2	24768	18	CET07C5	Caenorhabditis elegans 2.10e+02
	473	16	26.2	25030	18	CEF30F8	Caenorhabditis elegans 2.10e+02
	474	16	26.2	25642	18	CELF59G1	Caenorhabditis elegans 2.10e+02
	475	16	26.2	27368	18	CER08D7	Caenorhabditis elegans 2.10e+02
	476	16	26.2	28310	17	AC002548	*** SEQUENCING IN PROG 2.10e+02
	477	16	26.2	28865	27	HSN104C7	Human DNA sequence fro 2.10e+02
c	478	16	26.2	29468	27	HS431H6	Human DNA sequence *** 2.10e+02
	479	16	26.2	30182	18	CEF25B3	Caenorhabditis elegans 2.10e+02
c	480	16	26.2	30320	18	CEC05B5	Caenorhabditis elegans 2.10e+02
c	481	16	26.2	30640	18	CEK02C4	Caenorhabditis elegans 2.10e+02
c	482	16	26.2	30875	27	HSN20A6	Human DNA sequence fro 2.10e+02
c	483	16	26.2	31223	18	CEZK270	Caenorhabditis elegans 2.10e+02
c	484	16	26.2	31268	18	CEH08M01	Caenorhabditis elegans 2.10e+02
c	485	16	26.2	31546	18	CELC25A11	Caenorhabditis elegans 2.10e+02
	486	16	26.2	31664	18	CFF10A3	Caenorhabditis elegans 2.10e+02
	487	16	26.2	32110	18	CEF40G12	Caenorhabditis elegans 2.10e+02
c	488	16	26.2	32310	18	CEF34D10	Caenorhabditis elegans 2.10e+02
c	489	16	26.2	32561	18	CEC51E3	Caenorhabditis elegans 2.10e+02
	490	16	26.2	33013	18	DMU87286	Drosophila melanogaste 2.10e+02
c	491	16	26.2	33682	26	HSN73A10	Human DNA sequence fro 2.10e+02
c	492	16	26.2	34193	18	CEC38C10	Caenorhabditis elegans 2.10e+02
c	493	16	26.2	34576	18	CET23D8	Caenorhabditis elegans 2.10e+02
c	494	16	26.2	34671	18	CELF01G12	Caenorhabditis elegans 2.10e+02
c	495	16	26.2	35609	26	HSU90B3	Human DNA sequence fro 2.10e+02
	496	16	26.2	36000	24	SPCC1672	S.poobe chromosome III 2.10e+02
	497	16	26.2	36051	18	CET22A3	Caenorhabditis elegans 2.10e+02
c	498	16	26.2	36180	18	CET05G5	Caenorhabditis elegans 2.10e+02
	499	16	26.2	36772	18	CELT14B4	Caenorhabditis elegans 2.10e+02
	500	16	26.2	36798	26	HSE78G1	Human DNA sequence fro 2.10e+02
c	501	16	26.2	36812	18	CEF13B10	Caenorhabditis elegans 2.10e+02

	502	16	26.2	37274	26	AC005272	Homo sapiens chromosom 2.10e+02
	503	16	26.2	37666	18	CET24D5	Caenorhabditis elegans 2.10e+02
	504	16	26.2	37685	26	HSAC000371	Human cosmid g1980a181 2.10e+02
	505	16	26.2	37923	18	CELK07H8	Caenorhabditis elegans 2.10e+02
	506	16	26.2	38374	26	HSU95090	Homo sapiens chromosom 2.10e+02
c	507	16	26.2	38418	28	MUSPRNPA	Mus musculus short inc 2.10e+02
	508	16	26.2	38644	26	AC002508	Human Cosmid g5129z171 2.10e+02
c	509	16	26.2	38699	26	AC000070	Homo sapiens Chromosom 2.10e+02
c	510	16	26.2	38924	27	HSE110C7	Human DNA sequence fro 2.10e+02
	511	16	26.2	38969	18	CELF07G6	Caenorhabditis elegans 2.10e+02
	512	16	26.2	39165	18	CEF46B3	Caenorhabditis elegans 2.10e+02
	513	16	26.2	39220	18	CET12A7	Caenorhabditis elegans 2.10e+02
	514	16	26.2	39643	18	CELC25G6	Caenorhabditis elegans 2.10e+02
	515	16	26.2	39658	26	HSU231B4	Human DNA sequence fro 2.10e+02
c	516	16	26.2	39977	18	CEM162	Caenorhabditis elegans 2.10e+02
	517	16	26.2	40050	18	CELC29B4	C. elegans cosmid C29E 2.10e+02
c	518	16	26.2	40457	18	CELF54D7	Caenorhabditis elegans 2.10e+02
	519	16	26.2	40548	23	AB011480	Arabidopsis thaliana g 2.10e+02
c	520	16	26.2	40830	23	SPAC2C6	S.pombe chromosome I c 2.10e+02
c	521	16	26.2	41017	23	CHEGZ	Euglena gracilis Z Chl 2.10e+02
	522	16	26.2	41413	26	AC003050	Human cosmid g5129g127 2.10e+02
	523	16	26.2	41645	18	CELF09G8	C. elegans cosmid F09G 2.10e+02
	524	16	26.2	41645	18	CELC03H5	Caenorhabditis elegans 2.10e+02
c	525	16	26.2	41950	18	CELR03E9	Caenorhabditis elegans 2.10e+02
c	526	16	26.2	42345	17	CEF52D11	Caenorhabditis elegans 2.10e+02
	527	16	26.2	42572	27	AC005514	Homo sapiens chromosom 2.10e+02
	528	16	26.2	42771	26	AC000381	Homo sapiens chromosom 2.10e+02
	529	16	26.2	43309	26	AC002498	Human Cosmid g5129g124 2.10e+02
	530	16	26.2	43407	27	AC005584	Homo sapiens clone UWG 2.10e+02
c	531	16	26.2	43632	26	AC004144	Homo sapiens chromosom 2.10e+02
c	532	16	26.2	43772	26	AC004791	Homo sapiens chromoso 2.10e+02
c	533	16	26.2	45110	17	AC005032	*** SEQUENCING IN PROG 2.10e+02
c	534	16	26.2	45302	17	HSAJ9613	Homo sapiens *** SEQUE 2.10e+02
	535	16	26.2	45806	27	HSB33F2	Human DNA sequence fro 2.10e+02
	536	16	26.2	45931	18	CELC01B10	Caenorhabditis elegans 2.10e+02
	537	16	26.2	45950	17	AC005641	*** SEQUENCING IN PROG 2.10e+02
	538	16	26.2	45976	27	HSU81031	Homo sapiens cosmid 6E 2.10e+02
	539	16	26.2	45976	25	HSAC000028	00023, complete sequen 2.10e+02
c	540	16	26.2	46105	18	CELC55C3	Caenorhabditis elegans 2.10e+02
	541	16	26.2	48469	26	AC002484	Human cosmid clone LL1 2.10e+02
	542	16	26.2	54897	17	CEB0413	Caenorhabditis elegans 2.10e+02
c	543	16	26.2	55136	25	HUMVITDBP	Human vitamin D-bindin 2.10e+02
	544	16	26.2	57328	18	PFSC03065	Plasmodium falciparum 2.10e+02
c	545	16	26.2	58407	16	MII2CG	Methanococcus jannasch 2.10e+02
c	546	16	26.2	59383	18	PFSC03066	Plasmodium falciparum 2.10e+02
c	547	16	26.2	59782	18	PFSC03093	Plasmodium falciparum 2.10e+02
	548	16	26.2	61815	17	AC005453	*** SEQUENCING IN PROG 2.10e+02
	549	16	26.2	62607	27	HS653H13	Human DNA sequence fro 2.10e+02
	550	16	26.2	64767	26	U69730	Homo sapiens Xp22 Cosm 2.10e+02
c	551	16	26.2	69037	28	AC003949	Mus musculus chromosom 2.10e+02
c	552	16	26.2	69964	27	HS496C20	Human DNA sequence fro 2.10e+02

c	553	16	26.2	72847	18	AC004266	Drosophila melanogaste 2.10e+02
c	554	16	26.2	73132	26	HS833B2	Homo sapiens DNA seque 2.10e+02
c	555	16	26.2	73308	25	HUMHBB	Human beta globin regi 2.10e+02
c	556	16	26.2	74279	17	HS1174N9	Human DNA sequence *** 2.10e+02
	557	16	26.2	74371	26	AC005369	Homo sapiens chromosom 2.10e+02
	558	16	26.2	75475	23	AB005235	Arabidopsis thaliana g 2.10e+02
	559	16	26.2	75871	26	AC003953	Homo sapiens chromosom 2.10e+02
	560	16	26.2	77236	17	AC004581	*** SEQUENCING IN PROG 2.10e+02
	561	16	26.2	78616	28	AC000398	Genomic sequence from 2.10e+02
	562	16	26.2	78631	25	HSAC000057	Human BAC clone RG067M 2.10e+02
	563	16	26.2	79248	18	AC004640	Drosophila melanogaste 2.10e+02
	564	16	26.2	80507	27	AC004024	Homo sapiens 12q24.2 P 2.10e+02
	565	16	26.2	81347	24	AB015478	Arabidopsis thaliana g 2.10e+02
	566	16	26.2	82512	25	HUM133K23	Human BAC clone 133K23 2.10e+02
c	567	16	26.2	82976	27	AC005031	Homo sapiens P1 clone 2.10e+02
c	568	16	26.2	83078	24	AF072897	Arabidopsis thaliana B 2.10e+02
	569	16	26.2	83689	24	AB017067	Arabidopsis thaliana g 2.10e+02
c	570	16	26.2	83865	23	AB006700	Arabidopsis thaliana g 2.10e+02
	571	16	26.2	84702	17	AB018114	*** SEQUENCING IN PROG 2.10e+02
	572	16	26.2	84747	17	AC005661	*** SEQUENCING IN PROG 2.10e+02
	573	16	26.2	85214	24	ATAC005396	Arabidopsis thaliana c 2.10e+02
	574	16	26.2	85633	26	HSAC002088	Human BAC clone RG013P 2.10e+02
	575	16	26.2	87013	26	HS528L19	Human DNA sequence fro 2.10e+02
	576	16	26.2	87186	17	AC005454	*** SEQUENCING IN PROG 2.10e+02
c	577	16	26.2	87841	23	AB005247	Arabidopsis thaliana g 2.10e+02
c	578	16	26.2	88036	26	HUAC003119	Human Chromosome 16 BA 2.10e+02
c	579	16	26.2	88358	3	AC004718	*** SEQUENCING IN PROG 2.10e+02
c	580	16	26.2	89219	24	ATT6K22	Arabidopsis thaliana D 2.10e+02
	581	16	26.2	89521	17	HS751L1	Human DNA sequence *** 2.10e+02
	582	16	26.2	90694	17	HS257I9	Human DNA sequence *** 2.10e+02
	583	16	26.2	90832	27	AC003065	Homo sapiens Chromosom 2.10e+02
	584	16	26.2	91539	24	ATAC004261	Arabidopsis thaliana c 2.10e+02
c	585	16	26.2	91638	17	AC002489	*** SEQUENCING IN PROG 2.10e+02
	586	16	26.2	92802	17	AC004710	*** SEQUENCING IN PROG 2.10e+02
	587	16	26.2	92869	26	AC004992	Homo sapiens PAC clone 2.10e+02
	588	16	26.2	93158	23	ATAC002334	Arabidopsis thaliana c 2.10e+02
	589	16	26.2	94882	17	AC002317	*** SEQUENCING IN PROG 2.10e+02
	590	16	26.2	95167	23	ATAC003105	Arabidopsis thaliana c 2.10e+02
	591	16	26.2	95565	26	AC000379	Human Chromosome 15 pa 2.10e+02
c	592	16	26.2	96217	26	AC002410	Human BAC clone RG264L 2.10e+02
c	593	16	26.2	98957	24	ATAC004136	Arabidopsis thaliana c 2.10e+02
c	594	16	26.2	100000	26	AP000012	Homo sapiens genomic D 2.10e+02
c	595	16	26.2	100000	26	AP000046	Homo sapiens genomic D 2.10e+02
	596	16	26.2	100997	26	HS262D12	Homo sapiens DNA seque 2.10e+02
c	597	16	26.2	101574	26	HS206D15	Homo sapiens DNA seque 2.10e+02
	598	16	26.2	101615	26	AC004038	Homo sapiens chromosom 2.10e+02
	599	16	26.2	102204	26	AC003103	Homo sapiens chromosom 2.10e+02
	600	16	26.2	102846	27	AC002483	Homo sapiens PAC clone 2.10e+02
c	601	16	26.2	103319	26	AC004217	Homo sapiens 12q24.1 P 2.10e+02
c	602	16	26.2	105485	26	HUAC002301	Homo sapiens Chromosom 2.10e+02
	603	16	26.2	105754	17	AC005000	*** SEQUENCING IN PROG 2.10e+02

	604	16	26.2	105809	17	HS364I1	Human DNA sequence *** 2.10e+02
c	605	16	26.2	106729	27	AC005320	Human Chromosome 11p14 2.10e+02
c	606	16	26.2	108248	26	HS326I13	Homo sapiens DNA seque 2.10e+02
c	607	16	26.2	108695	26	AC002119	Homo sapiens chromosom 2.10e+02
c	608	16	26.2	109727	17	HS322P7	Human DNA sequence *** 2.10e+02
	609	16	26.2	110000	17	CEY116F11	Caenorhabditis elegans 2.10e+02
	610	16	26.2	110000	17	CEY57G11	Caenorhabditis elegans 2.10e+02
	611	16	26.2	110000	17	CEY113B8	Caenorhabditis elegans 2.10e+02
c	612	16	26.2	110000	17	CEY116A8	Caenorhabditis elegans 2.10e+02
c	613	16	26.2	110000	17	CEY53C10	Caenorhabditis elegans 2.10e+02
	614	16	26.2	110379	17	AC005560	*** SEQUENCING IN PROG 2.10e+02
	615	16	26.2	110858	25	HUM2G3A	Human PAC clone 2G3A f 2.10e+02
	616	16	26.2	111507	26	HS238J17	Human DNA sequence fro 2.10e+02
	617	16	26.2	111650	26	AC005215	Homo sapiens chromosom 2.10e+02
c	618	16	26.2	112088	26	AC005216	Homo sapiens chromosom 2.10e+02
c	619	16	26.2	112846	26	AC004991	Homo sapiens PAC clone 2.10e+02
	620	16	26.2	112863	17	HS482L3	Human DNA sequence *** 2.10e+02
	621	16	26.2	113241	26	HUAC002425	Homo sapiens Chromosom 2.10e+02
c	622	16	26.2	113704	25	HS267P19	Human DNA sequence fro 2.10e+02
c	623	16	26.2	114956	26	HSAC002072	Human PAC clone DJ218B 2.10e+02
	624	16	26.2	116156	23	ATAC002343	Arabidopsis thaliana B 2.10e+02
	625	16	26.2	116650	26	AC004003	Homo sapiens BAC clone 2.10e+02
c	626	16	26.2	117252	17	AC005091	*** SEQUENCING IN PROG 2.10e+02
c	627	16	26.2	117896	17	HS287L14	Human DNA sequence *** 2.10e+02
	628	16	26.2	120229	26	U90095	Human Chromosome 11p15 2.10e+02
	629	16	26.2	120321	17	HS341D10	Human DNA sequence *** 2.10e+02
c	630	16	26.2	121131	26	HS345B16	Human DNA sequence fro 2.10e+02
c	631	16	26.2	121266	17	AC005095	*** SEQUENCING IN PROG 2.10e+02
	632	16	26.2	122000	26	AC003093	Human BAC clone RG385F 2.10e+02
c	633	16	26.2	123234	26	AC002556	Human Chromosome 11 pa 2.10e+02
c	634	16	26.2	124181	17	HS285P3	Human DNA sequence *** 2.10e+02
	635	16	26.2	124752	17	AC004986	*** SEQUENCING IN PROG 2.10e+02
	636	16	26.2	125590	26	AC002382	Human BAC clone RG022J 2.10e+02
	637	16	26.2	125630	26	HSU91319	Human chromosome 16p13 2.10e+02
	638	16	26.2	126352	26	HS23K20	Human DNA sequence fro 2.10e+02
	639	16	26.2	126430	17	HS146H21	Human DNA sequence *** 2.10e+02
c	640	16	26.2	126586	17	AC004915	*** SEQUENCING IN PROG 2.10e+02
	641	16	26.2	126995	26	AC004773	Homo sapiens PAC clone 2.10e+02
c	642	16	26.2	127514	26	AC004417	Homo sapiens BAC clone 2.10e+02
	643	16	26.2	127766	27	AC005723	Homo sapiens chromosom 2.10e+02
c	644	16	26.2	128169	26	HS849L7	Human DNA sequence fro 2.10e+02
	645	16	26.2	129284	17	AC004892	*** SEQUENCING IN PROG 2.10e+02
	646	16	26.2	129914	26	AC003070	Homo sapiens chromosom 2.10e+02
c	647	16	26.2	130467	27	HS798A17	Human DNA sequence *** 2.10e+02
	648	16	26.2	130982	17	HS864I18	Human DNA sequence *** 2.10e+02
	649	16	26.2	131070	27	AC005295	Homo sapiens Xp22 bins 2.10e+02
c	650	16	26.2	131078	27	HSU80017	Homo sapiens basic tra 2.10e+02
	651	16	26.2	131673	17	HS336H9	Human DNA sequence *** 2.10e+02
c	652	16	26.2	131855	17	AC002518	*** SEQUENCING IN PROG 2.10e+02
	653	16	26.2	131903	26	HS360E18	Human DNA sequence fro 2.10e+02
c	654	16	26.2	132090	28	AC005402	Mus musculus clone UWG 2.10e+02

	655	16	26.2	133859	15	D90899	Synechocystis sp. PCC6 2.10e+02
	656	16	26.2	135352	17	HS354J5	Human DNA sequence *** 2.10e+02
c	657	16	26.2	135556	27	HSU85198	Homo sapiens BAC810, c 2.10e+02
	658	16	26.2	135928	27	HS434O14	Homo sapiens DNA seque 2.10e+02
	659	16	26.2	139259	18	CEY51A2D	Caenorhabditis elegans 2.10e+02
	660	16	26.2	140679	17	HS246H3	Human DNA sequence *** 2.10e+02
	661	16	26.2	141752	17	HS971N18	Human DNA sequence *** 2.10e+02
	662	16	26.2	143068	26	HSU95626	Homo sapiens ccr2b (cc 2.10e+02
c	663	16	26.2	143172	23	CLEGCGA	E.gracilis chloroplast 2.10e+02
c	664	16	26.2	146563	17	HS417M14	Human DNA sequence *** 2.10e+02
	665	16	26.2	146756	17	AC002348	*** SEQUENCING IN PROG 2.10e+02
	666	16	26.2	146963	26	HS964D12	Homo sapiens DNA seque 2.10e+02
	667	16	26.2	147806	17	HS111F4	Human DNA sequence *** 2.10e+02
	668	16	26.2	148416	26	HSAC002089	Human BAC clone RG308B 2.10e+02
c	669	16	26.2	150140	17	AC005478	*** SEQUENCING IN PROG 2.10e+02
c	670	16	26.2	151018	17	PFMAL1P6	Plasmodium falciparum 2.10e+02
c	671	16	26.2	151429	26	AC005279	Homo sapiens chromosom 2.10e+02
	672	16	26.2	151770	26	HSAC002066	Human BAC clone RG030H 2.10e+02
c	673	16	26.2	153479	17	AC005093	*** SEQUENCING IN PROG 2.10e+02
c	674	16	26.2	154848	26	AC002992	Homo sapiens chromosom 2.10e+02
c	675	16	26.2	155164	26	AC005165	Homo sapiens BAC clone 2.10e+02
	676	16	26.2	155821	18	AC005733	Drosophila melanogaste 2.10e+02
	677	16	26.2	157248	26	AC002312	Human Chromosome 15 pa 2.10e+02
c	678	16	26.2	157970	27	AC005609	Homo sapiens chromosom 2.10e+02
c	679	16	26.2	158676	17	AC005684	*** SEQUENCING IN PROG 2.10e+02
	680	16	26.2	159500	28	AC005742	Mus musculus chromosom 2.10e+02
	681	16	26.2	161453	27	AC004993	Homo sapiens PAC clone 2.10e+02
c	682	16	26.2	161973	26	HUAC002038	Homo sapiens chromosom 2.10e+02
c	683	16	26.2	162575	17	AC004086	*** SEQUENCING IN PROG 2.10e+02
	684	16	26.2	164074	17	AC005803	*** SEQUENCING IN PROG 2.10e+02
	685	16	26.2	164237	17	AC004838	*** SEQUENCING IN PROG 2.10e+02
c	686	16	26.2	164237	17	AC004838	*** SEQUENCING IN PROG 2.10e+02
c	687	16	26.2	166715	26	HS30801	Human genomic DNA sequ 2.10e+02
c	688	16	26.2	166983	26	AC002541	Human BAC clone RG043K 2.10e+02
c	689	16	26.2	167468	17	AC005033	*** SEQUENCING IN PROG 2.10e+02
	690	16	26.2	170760	17	AC004071	*** SEQUENCING IN PROG 2.10e+02
c	691	16	26.2	172151	25	HSAC000114	Human PAC clone DJ527C 2.10e+02
c	692	16	26.2	173452	26	AC003013	Human PAC clone DJ0205 2.10e+02
	693	16	26.2	173882	26	HSAF001550	Homo sapiens chromosom 2.10e+02
	694	16	26.2	174826	26	HUU91324	Homo sapiens Chromosom 2.10e+02
	695	16	26.2	177189	17	HS428A13	*** SEQUENCING IN PROG 2.10e+02
c	696	16	26.2	177189	17	HS428A13	*** SEQUENCING IN PROG 2.10e+02
	697	16	26.2	177328	17	HS462D8	Human DNA sequence *** 2.10e+02
c	698	16	26.2	180114	17	HS52D1	Human DNA sequence *** 2.10e+02
c	699	16	26.2	180388	25	HUMRETBLAS	Human retinoblastoma s 2.10e+02
	700	16	26.2	180527	17	HS269M15	Human DNA sequence *** 2.10e+02
	701	16	26.2	181098	26	AC004216	Homo sapiens 12q24 PAC 2.10e+02
c	702	16	26.2	181221	17	AC004835	*** SEQUENCING IN PROG 2.10e+02
c	703	16	26.2	184130	26	AC005209	Homo sapiens chromosom 2.10e+02
c	704	16	26.2	184949	26	AC005029	Homo sapiens BAC clone 2.10e+02
	705	16	26.2	185775	25	HUMFMR1S	Homo sapiens fragile X 2.10e+02

	706	16	26.2	189009	17	AC005746	*** SEQUENCING IN PROG 2.10e+02
c	707	16	26.2	189279	27	AC005028	Homo sapiens BAC clone 2.10e+02
c	708	16	26.2	189357	17	HS596H12	Human DNA sequence *** 2.10e+02
c	709	16	26.2	190000	17	AC005598	*** SEQUENCING IN PROG 2.10e+02
c	710	16	26.2	191439	17	AC005040	*** SEQUENCING IN PROG 2.10e+02
c	711	16	26.2	194020	26	HUAC004125	Homo sapiens Chromosom 2.10e+02
c	712	16	26.2	194252	17	HS105D16	Human DNA sequence *** 2.10e+02
	713	16	26.2	194985	28	AC002406	Mouse chromosome X BAC 2.10e+02
	714	16	26.2	195040	17	AC005235	*** SEQUENCING IN PROG 2.10e+02
c	715	16	26.2	196836	17	AC004836	*** SEQUENCING IN PROG 2.10e+02
c	716	16	26.2	199927	17	HS101A4	Human DNA sequence *** 2.10e+02
	717	16	26.2	201964	28	MMMHC29N7	Mus musculus major his 2.10e+02
c	718	16	26.2	201990	26	AC002385	Human BAC clone RG222A 2.10e+02
	719	16	26.2	205150	26	AC005274	Homo sapiens chromosom 2.10e+02
c	720	16	26.2	206880	26	AF020503	Homo sapiens PRA3B com 2.10e+02
c	721	16	26.2	207931	17	HS101G11	Human DNA sequence *** 2.10e+02
	722	16	26.2	208396	26	AC003101	Homo sapiens chromosom 2.10e+02
c	723	16	26.2	225035	17	AC004949	*** SEQUENCING IN PROG 2.10e+02
	724	16	26.2	226835	17	CEY49A3	Caenorhabditis elegans 2.10e+02
c	725	16	26.2	227328	17	HS323M22	Human DNA sequence *** 2.10e+02
	726	16	26.2	227538	28	AC002397	Mouse chromosome 6 BAC 2.10e+02
	727	16	26.2	236281	26	AC004673	Homo sapiens Xp22 BAC 2.10e+02
	728	16	26.2	238438	17	AC005046	*** SEQUENCING IN PROG 2.10e+02
c	729	16	26.2	244455	17	AC002470	*** SEQUENCING IN PROG 2.10e+02
	730	16	26.2	245439	28	MMHC135G15	Mus musculus major his 2.10e+02
c	731	16	26.2	247774	17	AC004803	*** SEQUENCING IN PROG 2.10e+02
c	732	16	26.2	247877	26	HUAE000661	Homo sapiens T-cell re 2.10e+02
c	733	16	26.2	251199	26	HUAE000659	Homo sapiens T-cell re 2.10e+02
c	734	16	26.2	257867	18	AC005557	Drosophila melanogaste 2.10e+02
c	735	16	26.2	259474	26	HUAC004605	Homo sapiens Chromosom 2.10e+02
c	736	16	26.2	275138	17	CEY54E5	Caenorhabditis elegans 2.10e+02
c	737	16	26.2	286715	17	CEY45F3	Caenorhabditis elegans 2.10e+02
c	738	16	26.2	307360	17	HS796E4	Human DNA sequence *** 2.10e+02
	739	16	26.2	313573	18	CEY57G11C	Caenorhabditis elegans 2.10e+02
c	740	16	26.2	313573	18	CEY57G11C	Caenorhabditis elegans 2.10e+02
c	741	16	26.2	323792	17	AC004965	*** SEQUENCING IN PROG 2.10e+02
	742	16	26.2	328808	17	CEY44A6	Caenorhabditis elegans 2.10e+02
	743	16	26.2	329751	17	HS407F11	Human DNA sequence *** 2.10e+02
	744	16	26.2	331519	17	CEY39B6	Caenorhabditis elegans 2.10e+02
	745	15	24.6	222	25	HSDNASAAI	H.sapiens DNA sequence 6.82e+02
	746	15	24.6	255	33	HIVV3ACE	Human immunodeficiency 6.82e+02
c	747	15	24.6	278	33	HJV02	highlands j virus rna, 6.82e+02
	748	15	24.6	283	33	HIVU49589	Human immunodeficiency v 6.82e+02
	749	15	24.6	288	33	HIVED87DP	Human immunodeficiency 6.82e+02
	750	15	24.6	300	33	ALEEEN	Equine encephalitis vi 6.82e+02
	751	15	24.6	315	33	AF023381	HIV-1 patient 196, sam 6.82e+02
	752	15	24.6	324	33	AF016563	HIV-1 strain HI66c1/1 6.82e+02
	753	15	24.6	328	33	HIVY14419	HIV-1 virus 93CUPL05 g 6.82e+02
	754	15	24.6	332	33	ALEEES	3' terminal sequence o 6.82e+02
	755	15	24.6	351	33	ALWEEX	3' terminal sequence o 6.82e+02
c	756	15	24.6	351	33	EEW3E	western equine encepha 6.82e+02

c	757	15	24.6	370	28	MMTPM1	M.musculus Tpm1 gene f	6.82e+02
c	758	15	24.6	390	30	HSPE05A03	H.sapiens flow-sorted	6.82e+02
c	759	15	24.6	467	30	G27176	human STS SHGC-30049.	6.82e+02
	760	15	24.6	503	25	HSPA17G1	H.sapiens flow-sorted	6.82e+02
	761	15	24.6	558	30	G39210	E17294 Zebrafish AB Da	6.82e+02
c	762	15	24.6	631	30	HUM48TS384	Human chromosome 4 seq	6.82e+02
	763	15	24.6	666	33	AF041119	HIV-1 isolate L6099 fr	6.82e+02
	764	15	24.6	675	15	BSSFPG	B.subtilis sfp gene.	6.82e+02
	765	15	24.6	722	33	HIVU69378	HIV-1 patient E, sampl	6.82e+02
	766	15	24.6	727	27	AF036342	Homo sapiens cystatin	6.82e+02
	767	15	24.6	734	20	HBU24672	Heterixalus boettgeri	6.82e+02
	768	15	24.6	867	27	AF031824	Homo sapiens leukocyst	6.82e+02
	769	15	24.6	924	33	HIV1U19694	Human immunodeficiency	6.82e+02
c	770	15	24.6	1095	18	BCU05814	Blastocrithidia culici	6.82e+02
c	771	15	24.6	1136	28	AF044279	Rattus norvegicus stra	6.82e+02
	772	15	24.6	1233	21	I36744	Sequence 4 from patent	6.82e+02
c	773	15	24.6	1333	23	GMU47143	Glycine max nonsymbiot	6.82e+02
	774	15	24.6	1400	21	I25908	Sequence 1 from patent	6.82e+02
c	775	15	24.6	1404	20	S76875	casein kinase 2 alpha	6.82e+02
	776	15	24.6	1409	18	S61734S2	omb (biD4)=optomotor-b	6.82e+02
	777	15	24.6	1503	33	HIV1U29694	Human immunodeficiency	6.82e+02
c	778	15	24.6	1561	28	RNU50447	Rattus norvegicus clon	6.82e+02
	779	15	24.6	1610	27	HSU31628	Human interleukin-15 r	6.82e+02
c	780	15	24.6	1619	28	RATMHCIMRA	Rattus norvegicus MHC	6.82e+02
c	781	15	24.6	1788	24	AF052405	Pandora neoaphidis sma	6.82e+02
	782	15	24.6	1905	23	VFLEB6	Vicia faba VfLEB6 gene	6.82e+02
c	783	15	24.6	1906	23	EAU35393	Entomophaga aulicae sp	6.82e+02
	784	15	24.6	1937	18	DMMIPMY	D.melanogaster mRNA fo	6.82e+02
c	785	15	24.6	1978	23	DDIDDKA	D.discoideum protein k	6.82e+02
c	786	15	24.6	2047	18	BMERB2DNA	B.mori ErB.2 gene for	6.82e+02
	787	15	24.6	2098	18	DDSP60	Dictyostelium discoide	6.82e+02
	788	15	24.6	2252	28	AF047716	Mus musculus A-kinase	6.82e+02
	789	15	24.6	2286	15	BPTETAQ3	B.fragilis tetA(Q)3 ge	6.82e+02
c	790	15	24.6	2647	23	SVGMAF	S.vulgare gene for gam	6.82e+02
	791	15	24.6	2657	15	BTTETQG	B.thetaiotaomicron tet	6.82e+02
c	792	15	24.6	2696	23	DDIPRRNGPK	Dictyostelium discoide	6.82e+02
c	793	15	24.6	2881	28	RATCBPTP	Rat mRNA for tyrosine	6.82e+02
c	794	15	24.6	2955	33	AF020660	Tomato spotted wilt vi	6.82e+02
	795	15	24.6	2984	18	TRBANTA	Trypanosoma cruzi anti	6.82e+02
	796	15	24.6	3354	23	YSCAMD	Yeast (S.cerevisiae) A	6.82e+02
c	797	15	24.6	3429	19	SSATPAR	Pig mRNA for (Na+, K+)	6.82e+02
	798	15	24.6	3510	26	HSAC001053	Homo sapiens (subclone	6.82e+02
c	799	15	24.6	3722	33	NPHRLGE	A.californica nuclear	6.82e+02
c	800	15	24.6	3723	28	MUSALCR05	Mus musculus alphaA-cr	6.82e+02
c	801	15	24.6	3726	26	HSL81575	Homo sapiens (subclone	6.82e+02
	802	15	24.6	3842	15	BBU45425	Borrelia burgdorferi 2	6.82e+02
c	803	15	24.6	3928	20	CHKAFTROP4	Chicken tropomyosin ge	6.82e+02
c	804	15	24.6	4144	33	ZEQPOLY	Eastern equine encepha	6.82e+02
c	805	15	24.6	4207	23	SCYDR074W	S.cerevisiae chromosom	6.82e+02
c	806	15	24.6	4259	33	TOEEE26S	Eastern equine encepha	6.82e+02
c	807	15	24.6	4472	15	HPU59625	Helicobacter pylori co	6.82e+02

c	808	15	24.6	4490	26	AF009242	Homo sapiens proline-r	6.82e+02
	809	15	24.6	4573	25	HUM32DC97%	Homo sapiens (subclone	6.82e+02
c	810	15	24.6	4654	28	AF025309	Rattus norvegicus MHC	6.82e+02
	811	15	24.6	4937	18	DDU64830	Dictyostelium discoide	6.82e+02
	812	15	24.6	6237	28	MUSMHP36BG	Mouse MHC class I tum-	6.82e+02
c	813	15	24.6	6746	23	ASTPHT3A	Avena sativa type 3 ph	6.82e+02
	814	15	24.6	7149	18	PFSC03031	Plasmodium falciparum	6.82e+02
c	815	15	24.6	7544	18	DMXDH	D. melanogaster Xdh ge	6.82e+02
c	816	15	24.6	7575	18	PFSC03082	Plasmodium falciparum	6.82e+02
c	817	15	24.6	8607	33	PAVCGS	Canine oral papillomav	6.82e+02
	818	15	24.6	8892	25	HUMMHA3C7B	Homo sapiens MHC class	6.82e+02
	819	15	24.6	10168	18	PFSC03091	Plasmodium falciparum	6.82e+02
c	820	15	24.6	10332	15	HPAE000527	Helicobacter pylori se	6.82e+02
	821	15	24.6	10798	23	LEBFRUCG	L.esculentum gene for	6.82e+02
	822	15	24.6	10965	21	I29074	Sequence 4 from patent	6.82e+02
c	823	15	24.6	11678	33	U01034	Eastern equine encepha	6.82e+02
c	824	15	24.6	11763	28	MUSBA	Mouse gene for basigin	6.82e+02
c	825	15	24.6	12054	18	DRODACHSOU	Drosophila melanogaste	6.82e+02
	826	15	24.6	12206	16	AF000366	Borrelia burgdorferi o	6.82e+02
	827	15	24.6	12222	19	BOVLYSOZMA	Bos taurus lysozyme ge	6.82e+02
c	828	15	24.6	12590	18	CELC04F2	Caenorhabditis elegans	6.82e+02
c	829	15	24.6	13133	23	ALRIBP	A.longa plastid genes	6.82e+02
	830	15	24.6	13682	15	YETRSATOH	Y.enterocolitica 6471/	6.82e+02
c	831	15	24.6	13747	18	AF015193	Onchocerca volvulus mi	6.82e+02
	832	15	24.6	14477	18	DMD02279	Drosophila melanogaste	6.82e+02
c	833	15	24.6	15459	18	CEK02B7	Caenorhabditis elegans	6.82e+02
c	834	15	24.6	20048	18	CEC33A11	Caenorhabditis elegans	6.82e+02
	835	15	24.6	21011	18	CELF12A10	Caenorhabditis elegans	6.82e+02
	836	15	24.6	21643	18	CEK07C10	Caenorhabditis elegans	6.82e+02
c	837	15	24.6	22429	18	CELF58F12	Caenorhabditis elegans	6.82e+02
	838	15	24.6	23016	18	TRBKPGEN	Trypanosoma brucei bru	6.82e+02
	839	15	24.6	25593	18	CEH21P03	Caenorhabditis elegans	6.82e+02
c	840	15	24.6	26511	18	CET07D4	Caenorhabditis elegans	6.82e+02
c	841	15	24.6	26790	18	CEC14H10	Caenorhabditis elegans	6.82e+02
c	842	15	24.6	27809	18	CEF38H4	Caenorhabditis elegans	6.82e+02
c	843	15	24.6	28678	18	CELT04D1	Caenorhabditis elegans	6.82e+02
	844	15	24.6	28883	18	CELZC395	Caenorhabditis elegans	6.82e+02
c	845	15	24.6	29039	18	CEF56F3	Caenorhabditis elegans	6.82e+02
	846	15	24.6	29132	18	CEF10B5	Caenorhabditis elegans	6.82e+02
c	847	15	24.6	29404	18	CEF55C9	Caenorhabditis elegans	6.82e+02
	848	15	24.6	30320	18	CEC05B5	Caenorhabditis elegans	6.82e+02
c	849	15	24.6	30392	18	CEZK1037	Caenorhabditis elegans	6.82e+02
	850	15	24.6	32433	18	CELF08B1	Caenorhabditis elegans	6.82e+02
	851	15	24.6	32632	18	CEF07A5	Caenorhabditis elegans	6.82e+02
c	852	15	24.6	34500	23	SC8337	S.cerevisiae chromosom	6.82e+02
c	853	15	24.6	34831	18	CEC47G2	Caenorhabditis elegans	6.82e+02
	854	15	24.6	35253	18	CELC24A8	Caenorhabditis elegans	6.82e+02
	855	15	24.6	35409	18	CEE02H4	Caenorhabditis elegans	6.82e+02
	856	15	24.6	35727	18	CEW09G3	Caenorhabditis elegans	6.82e+02
c	857	15	24.6	35727	18	CEW09G3	Caenorhabditis elegans	6.82e+02
	858	15	24.6	35875	18	CEZC15	Caenorhabditis elegans	6.82e+02

	859	15	24.6	36256	18	CEF38A6	Caenorhabditis elegans	6.82e+02
c	860	15	24.6	36532	18	CEF57B1	Caenorhabditis elegans	6.82e+02
c	861	15	24.6	36808	17	CEH36F17	Caenorhabditis elegans	6.82e+02
c	862	15	24.6	36958	18	CEF11E6	Caenorhabditis elegans	6.82e+02
	863	15	24.6	36974	18	CEK01D12	Caenorhabditis elegans	6.82e+02
	864	15	24.6	37261	18	CEC17H1	Caenorhabditis elegans	6.82e+02
	865	15	24.6	37614	18	CELT10H10	Caenorhabditis elegans	6.82e+02
c	866	15	24.6	37963	17	AC003094	*** SEQUENCING IN PROG	6.82e+02
c	867	15	24.6	38159	18	CEC54E10	Caenorhabditis elegans	6.82e+02
c	868	15	24.6	39073	23	SPAC2E11	S.pombe chromosome I c	6.82e+02
c	869	15	24.6	39090	26	D86996	Human (lambda) DNA for	6.82e+02
	870	15	24.6	39216	18	CEB0250	Caenorhabditis elegans	6.82e+02
c	871	15	24.6	39599	18	CEM01G12	Caenorhabditis elegans	6.82e+02
c	872	15	24.6	39757	18	CEC04C11	Caenorhabditis elegans	6.82e+02
	873	15	24.6	39925	18	CEK06B4	Caenorhabditis elegans	6.82e+02
	874	15	24.6	39962	27	AC005616	Homo sapiens chromosom	6.82e+02
	875	15	24.6	40065	23	SPBC16E9	S.pombe chromosome II	6.82e+02
c	876	15	24.6	40876	18	CELM02D8	Caenorhabditis elegans	6.82e+02
	877	15	24.6	40909	18	CEB0464	Caenorhabditis elegans	6.82e+02
	878	15	24.6	41100	24	SPAC26H5	S.pombe chromosome I c	6.82e+02
c	879	15	24.6	41277	27	AC004180	Homo sapiens clone UWG	6.82e+02
c	880	15	24.6	41345	18	CEF15G9	Caenorhabditis elegans	6.82e+02
	881	15	24.6	41588	18	CELD1014	Caenorhabditis elegans	6.82e+02
	882	15	24.6	41615	27	AC005625	Homo sapiens chromosom	6.82e+02
c	883	15	24.6	41645	18	CELC03H5	Caenorhabditis elegans	6.82e+02
c	884	15	24.6	41811	18	CELR05G6	Caenorhabditis elegans	6.82e+02
c	885	15	24.6	42393	27	AC005780	Homo sapiens chromosom	6.82e+02
c	886	15	24.6	43555	18	CEC40H5	Caenorhabditis elegans	6.82e+02
	887	15	24.6	44318	18	CELB0273	Caenorhabditis elegans	6.82e+02
	888	15	24.6	44380	15	AE001137	Borrelia burgdorferi (	6.82e+02
	889	15	24.6	44406	26	AC005401	Homo sapiens clone UWG	6.82e+02
c	890	15	24.6	45125	18	CELB0348	Caenorhabditis elegans	6.82e+02
	891	15	24.6	48689	24	T2H3	Arabidopsis thaliana B	6.82e+02
c	892	15	24.6	55186	18	AC001655	Drosophila melanogaste	6.82e+02
	893	15	24.6	65691	18	PFMAL3P1	Plasmodium falciparum	6.82e+02
	894	15	24.6	68940	26	AF019664	Homo sapiens cosmids 5	6.82e+02
c	895	15	24.6	71889	17	AC004325	*** SEQUENCING IN PROG	6.82e+02
	896	15	24.6	72119	23	SCD9717	Saccharomyces cerevisi	6.82e+02
	897	15	24.6	78661	26	AC005350	Homo sapiens chromosom	6.82e+02
c	898	15	24.6	82212	23	ATAC004484	Arabidopsis thaliana c	6.82e+02
	899	15	24.6	83371	24	ATM4I22	Arabidopsis thaliana D	6.82e+02
c	900	15	24.6	83650	17	AC004311	*** SEQUENCING IN PROG	6.82e+02
	901	15	24.6	83675	27	AC004830	Homo sapiens PAC clone	6.82e+02
	902	15	24.6	85641	26	AC005217	Homo sapiens chromosom	6.82e+02
	903	15	24.6	86004	27	HS51V23	Human DNA sequence fro	6.82e+02
c	904	15	24.6	86567	18	AC004276	Drosophila melanogaste	6.82e+02
c	905	15	24.6	86710	23	ATF23E12	Arabidopsis thaliana D	6.82e+02
c	906	15	24.6	87034	26	AC004587	Human Chromosome 15q26	6.82e+02
	907	15	24.6	87114	26	AC005353	Homo sapiens chromosom	6.82e+02
c	908	15	24.6	89840	24	ATAC004705	Arabidopsis thaliana c	6.82e+02
c	909	15	24.6	92802	17	AC004710	*** SEQUENCING IN PROG	6.82e+02

c	910	15	24.6	92944	17	HSU242P8	Human DNA sequence ***	6.82e+02
c	911	15	24.6	93060	24	ATAC004077	Arabidopsis thaliana c	6.82e+02
	912	15	24.6	94038	17	HS246D7	Human DNA sequence ***	6.82e+02
	913	15	24.6	94319	26	HS341I10	Human DNA sequence fro	6.82e+02
c	914	15	24.6	95235	17	HS703H14	Human DNA sequence ***	6.82e+02
c	915	15	24.6	95519	24	AF071527	Arabidopsis thaliana B	6.82e+02
c	916	15	24.6	97494	24	T15F16	Arabidopsis thaliana B	6.82e+02
c	917	15	24.6	99359	23	ATAC003673	Arabidopsis thaliana c	6.82e+02
c	918	15	24.6	100000	26	AP000010	Homo sapiens genomic D	6.82e+02
	919	15	24.6	101371	23	AC002534	Arabidopsis thaliana B	6.82e+02
c	920	15	24.6	101486	17	AC005228	*** SEQUENCING IN PROG	6.82e+02
	921	15	24.6	102473	17	AC002320	*** SEQUENCING IN PROG	6.82e+02
	922	15	24.6	102591	26	AC002978	Homo sapiens 12q24 PAC	6.82e+02
c	923	15	24.6	102846	27	AC002483	Homo sapiens PAC clone	6.82e+02
c	924	15	24.6	106696	26	HUAC003661	Homo sapiens Chromosom	6.82e+02
c	925	15	24.6	107183	23	AC002983	A. thaliana BAC T3F12	6.82e+02
	926	15	24.6	110000	17	CEY53H1	Caenorhabditis elegans	6.82e+02
	927	15	24.6	110000	17	CEY43F8	Caenorhabditis elegans	6.82e+02
c	928	15	24.6	110000	17	CEY51A2	Caenorhabditis elegans	6.82e+02
c	929	15	24.6	110000	17	CEY60A9	Caenorhabditis elegans	6.82e+02
	930	15	24.6	114231	27	HS390C10	Homo sapiens DNA seque	6.82e+02
c	931	15	24.6	117478	17	CEY69H2	Caenorhabditis elegans	6.82e+02
c	932	15	24.6	117752	26	AC004743	Homo sapiens PAC clone	6.82e+02
c	933	15	24.6	118610	17	AC002344	*** SEQUENCING IN PROG	6.82e+02
	934	15	24.6	118684	26	AC004259	Human Chromosome 15q11	6.82e+02
c	935	15	24.6	120452	27	AC005753	Homo sapiens chromosom	6.82e+02
c	936	15	24.6	122121	26	AC004543	Homo sapiens BAC clone	6.82e+02
	937	15	24.6	122809	17	HS881E24	Human DNA sequence ***	6.82e+02
	938	15	24.6	127247	25	HS326L13	Human DNA sequence fro	6.82e+02
c	939	15	24.6	127356	17	AC005063	*** SEQUENCING IN PROG	6.82e+02
	940	15	24.6	129295	17	HS48A11	Human DNA sequence ***	6.82e+02
	941	15	24.6	130020	26	HUAC004525	Homo sapiens Chromosom	6.82e+02
c	942	15	24.6	132070	26	AC003663	Homo sapiens chromosom	6.82e+02
c	943	15	24.6	132783	26	AC000378	Human Chromosome 11 pa	6.82e+02
c	944	15	24.6	134478	17	HS154A10	Human DNA sequence ***	6.82e+02
	945	15	24.6	135686	26	HS473J6	Human DNA sequence fro	6.82e+02
	946	15	24.6	136746	17	AC004928	*** SEQUENCING IN PROG	6.82e+02
c	947	15	24.6	137662	17	CEY37C11	Caenorhabditis elegans	6.82e+02
c	948	15	24.6	139904	26	HS97D16	Homo sapiens DNA seque	6.82e+02
c	949	15	24.6	139958	27	HUAC002519	Human Chromosome 16 BA	6.82e+02
	950	15	24.6	140704	17	AC004916	*** SEQUENCING IN PROG	6.82e+02
c	951	15	24.6	141460	17	AC004829	*** SEQUENCING IN PROG	6.82e+02
	952	15	24.6	142640	17	AC004936	*** SEQUENCING IN PROG	6.82e+02
	953	15	24.6	143427	26	AC002432	Human BAC clone RG317G	6.82e+02
	954	15	24.6	146500	26	AC004453	Homo sapiens PAC clone	6.82e+02
c	955	15	24.6	146963	26	HS964D12	Homo sapiens DNA seque	6.82e+02
c	956	15	24.6	147135	26	HS368A4	Human DNA sequence fro	6.82e+02
c	957	15	24.6	147476	17	AC004157	*** SEQUENCING IN PROG	6.82e+02
c	958	15	24.6	149687	27	AC004981	Homo sapiens PAC clone	6.82e+02
c	959	15	24.6	151855	17	HS191J18	Human DNA sequence ***	6.82e+02
	960	15	24.6	152392	17	HS1018D12	Human DNA sequence ***	6.82e+02

```
    961      15    24.6 153692 17  CEY26D4      Caenorhabditis elegans  6.82e+02
c   962      15    24.6 153836 27  AC005344     Homo sapiens 12p13.3 P  6.82e+02
    963      15    24.6 154152 17  AC005057     *** SEQUENCING IN PROG  6.82e+02
    964      15    24.6 156399 26  AC002456     Human BAC clone RG013L  6.82e+02
c   965      15    24.6 159500 28  AC005742     Mus musculus chromosom  6.82e+02
c   966      15    24.6 160048 28  AC002109     Genomic sequence from   6.82e+02
c   967      15    24.6 160531 17  HS127F18     Human DNA sequence ***  6.82e+02
c   968      15    24.6 163023 17  AC004815     *** SEQUENCING IN PROG  6.82e+02
c   969      15    24.6 163908 27  AC005411     Homo sapiens chromosom  6.82e+02
    970      15    24.6 164536 17  HS134N8      Human DNA sequence ***  6.82e+02
c   971      15    24.6 165015 17  HS1158E12    Human DNA sequence ***  6.82e+02
    972      15    24.6 165612 17  HS545L17     Human DNA sequence ***  6.82e+02
    973      15    24.6 169998 26  HS279N11     Human DNA sequence fro  6.82e+02
c   974      15    24.6 170000 17  AC004066     *** SEQUENCING IN PROG  6.82e+02
c   975      15    24.6 170128 17  AC005059     *** SEQUENCING IN PROG  6.82e+02
c   976      15    24.6 170215 17  AC004127     *** SEQUENCING IN PROG  6.82e+02
    977      15    24.6 173970 18  AC005286     Drosophila melanogaste  6.82e+02
c   978      15    24.6 174076 17  AC005743     *** SEQUENCING IN PROG  6.82e+02
c   979      15    24.6 178098 26  HS168L15     Homo sapiens DNA seque  6.82e+02
    980      15    24.6 184949 26  AC005029     Homo sapiens BAC clone  6.82e+02
c   981      15    24.6 186550 17  CEY44F5      Caenorhabditis elegans  6.82e+02
c   982      15    24.6 190482 26  AC004106     Homo sapiens Xp22 BAC   6.82e+02
c   983      15    24.6 193161 27  AC005500     , complete sequence.    6.82e+02
c   984      15    24.6 195033 17  HS608E8      Human DNA sequence ***  6.82e+02
    985      15    24.6 195364 26  HS431A14     Human DNA sequence fro  6.82e+02
c   986      15    24.6 196405 17  AC005053     *** SEQUENCING IN PROG  6.82e+02
c   987      15    24.6 196473 17  HS473B4      Human DNA sequence ***  6.82e+02
c   988      15    24.6 199878 17  AC005027     *** SEQUENCING IN PROG  6.82e+02
    989      15    24.6 207730 15  BSUB0011     Bacillus subtilis comp  6.82e+02
    990      15    24.6 209510 15  BSUB0018     Bacillus subtilis comp  6.82e+02
c   991      15    24.6 224525 17  CEY70C5      Caenorhabditis elegans  6.82e+02
    992      15    24.6 227328 17  HS323M22     Human DNA sequence ***  6.82e+02
    993      15    24.6 227892 17  AC005587     *** SEQUENCING IN PROG  6.82e+02
    994      15    24.6 230023 17  HS522P13     Human DNA sequence ***  6.82e+02
c   995      15    24.6 250448 17  CEY36E3      Caenorhabditis elegans  6.82e+02
c   996      15    24.6 250611 28  MMAE000663   Mus musculus TCR beta   6.82e+02
    997      15    24.6 253505 16  AP000007     Pyrococcus horikoshii   6.82e+02
c   998      15    24.6 269078 17  CEY76A2      Caenorhabditis elegans  6.82e+02
c   999      15    24.6 276261 17  HSAC000406   *** SEQUENCING IN PROG  6.82e+02
c  1000      15    24.6 281873 17  AC005533     *** SEQUENCING IN PROG  6.82e+02
```

ALIGNMENTS

```
RESULT    1
LOCUS       HUAC002425 113241 bp    DNA             PRI       10-MAR-1998
DEFINITION  Homo sapiens Chromosome 16 BAC clone CIT987SK-A-575C2, complete
            sequence.
ACCESSION   AC002425
NID         g2947051
KEYWORDS    HTG.
```



```
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 113241)
  AUTHORS   Adams,M.D., Loftus,B.J., Zhou,L., Crosby,M., Fuhrmann,J.,
            Mason,T.M., Brandon,R., Kim,U.J., Kerlavage,A.R. and Venter,J.C.
  TITLE     Homo sapiens Chromosome 16 BAC clone CIT987SK-A-575C2
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 113241)
  AUTHORS   Adams,M.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-AUG-1997) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
REFERENCE   3  (bases 1 to 113241)
  AUTHORS   Adams,M.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAR-1998) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
COMMENT     On Mar 10, 1998 this sequence version replaced gi:2924794.
            Address all correspondence to: Mark Adams The Institute for Genomic
            Research 9712 Medical Center Dr,
            Rockville, MD 20850, USA e-mail address: mdadams@tigr.org.  The bac
            location is on chromosome .   The orientation of the sequence is
            from SP6 end to T7 end. Genes were identified by a combination of
            five methods including: XGRAIL (available by anonymous ftp from
            arthur.epm.ornl.gov), Genefinder (Phil Green, University of
            Washington), Genscan  (Chris Burge,
            http://gnomic.stanford.Edu/~chris/GENSCANW.html)searches of the
            complete sequence against a peptide database,and the Human gene
            Index database at TIGR (http://www.tigr.org/tdb/hgi/hgi.html).
            Genes without pepetide homolgy having spliced EST hits are termed
            'Unknown gene product'. Genes encoding tRNAs are predicted by
            tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
FEATURES                Location/Qualifiers
     source             1. .113241
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="16"
                        /map="16p11.2"
                        /clone="A-575C2"
     mRNA               join(35149. .35370,35957. .36171)
                        /gene="A-575C2.2"
     mRNA               join(35149. .35370,36011. .36160,36497. .36825)
                        /gene="A-575C2.2"
     gene               35149. .36825
                        /gene="A-575C2.2"
     CDS                join(35259. .35370,36011. .36147)
                        /gene="A-575C2.2"
                        /codon_start=1
                        /product="Gene product with similarity to Rat P8"
```



```
                        /db_xref="PID:g2947054"
                        /translation="MATFPPATSAPQQPPGPEDEDSSLDESDLYSLAHSYLGGGGRKG
                        RTKREAAANTNRPSPGGHERKLVTKLQNSERKKRGARR"
     CDS                join(35259. .35370,35957. .36147)
                        /gene="A-575C2.2"
                        /codon_start=1
                        /product="Gene product with similarity to Rat P8"
                        /db_xref="PID:g2947053"
                        /translation="MATFPPATSAPQQPPGPEDEDSSLDESDLYSLAHSYLGPLIMPM
                        PTSPLTPALVTGGGGRKGRTKREAAANTNRPSPGGHERKLVTKLQNSERKKRGARR"
     mRNA               join(82211. .82433,82587. .82665,84762. .84858,85486. .85557,
                        86407. .86486,86607. .86692,87498. .87570,87658. .87801,
                        90030. .90142,91476. .91522,91603. .91671,91766. .91821,
                        91950. .92043,96271. .96411,96513. .96868)
                        /gene="A-575C2.5"
     mRNA               join(82211. .82433,82587. .82665,84762. .84858,85486. .85557,
                        86407. .86486,86607. .86692,87498. .87570,87658. .87801,
                        90030. .90142,91476. .91522,91603. .91671,91766. .91821,
                        91950. .92043,96271. .96411,96513. .96684,107246. .107493)
                        /gene="A-575C2.5"
     gene               82211. .107493
                        /gene="A-575C2.5"
     CDS                join(82388. .82433,82587. .82665,84762. .84858,85486. .85557,
                        86407. .86486,86607. .86692,87498. .87570,87658. .87801,
                        90030. .90142,91476. .91522,91603. .91671,91766. .91821,
                        91950. .92043,96271. .96411,96513. .96632)
                        /gene="A-575C2.5"
                        /codon_start=1
                        /product="CLN3"
                        /db_xref="PID:g2947055"
                        /translation="MGGCAGSRRRFSDSEGEETVPEPRLPLLDHQGAHWKNAVGFWLL
                        CLCNNFSYVVWLSAAHDILSHKRTSGNQSHVDPGPTPIPHNSSSRFDCNSVSTAAVLL
                        ADILPTLVIKLLAPLGLHLLPYSPRVLVSGICAAGSFVLVAFSHSVGTSLCGVVFASI
                        SSGLGEVTFLSLTAFYPRAVISWWSSGTGGAGLLGALSYLGLTQAGLSPQQTLLSMLG
                        IPALLLASYFLLLTSPEAQDPGGEEEAESAARQPLIRTEAPESKPGSSSSLSLRERWT
                        VPKGLLWYIVPLVVVYFAEYFINQGLFELLPFWNTSLSHAQQYRWYQMLYQAGVFASR
                        SSLRCCRIRFTWALALLQCLNLVFLLADVWFGFLPSIYLVPLIILYEGLLGGAAYVNT
                        FHNIALETSDEHREFAMAATCISDTLGISLSGLLALPLHDFLCQLS"
     mRNA               join(111070. .111169,111282. .111481)
                        /gene="A-575C2.4"
     gene               111070. .111481
                        /gene="A-575C2.4"
     CDS                join(111121. .111169,111282. .111481)
                        /gene="A-575C2.4"
                        /codon_start=1
                        /product="Unknown gene product"
                        /db_xref="PID:g2947052"
                        /translation="MDVQQRAWRSNRSRQKGLRSICMHTKKRVSSFRGNKIGLKDVIT
                        LRRHVETKVRAKIRKREVTTKINRHDKINGKRKTARKQ"
BASE COUNT    27906 a  28459 c  27774 g  29102 t
ORIGIN
```



```
  Query Match           72.1%;  Score 44;  DB 26;  Length 113241;
  Best Local Similarity  97.8%;  Pred. No. 4.05e-17;
  Matches   44;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db  75184 TTCCCTCAGTCTTCCTTGCAGTCTTTATTGTTTTATTTTGTTTTT 75228
          |||| |||||||||||||||||||||||||||||||||||||||||
Cp      61 TTCCNTCAGTCTTCCTTGCAGTCTTTATTGTTTTATTTTGTTTTT 17


RESULT    2
LOCUS       HUAC002544 156641 bp    DNA             PRI       24-JUL-1998
DEFINITION  Homo sapiens Chromosome 16 BAC clone CIT987SK-A-761H5, complete
            sequence.
ACCESSION   AC002544
NID         g3337382
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 156641)
  AUTHORS   Adams,M.D., Loftus,B.J., Zhou,L., Crosby,M., Fuhrmann,J.,
            Mason,T.M., Brandon,R., Kim,U.J., Kerlavage,A.R. and Venter,J.C.
  TITLE     Homo sapiens Chromosome 16 BAC clone CIT987SK-A-761H5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 156641)
  AUTHORS   Adams,M.D. and Loftus,B.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-SEP-1997) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA, Email:
            mdadams@tigr.org
REFERENCE   3  (bases 1 to 156641)
  AUTHORS   Adams,M.D. and Loftus,B.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUL-1998) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
COMMENT     On Jul 24, 1998 this sequence version replaced gi:2959563.
            Address all correspondence to: Mark Adams The Institute for Genomic
            Research 9712 Medical Center Dr, Rockville, MD 20850, USA e-mail
            address: humgen@tigr.org.   The orientation of the sequence is from
            SP6 end to T7 end. Genes were identified by a combination of five
            methods including: XGRAIL (available by anonymous ftp from
            arthur.epm.ornl.gov), Genefinder (Phil Green, University of
            Washington), Genscan  (Chris Burge,
            http://gnomic.stanford.Edu/~chris/GENSCANW.html)searches of the
            complete sequence against a peptide database,and the Human gene
            Index database at TIGR (http://www.tigr.org/tdb/hgi/hgi.html).
            Genes without pepetide homolgy having spliced EST hits are termed
            'Unknown gene product'. Genes encoding tRNAs are predicted by
            tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
```

```
FEATURES             Location/Qualifiers
     source          1. .156641
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /map="16p11.2"
                     /clone="A-761H5"
     STS             7426. .7531
                     /note="44231, SHGC-36058, Chr. 16, Homo sapiens"
                     /db_xref="dbSTS:G30050"
     gene            27016. .108426
                     /gene="A-761H5.3"
                     /pseudo
     STS             27383. .27567
                     /gene="A-761H5.3"
                     /note="25260, CHLC.GCT16G06, Chr. -, Homo sapiens"
                     /db_xref="dbSTS:G16020"
     mRNA            complement(join(32539. .32821,32917. .33079,33182. .33269,
                     33358. .33423,33522. .33681,35998. .36361,40957. .41181,
                     41407. .41597,43190. .43283,43419. .43544,43739. .43880,
                     44649. .44752,44877. .45036,52772. .52899,53278. .53387,
                     53516. .53570,53760. .53885,54108. .54297,54440. .54505,
                     56420. .56550,56689. .56743))
                     /gene="A-761H5.1"
     gene            complement(32539. .56743)
                     /gene="A-761H5.1"
     CDS             complement(join(32739. .32821,32917. .33079,33182. .33269,
                     33358. .33423,33522. .33681,35998. .36361,40957. .41181,
                     41407. .41597,43190. .43283,43419. .43544,43739. .43880,
                     44649. .44752,44877. .45036,52772. .52899,53278. .53387,
                     53516. .53570,53760. .53885,54108. .54297,54440. .54505,
                     56420. .56520))
                     /gene="A-761H5.1"
                     /codon_start=1
                     /product="Translation initiation factor eIF-p110"
                     /db_xref="PID:g3337383"
                     /translation="MSRFFTTGSDSESESSLSGEELVTKPVGGNYGKQPLLLSEDEED
                     TRRVVRSAKDRRFEELTNLIRTIRNAMKIRDVTKCLEEFELLGKAYGKAKSIVDKEGV
                     PRFYIRILADLEDYLNELWEDKEGKKKMNKNNAKALSTLRQKIRKYNRDFESHITSYK
                     QNPEQSADEDAEKNEEDSEGSSDEDEDEDGVSAATPLKKKSEAPSGESRKFLKKMDDE
                     DEDSEDSEDDEDWDTGSTSSDSDSEEEEGKQTALASRFLKKAPTTDEDKKAAEKKRED
                     KAKKKHDRKSKRLDEEEEDNEGGEWERVRGGVPLVKEKPKMFAKGTEITHAVVIKKLN
                     EILQARGKKGTDRAAQIELLQLLVQIAAENNLGEGVIVKIKPNIIASLYDYNPNLATY
                     MKPEMWGKCLMCINELKDTLFANPNIFVGENILEESEWLHNADQPLRVRGCILTLVER
                     MDEEPTKIMQWTDPHSQETVEHLKDEAQVCAIIERVQRYLEEKGTTEEVCRIYLLRIL
                     HTYYKFDYKAHQRQLTPPEGSSKSEQDQAENEGEDSAVLMERLCKYIYAKDRTDRIRT
                     CAILCHIYHHALHSRWYQARDLMLMSHLQDNIQHADPPVQILYNRTMVQLGICAFRQG
                     LTKDAHNALLDIQSSGRAKELLGQGLLLRSLQERNQEQEKVERRRQVPFHLHINLELL
                     ECVYLVSAMLLEIPYMAAHESDARRRMISKQFHHQLRVGERQPLLGPPESMREHVVAA
                     SKAMKMGDWKTCHSFIINEKMNGKVWDLFPEADKVRTMLVRKIQEESLRTYLPTTSSV
                     YDSISMETLSDMPELDLPTVHSIISKMIINEELMASLDQPTQTVVMHRTEPTAQQNLA
```



```
                     LQLAEKLGSLVENNERVFDHKQGTYGGYFRDQKDGYRKNEGYMRRGGYRQQQSQTAY"
     STS             57091. .57208
                     /gene="A-761H5.3"
                     /note="7608, STS1-cSRL-24g1-uA/cSRL-24g1-uE, Chr. -, Homo
                     sapiens"
                     /db_xref="dbSTS:G02122"
     mRNA            join(<66573. .66995,67730. .>68008)
                     /gene="A-761H5.2"
     gene            <66573. .>68008
                     /gene="A-761H5.2"
     CDS             join(66573. .66995,67730. .68008)
                     /gene="A-761H5.2"
                     /codon_start=1
                     /product="CDC37-like gene"
                     /db_xref="PID:g3337384"
                     /translation="MALGGSTTSLRTTVQGVSWVAEAKCWTTACGATLSDDEEERHPN
                     THRANLFRWQHQARVVELMEQPQKEKEELDRGCWEYKCKMAECQRKLKELEVSEGARVE
                     VEWLHEELSWEQKLEEMPKKKSMTWNVDTLSKSSFRLGEMDPTRAKFPMQQCIDAGLR
                     VPNSEVSEAKEGEEECSGTGCWKLFPRQAGMSVRDPPQLQLPPVCKPLGAPFQKTKID
                     TISPAPGFLHFRCSP"
     mRNA            complement(join(<109422. .110066,111426. .111461,
                     111571. .111631,115712. .115907,116020. .116119,
                     123403. .>123468))
                     /gene="A-761H5.4"
     gene            complement(<109422. .>123468)
                     /gene="A-761H5.4"
     CDS             complement(join(109422. .110066,111426. .111461,
                     111571. .111631,115712. .115907,116020. .116119,
                     123403. .123468))
                     /gene="A-761H5.4"
                     /codon_start=1
                     /product="Unknown gene product splice form-1"
                     /db_xref="PID:g3337385"
                     /translation="MRLRFWLLIWLLLGFISHQPTPVSFLKTIFWSRNGHDGSMDVQQ
                     RAWRSNXSRQKGLRSICMHTKKRVSSFRGNKIGLKDVITLRRHVETKVRXKIRKRKVT
                     TKINRHDKINGKRKTARKQKMPQRAQELRRRAEDYHKCKIPPSARKPLCNWVRMAAAE
                     HRHSSGLPYWLYLTAETLKNRMGRQPPPPTQQHSITDNSLSLKTPPECLLTPLPPSVD
                     DNIKECPLAPLPPSPLPPSVDDNLKECLFVPLPPSPLPPSVDDNLKECLFVPLPPSPL
                     PPSVDDNLKTPPLATQEAEVEKPPKPKRWRVDEVEQSPKPKRRRVDEVEQSPKPKRQR
                     EAEAQQLPKPKRRRLSKLRTRHCTQAWAIRINP"
     STS             109425. .109537
                     /gene="A-761H5.4"
                     /note="40991, SHGC-32146, Chr. 16, Homo sapiens"
                     /db_xref="dbSTS:G27136"
     mRNA            complement(join(110411. .111461,111571. .111631,
                     115712. .115907,116020. .116119,123403. .>123468))
                     /gene="A-761H5.5"
     gene            complement(110411. .>123468)
                     /gene="A-761H5.5"
     CDS             complement(join(111219. .111461,111571. .111631,
                     115712. .115907,116020. .116119,123403. .123468))
```



```
                     /gene="A-761H5.5"
                     /codon_start=1
                     /product="Unknown gene product splice form-2"
                     /db_xref="PID:g3337386"
                     /translation="MRLRFWLLIWLLLGFISHQPTPVSFLKTIFWSRNGHDGSMDVQQ
                     RAWRSNXSRQKGLRSICMHTKKRVSSFRGNKIGLKDVITLRRHVETKVRXKIRKRKVT
                     TKINRHDKINGKRKTARKQKMPQRAQELRRRAEDYHKCKIPPSARKPLCNWVSLFVFL
                     AFEHSLPGQDMDTFFSLQLCAQALQREMAERKAATRHHSPIPVGNRVVQKHLHPHPVG
                     PLI"
     STS             121375. .121480
                     /gene="A-761H5.5"
                     /note="44231, SHGC-36058, Chr. 16, Homo sapiens"
                     /db_xref="dbSTS:G30050"
     mRNA            complement(join(<130544. .130663,130765. .130905,
                     135131. .135224,135353. .135408,135503. .135571,
                     135652. .135698,137020. .137132,139362. .139505,
                     139593. .139665,140471. .140556,140677. .140756,
                     141606. .141677,142305. .142401,144492. .144570,
                     144724. .144946))
                     /gene="A-761H5.6"
     gene            complement(<130544. .144946)
                     /gene="A-761H5.6"
     CDS             complement(join(130544. .130663,130765. .130905,
                     135131. .135224,135353. .135408,135503. .135571,
                     135652. .135698,137020. .137132,139362. .139505,
                     139593. .139665,140471. .140556,140677. .140756,
                     141606. .141677,142305. .142401,144492. .144570,
                     144724. .144769))
                     /gene="A-761H5.6"
                     /codon_start=1
                     /product="CLN"
                     /db_xref="PID:g3337387"
                     /translation="MGGCAGSRRRFSDSEGEETVPEPRLPLLDHQGAHWKNAVGFWLL
                     GLCNNFSYVVMLSAAHDILSHKRTSGNQSHVDPGPTPIPHNSSSRFDCNSVSTAAVLL
                     ADILPTLVIKLLAPLGLHLLPYSPRVLVSGICAAGSFVLVAFSHSVGTSLCGVVFASI
                     SSGLGEVTFLSLTAFTPRAVISWWSSGTGGAGLLGALSTYLGLTQAGLSPQQTLLSMLG
                     IPALLLASYFLLLTSPEAQDPGGEEEAESAARQPLIRTEAPESKPGSSSSLSLRERWT
                     VFKGLLWYIVPLVVVYFAETYFINQGLFELLFFWNTSLSHAQQYRWYQMLYQAGVFASR
                     SSLRCCRIRFTWALALLQCLNLVFLLADVWFGFLPSIYLVFLIILYEGLGGAAYVNT
                     FHNIALETSDEHREFAMAATCISDTLGISLSGLLALPLHDFLCQLS"
BASE COUNT    41602 a  38891 c  38211 g  37878 t     59 others
ORIGIN

  Query Match          72.1%;  Score 44;  DB 26;  Length 156641;
  Best Local Similarity  97.8%;  Pred. No. 4.05e-17;
  Matches    44;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db 151902 AAAAACAAAATAAAACAATAAAGACTGCAAGGAAGACTGAGGGAA 151946
          ||||||||||||||||||||||||||||||||||||||||||| ||||
Qy     17 AAAAACAAAATAAAACAATAAAGACTGCAAGGAAGACTGANGGAA 61
```



```
RESULT    3
LOCUS       HS394I7    116271 bp    DNA              HTG       23-SEP-1998
DEFINITION  Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone 394I7;
            HTGS phase 1.
ACCESSION   AL023585
NID         g3649712
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 116271)
  AUTHORS   Grafham,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-SEP-1998) Wellcome Trust Genome Campus, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquires:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 25, 1998 this sequence version replaced gi:3135973.
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and the release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc. Order of segments is not known; 800 n's separate
            segments. Unfinished: dJ394I7  Contig_ID: 00703  acc=AL023585
            Length: 2083 bp Unfinished: dJ394I7  Contig_ID: 02076  acc=AL023585
             Length: 3380 bp Unfinished: dJ394I7  Contig_ID: 01595
            acc=AL023585  Length: 30283 bp Unfinished: dJ394I7  Contig_ID:
            00482  acc=AL023585  Length: 76304 bp Unfinished: dJ394I7
            Contig_ID: 00546  acc=AL023585  Length: 1021 bp.
            ***                                                    ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                    ***
            * This sequence is unfinished. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
FEATURES              Location/Qualifiers
     source           1. .116271
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="1"
                      /clone="394I7"
BASE COUNT    31632 a  25890 c  25318 g  30228 t   3203 others
ORIGIN

  Query Match          36.1%;  Score 22;  DB 17;  Length 116271;
  Best Local Similarity 100.0%;  Pred. No. 7.55e-02;
  Matches    22;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   5138 AAAAACAAAATAAAACAATAAA 5159
```

```
                    ||||||||||||||||||||||
Qy      17 AAAAACAAAATAAAACAATAAA 38


RESULT    4
LOCUS       HSPA32H10      268 bp    DNA             STS       21-MAY-1998
DEFINITION  H.sapiens flow-sorted chromosome 6 HindIII fragment, SC6pA32H10,
            sequence tagged site.
ACCESSION   Z94233
NID         g1945227
KEYWORDS    STS; single read.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 268)
  AUTHORS   Mungall,A.J., Huckle,E., Langford,C., Ross,M.T. and Hunt,S.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-APR-1997) The Sanger Centre, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail contact:
            humquery@sanger.ac.uk
COMMENT     Vector: pBSIISK+.
FEATURES             Location/Qualifiers
     source          1. .268
                     /organism="Homo sapiens"
                     /note="The estimated purity of the flow-sorted chromosome
                     6 library is >97%"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /sex="Female"
                     /dev_stage="adult"
                     /tissue_type="EBV lymphoblastoid cell line"
                     /clone_lib="SC6pA"
                     /clone="SC6pA32H10"
BASE COUNT      129 a     48 c     33 g     58 t
ORIGIN

  Query Match           34.4%;  Score 21;  DB 30;  Length 268;
  Best Local Similarity 100.0%;  Pred. No. 3.07e-01;
  Matches    21;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      17 AAAAACAAAATAAAACAATAA 37
           |||||||||||||||||||||
Qy      17 AAAAACAAAATAAAACAATAA 37


RESULT    5
LOCUS       CELF42C5    41351 bp    DNA             INV       17-NOV-1995
DEFINITION  Caenorhabditis elegans cosmid F42C5.
ACCESSION   U40799
NID         g1065935
```



```
KEYWORDS    .
SOURCE      Caenorhabditis elegans strain=Bristol N2.
  ORGANISM  Caenorhabditis elegans
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Nematoda;
            Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea;
            Rhabditidae; Caenorhabditis.
REFERENCE   1  (bases 1 to 41351)
  AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
  TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
  JOURNAL   Nature 368 (6466), 32-38 (1994)
  MEDLINE   94150718
REFERENCE   2  (bases 1 to 41351)
  AUTHORS   Du,Z.
  TITLE     The sequence of C. elegans cosmid F42C5
  JOURNAL   Unpublished (1995)
REFERENCE   3  (bases 1 to 41351)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-1995) Robert Waterston
COMMENT               Submitted by:
                      Genome Sequencing Center
                      Department of Genetics, Washington University,
                      St. Louis, MO  63110, USA, and
                      Sanger Centre, Hinxton Hall
                      Cambridge CB10 1RQ, England
                      e-mail: rw@nematode.wustl.edu and jes@sanger.ac.uk
                      NEIGHBORING COSMID INFORMATION:
            The 5' cosmid is C06E4, 200 bp overlap;3' end lies in a gap
            followed by the cosmid T12B3. Actual start of this cosmid is at
            base position 1 of CELF42C5; actual end is at base position 41351
            of CELF42C5.

            NOTES:
            Coding sequences below are predicted from computer analysis, using
            the program Genefinder(P. Green and L. Hillier, ms in preparation).
FEATURES             Location/Qualifiers
     source          1. .41351
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /map="IV"
```



```
     gene            complement(979. .2164)
                     /gene="F42C5.3"
     CDS             complement(join(979. .1116,1265. .1333,1382. .1501,
                     1545. .1778,1973. .2164))
                     /gene="F42C5.3"
                     /codon_start=1
                     /evidence=not_experimental
                     /db_xref="PID:g1065936"
                     /translation="MEDDHSKEMDRLRSQAAKVQSEKETFDRKRRETLEKHKKELDEL
                     EKKNKEEDDLREQNMNLWNKNLDQQIALGNELNNKYTEISNQNSRLQIKIGQEEDIR
                     VFRIKLLDVSKVWTDVKVNYQDYLRNTLDEHSNSNKSDVLKEIDTLIYNKEKLNEVLI
                     TAKKLLGKCQKFTTSDSFKVINDSLTELMRFKFEDDILILEKLTIIKKNGSAEQSFLTK
                     MDETIDKYNEMVNELPGLQLKSVEPIHQAAIQ"
     gene            complement(2977. .6924)
                     /gene="F42C5.4"
     CDS             complement(join(2977. .3636,3722. .3835,3885. .3997,
                     4245. .4323,4369. .5030,5213. .5405,5480. .5670,5778. .5846,
                     6867. .6924))
                     /gene="F42C5.4"
                     /note="short region of similarity to S. cerevisiae nuclear
                     polyadenylated RNA-binding protein NAB3 (SP:P38996)"
                     /codon_start=1
                     /evidence=not_experimental
                     /db_xref="PID:g1065937"
                     /translation="MEWKRVEARLEISHGSIATEMLHFVSLVCSGCDSLGDRDRFFDN
                     DKKRPCVSSCNHSLCSQCFDKFTECPICKEPLVRSENYQARKIVEHFKENPTMIIEKW
                     WNSTANGSETCSHCHNDAKKPRFCVTCGLQKKYLMMCRVPKDGIERDDAEEQEEWKRR
                     QEEETRKSKKYLIVNEFLPKSNSWYDDKKMRPVPGCPFNVMANQLLCADCCFEHHEDH
                     ALRTSEQLMYIDADWNKSALKIVSDPIWYELRSRKSKCLIKTMRLHRTCEKLIHFVKC
                     FYTYTEPSDRESHVKGRWRSDVNKVFESLSRKNVDSVSRETVDELIQSLSIQLKYLTY
                     HDDCKDDIWNQMHKISFGNQVEKLFVKVINELEDREIKDCPLTDMDKQNIQNKVEEV
                     SKHQIFRIYTAFCWNYTEKGVFHLFDNNEHKPCVCPNCNISTCLECLKSNWLYKCSPC
                     GEAFFEVGLGIVREFNEDVNILELVGLYKRNCVDFLEEWWRCEPSHLGPCLSCSSYSE
                     ELEICVFCEFLDKHKALKTKPKIENSIQMRQNLSLEFNYSGLNDFPIRWQCSDCKKRL
                     HSNWVHSQCDSSIPEHWKGTWRRGCNHILTRILRSKCKCYDANADECETYNSITLKSIR
                     NYEFAMKVATMGLIFRILKSGIEENITCKIKRGEMLNTYGLLKSKCRYFFINSMKEEI
                     NEEMNRLSNEINKELDNLKIQWKNYNGITI"
     gene            complement(7399. .8928)
                     /gene="F42C5.5"
     CDS             complement(join(7399. .7449,7500. .7864,7915. .8305,
                     8764. .8928))
                     /gene="F42C5.5"
                     /codon_start=1
                     /evidence=not_experimental
                     /db_xref="PID:g1065938"
                     /translation="MKIYREESQQTTVVGEPVDEFVHLQTTRGKETYQQDLDDMNLDD
                     EKIERFIRNGEKNRFTNVTCPDEILDRENQEFFVHAMTFRTPYGNFLISQAPNDSTNA
                     DHLHLMWIYNVQTVVCLVLPEEAGGYFVPEEGETPTSFQKYQIKTIGIPDEKQGVTVY
                     QCELKSYRAPKKLNRRMVYIICCDTSVGSIRSPRQQTVIMEYMWAFEETNAIENEVAL
                     ADATTTVLVHGIAGTRRCASFVVAAVMCKQLLETGQFSAMETWFEMKRRRAHTCTRKH
                     DFPSSVYTFFAPCVQSSAVSETNENYAKTMEIVKDLLVKKNKEEKND"
```



```
     gene            19671. .23147
                     /gene="F42C5.2"
     CDS             join(19671. .19844,21175. .21253,21319. .21427,21912. .22163,
                     22382. .22873,22922. .22974,23026. .23147)
                     /gene="F42C5.2"
                     /note="similar to MU-typ opiod and somatostatin receptors"
                     /codon_start=1
                     /evidence=not_experimental
                     /db_xref="PID:g1065939"
                     /translation="MTKKNWISIKNYANGFHQKLCKFDEPVTANNVKCVTIVDLRLKL
                     ELEKQKGSSFSNGILEIFYSLIYLFPLKDKSFPEMFIFENDIDMSFAKGFMTVVYIVV
                     CLVGTPGNLWIIYKLFRAKLWSGASVQLTVSQRSRITIFALACSDMFLLLTLPATAGY
                     NYHGTWIFGSAACYIIRSIEIFAKLFSVVLLTVMSLERYIIVCTRLRHIYRAWNSLVP
                     LAVGTIFGVLVPTIIHYFTLQHFSVPFDPDVTWCLPLMSNEVFNLFAQTTPVVGFII
                     PFAIMACYIMLVRHVRTKYKMRRAQTTTTLAKPGKEPRYMSEVKKSIWRIAVFHFVC
                     WAPFWGFTMLPNYIYQIDQFLHGDNEEESGGESIPLVYCRLVSNCLPYIINAAGNWVLY
                     ALLNVDVRKHIYNQPKKRKRFTLKFNPVNTSSNC"
     gene            complement(23283. .24752)
                     /gene="F42C5.6"
     CDS             complement(join(23283. .23477,23525. .23665,23714. .23748,
                     24683. .24752))
                     /gene="F42C5.6"
                     /codon_start=1
                     /evidence=not_experimental
                     /db_xref="PID:g1065940"
                     /translation="MHSGLCCIYDLHFSLCLSLPSSPEETQTEGSIGTTTVVKIESDA
                     IWCPTHRVGNQCPEETLLSFYKCCGHLNKECCFHLRVWIVIMIIVLPICLVLPASTFL
                     IRRILCSSSQQPTGRAPSAADRVRQLAAEDSTTFRRDSDDMVML"
     gene            complement(24822. .26102)
                     /gene="F42C5.7"
     CDS             complement(join(24822. .25001,25049. .25120,25175. .25555,
                     25628. .25703,26059. .26102))
                     /gene="F42C5.7"
                     /codon_start=1
                     /evidence=not_experimental
                     /db_xref="PID:g1065941"
                     /translation="MEAVPKDVRVKMAKRKLRRKFEITASETGYYNFGAKLRVWMLLP
                     ISILFSIIPIASSTFGGCCSWGPPPCPPPPPPMCAPPPLPCPPPPICPPQFCPPPPMC
                     PPPPPPPPPPMCPPPPPPMPSYSPCQSYAPAPVFNQYAMQPANDCCCRCGSPCRFMAR
                     HRTHGSKLFTTEDPEEDPTCNSKKLRRVMERMNNGDPSISKRAIQKAVEEKMFGKFNV
                     ICARGDFSTVATTETYCQVANDDVTCYAFRPM"
     gene            complement(27841. .28656)
                     /gene="F42C5.8"
     CDS             complement(join(27841. .27950,27997. .28302,28353. .28452,
                     28546. .28656))
                     /gene="F42C5.8"
                     /note="coded for by C. elegans cDNA yk51c6.5; coded for by
                     C. elegans cDNA yk107e3.5; coded for by C. elegans cDNA
                     yk103h6.5; coded for by C. elegans cDNA yk55e5.5; coded
                     for by C. elegans cDNA yk78a4.5; coded for by C. elegans
                     cDNA yk107e3.3; coded for by C. elegans cDNA yk103h6.3;
```

```
                     coded for by C. elegans cDNA yk55e5.3; coded for by C.
                     elegans cDNA yk78a4.3;  Similar to 40S ribosomal protein
                     S8."
                     /codon_start=1
                     /db_xref="PID:g1065942"
                     /translation="MGISRDSWHKRYKTGATQPVPHKKRKFELGRPAANTKIGAHRVR
                     LVRTRGGNKKYRALRLDSGNFSWASEQTTRKTRIVDTMTNATWNELVRTKTLVKGAII
                     SVDAAPFRQWYEAHTALPLARKKNAKLSEEDNAILNKKRSHHTMKYTERQKTAAVDA
                     LLIEQFNTGRLLARISSSPGQVGQANGYILEGKELDFYLRKIRAKKAK"
     gene            28209. .29786
                     /gene="F42C5.1"
     CDS             join(28209. .28440,29481. .29696,29749. .29786)
                     /gene="F42C5.1"
                     /codon_start=1
                     /evidence=not_experimental
                     /db_xref="PID:g1065943"
                     /translation="MAAPLTRVLVRTNSLFVALYMVSTIRVLRVVSEKSEYERFIKYN
                     NETYAQRPRRSFRSRGEELCTFRFLHECERDERGRSFRKTMTGAVQGTLEAVEEKCRS
                     LQHWRLMRRSEILKFYRCSISWFFFSWWAYLGSKICCVIPPWTDVNLTIPGIRDAPPA
                     K"
     gene            complement(29490. .31891)
                     /gene="F42C5.9"
     CDS             complement(join(29490. .29582,29645. .29792,29842. .30089,
                     30136. .30396,30445. .30934,31019. .31242,31697. .31891))
                     /gene="F42C5.9"
                     /note="coded for by C. elegans cDNA CESAA62F; coded for by
                     C. elegans cDNA CEESL37R; coded for by C. elegans cDNA
                     CEESV61F; coded for by C. elegans cDNA CEESX16F; coded for
                     by C. elegans cDNA CEESL37F; coded for by C. elegans cDNA
                     CEESX39F;  short region of similarity to actins"
                     /codon_start=1
                     /db_xref="PID:g1065944"
                     /translation="MPDLEINGVEMENGVGRSDSLKAEEANALREWDPKHLLKALYEI
                     SYEPRVETKRNRFVTMEGHVEVPADEVTKSDGLDNEGWMKQYIRMKEGRLQIFASHYA
                     GEAPAQELVLSGADVDANKDERTLTIHGGREHVKYYCRVPSAVFDKWRQSPLSHCASS
                     LIDAYVKPIPRAFQHLTERILILELGSCGTRAGVLTTEPSLPQSPFPAIAVRTDPGKI
                     FVGEDAYRPEIRHNGDPVRPIASATDPSVERYSIDKDVLKAVIGEVTKELVEPWKYAV
                     LLSIPQNIPTVLIGELLQICLEDIKPQSAGITRQPSLILYAYDVTTGVVVDIGERVNI
                     VPVIDGYVVESAICSIPTGSQQMGDSLRASLSAHNKGLYAPQSPVEKLILRFALEQTT
                     YVPEDYQKEEKSENKETRISLDAFTPSPGMQTRFDIDNSRFLCTEGLFKPKKWGLDTK
                     GLHQLIHEAILQCPIDSRRTLYRNIYLAGGASLMPGIAEKLEQELSTVVPNTIHTQVN
                     ISPWRYNAAYLGAQIVASASTFEDSCISLQQLPKPLAQLQSSSF"
     gene            complement(35688. .40527)
...
Note: remainder of annotations omitted.

  Query Match           32.8%;  Score 20;  DB 18;  Length 41351;
  Best Local Similarity 100.0%;  Pred. No. 1.22e+00;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   7310 AAATAAAACAATAAAGACTG 7329
```



```
              ||||||||||||||||||||
Qy      24 AAATAAAACAATAAAGACTG 43


RESULT    6
LOCUS     AC005545    43514 bp    DNA             PRI       03-SEP-1998
DEFINITION  Homo sapiens chromosome 19, cosmid R26634, complete sequence.
ACCESSION   AC005545
NID       g3478638
KEYWORDS  HTG.
SOURCE    human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 43514)
  AUTHORS   Lamerdin,J.E., McCready,P.M., Skowronski,E., Adamson,A.W.,
            Burkhart-Schultz,K., Gordon,L., Kyle,A., Ramirez,M., Stilwagen,S.,
            Phan,H., Velasco,N., Do,L., Regala,W., Terry,A., Garnes,J.,
            Danganan,L., Poundstone,P., Christensen,M., Georgescu,A., Avila,J.,
            Liu,S., Attix,C., Andreise,T., Trankheim,M., Amico-Keller,G.,
            Coefield,J., Duarte,S., Lucas,S., Bruce,R., Thomas,P., Quan,G.,
            Kronmiller,B., Arellano,A., Montgomery,M., Ow,D., Nolan,M.,
            Trong,S., Kobayashi,A., Olsen,A.S. and Carrano,A.V.
  TITLE     Sequence analysis of a 3.5 Mb contig in human 19p13.3 containing a
            serine protease gene cluster
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 43514)
  AUTHORS   Lamerdin,J.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-AUG-1998) Joint Genome Institute, Lawrence Livermore
            National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
REFERENCE   3  (bases 1 to 43514)
  AUTHORS   Lamerdin,J.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-SEP-1998) Joint Genome Institute, Lawrence Livermore
            National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
COMMENT     Map and sequence oriented from p telomere to centromere.  Cosmid
            R26634 overlaps cosmid R26660 (AC005328) to the left from bases 1
            to 4,683 of this accession, and overlaps cosmid F8682 (AC005257) to
            the right from bases 41,662 to 43,514. Additional chromosome 19 map
            and sequence information may be obtained at:
            http://www-bio.llnl.gov/bbrp/genome/genome.html.
FEATURES              Location/Qualifiers
     source           1. .43514
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="R26634"
                      /chromosome="19"
                      /map="19p13.3 between CDC34 and D19S342"
                      /cell_line="5HL2-B"
                      /clone_lib="LL19NC03 R chromosome 19-specific cosmid
```



```
                     library"
                     /note="Cosmid library constructed at LLNL from flow-sorted
                     chromosomes from hybrid 5HL2-B, which carries chromosome
                     19 as its only human chromosome."
     misc_feature    46. .152
                     /note="DDS similarity to overlapping ESTs:~(46. .152)
                     AA431731 zw77g03.s1 Soares testis NHT Homo sapiens cDNA
                     clone 782260 3'; (391. .340); 100% identity.~~(111. .152)
                     AA609572 af15c02.s1 Soares testis NHT Homo sapiens cDNA
                     clone 1031714 3'; (372. .331); 100% identity."
     misc_feature    complement(273. .608)
                     /note="DDS similarity to overlapping ESTs:~(608. .273)
                     T05239 EST03128 Homo sapiens cDNA clone HFBCT75. Score:
                     621  Identity: 325/337 (96%).~~(718. .308) AA496009
                     zv72a03.s1 Soares total fetus Nb2HF8 9w Homo sapiens cDNA
                     clone 759148 3'; Score: 755  Identity: 398/408 (97%)."
     misc_feature    378. .707
                     /note="DDS similarity to overlapping ESTs:~AA431731
                     zw77g03.s1 Soares testis NHT Homo sapiens cDNA clone
                     782260 3'; (339. .11); 99% identity.~~AA609572 af15c02.s1
                     Soares testis NHT Homo sapiens cDNA clone 1031714 3';
                     (330. .5); 99% identity.~~AA448181 zw83c05.s1 Soares testis
                     NHT Homo sapiens cDNA clone 782792 3'; (336. .4); 99%
                     identity.~"
     CDS             <378. .468
                     /note="Hypothetical human protein with no significant
                     similarity to known proteins"
                     /codon_start=2
                     /evidence=not_experimental
                     /product="R26660_2, partial CDS"
                     /db_xref="PID:g3478640"
                     /translation="LVSPALRCLEPPHLANLTLEDAAECLKQH"
     misc_feature    1771. .3383
                     /note="DDS similarity to overlapping ESTs:~(3391. .3154)
                     AA630776 acl4c10.s1 Stratagene HeLa cell s3 937216 Homo
                     sapiens cDNA clone 856434 3' similar to TR:G163328 G163328
                     LEUKEMIA VIRUS CELL RECEPTOR. ; (262. .1); 99%
                     identity.~~(1771. .2449) AA167736 sq40d04.s1 Stratagene hNT
                     neuron (#937233) Homo sapiens cDNA clone 632167 3'; Score:
                     1207  Identity: 655/674 (97%).~~(2109. .2638) AA843415
                     ak07f04.s1 Soares parathyroid tumor NbHPA Homo sapiens
                     cDNA clone IMAGE:1405279 3'; Score: 969  Identity: 516/547
                     (94%).~~(2814. .2510) AA364678 EST75336 Pineal gland II
                     Homo sapiens cDNA 5' end; Score: 582  Identity: 301/305
                     (98%).~~(3200. .2766) H73448 yu02f02.r1 Homo sapiens cDNA
                     clone 232635 5'. Score: 717  Identity: 413/446
                     (92%).~~(3358. .2857) H28923 ym33c10.r1 Homo sapiens cDNA
                     clone 49874 5'. Score: 756  Identity: 465/518
                     (89%).~~(3383. .2950) W69851 zd48c04.r1 Soares fetal heart
                     NbHH19W Homo sapiens cDNA clone 343878 5' similar to
                     PIR:A45716 A45716 leukemia virus cell receptor ;~Score:
```



```
                     744  Identity: 418/440 (95%). ~~Additional EST
                     matches:~AA617872, AA716429, W63772, AA305631, AA321888,
                     AA371945, R18826, T33092, T30887, F01584 and many others"
     mRNA            complement(2119. .>39837)
                     /note="Human AP-3 complex delta subunit mRNA"
                     /product="Homo sapiens delta-adaptin mRNA"
     CDS             complement(join(3296. .3391,9810. .9889,10209. .10330,
                     10996. .11081,11259. .11347,11830. .12019,12408. .12455,
                     12802. .12951,13983. .14054,14459. .14536,15248. .15425,
                     15871. .15944,16342. .16541,16661. .16736,17330. .17401,
                     17728. .17869,18345. .18490,19724. .19955,21985. .22215,
                     22286. .22434,22857. .23002,24481. .24529,24953. .25002,
                     28275. .28324,30213. .30244,31625. .31660,33594. .33701,
                     38134. .38214,38850. .38930,39742. .>39837))
                     /note="vesicle coat component, similar to alpha and gamma
                     adaptins"
                     /codon_start=1
                     /product="AP-3 complex delta subunit, partial CDS"
                     /db_xref="PID:g3522925"
                     /translation="AKYISQCIDEIKQELKQDNIAVKANAVCKLTYLQMLGYDISWAA
                     FNIIEVMSASKFTPKRIGYLAASQSFHEGTDVIMLTTNQIRKDLSSPSQYDTGVALTG
                     LSCFVTPDLARDLANDIMTLMSHTKPYIRKKALIEPLTNLIHSTSAMSLLYECVNTVI
                     AVLISLSSGMPNHSASIQLCVQKLRILIEDSDQNLKYLGLLAMSKILKTHPKSVQSHK
                     DLILQCLDDKDESIRLRALDLLYGMVSKKNLMEIVKKLMTHVDKAEGTTYRDELLTKI
                     IDICSQSNTQYITNFEWYISILVELTRLEGTRHGHLIAAQKLDVAIIVTKAIRKFAVSQ
                     MSALLSAHLLASTQRNGICEVLVAAAWICGEPSEHLQEPHHTLEAMLRPRVTTLPG
                     HIQAVTVQNVVKLYASILQQKEQAGEAEGAQAVTQLMVDRLPQFVQSADLEVQERASC
                     ILQLVKHIQKLQAKDVPVAEEVSALFAGELNPVAPKAQKKVPVPEGLDLDAWINEPLS
                     DSESEDERPRAVPHEEEQRRPKHRPSEADEEELARRREARKQEQANNPFYIKSSPSPQ
                     KRYQDTPGVEHIPVVQIDLSVPLKVPGLPMSDQYVKLEEERRHRQKLEKDKRRKKRKE
                     KEKKGKRRHSSLPTESDEDIAPAQQVDIVTEEMPENALPSDEDDKDPNDPYRALDIDL
                     DKPLADSEKLPIQKHRNTETSKSPEKDVPMVEKKSKKPKKKEKKHKEKERDKEKKKEK
                     EKKAEDLDFWLSTTPPPAPAPAPVPSTDECEDAKTEAQGEEDDAEGQDQDKKSPKP
                     KKKKHRKEKEERTKGKKKSKKQPPGSEEAAGEPVQNGAPEEEQLPPESSYSLLAENSY
                     VKMTCDIRGSLQEDSQVTVAIVLENRSSSILKGMELSVLDSLNARMARPQGSSVHDGV
                     PVPFQLPPGVSNEAQVTVPTIQSIVMAQKLKGTLSFIAKNDEGATHEKLDFRLHFCSS
                     YLITPCYSDAFALLESGDLSWSSIKVDGIRMSPQNLLAKICFHHHPSVVERVDSCA
                     SMYSRSIQGHHVCLLVKKGENSVSVDGKCSDSTLLSNLLEEMKATLAKC"
     CDS             complement(join(3296. .3391,9810. .9889,10209. .10330,
                     10996. .11081,11259. .11347,11830. .12019,12408. .12455,
                     12802. .12951,15248. .15425,15871. .15944,16342. .16541,
                     16661. .16736,17330. .17401,17728. .17869,18345. .18490,
                     19724. .19955,21985. .22215,22286. .22434,22857. .23002,
                     24481. .24529,24953. .25002,28275. .28324,30213. .30286,
                     30441. .30580,31531. .31660,33594. .33701,38134. .38214,
                     38850. .38930,39742. .>39837))
                     /function="vesicle coat component"
                     /function="similar to alpha and gamma adaptins"
                     /note="AP-3 complex delta subunit; Human homolog of mouse
                     bovine leukemia virus receptor"
                     /codon_start=1
```

```
                     /product="delta-adaptin, partial CDS"
                     /db_xref="PID:g3478639"
                     /translation="ARYISQCIDEIKQELKQDNIAVKANAVCKLTYLQMLGYDISWAA
                     FNIIEVMSASKFTFKRIGYLAASQSFHEGTDVIMLTTNQIRKDLSSPSQYDTGVALTG
                     LSCFVTPDLARDLANDIMTLMSHTKPYIRRKAVLIMTKVFLKYPESLRPAFPRLKEKL
                     EDPDPGVQSAAVNVICELARRNPKNYLSLAPLFFKLMTSSTNNWVLIKIIKLFGALTP
                     LEPRLGKKLIEPLTNLIHSTSAMSLLYECVNTVIAVLISLSSGMPNHSASIQLCVQKL
                     RILIEDSDQNLKYLGLLAMSKILKTHPKSVQSHKDLILQCLDDKDESIRLRALDLLYG
                     MVSKKNLMEIVKKLMTHVDKAEGTTYRDELLTKIIDICSQSNYQYITNFEWYISILVE
                     LTRLEGTRHGHLIAAQMLDVAIRVKAIRKFAVSQMSALLDSAHLLASSTQRNGICEVL
                     YAAAWICGEFSEHLQEPHHTLEAMLRPRVTTLPGHIQAVTVQNVVKLYASILQQKEQA
                     GEAEGAQAVTQLMVDRLPQPVVQSADLEVQERASCILQLVKHIQKLQAKDVPVAEEVSA
                     LFAGELHPVAPKAQKEVPVPEGLDLDAWINEPLSDSESEDERPRAVPHEEEQRRPKHR
                     PSEADEEELARRREARKQEQANNPFYIKSSPSPQKRYQDTPGVEHIPVVQIDLSVPLK
                     VPGLPMSDQYVKLEEERRHRQKLEKDKRRKKRKEKEKKGKRRHSSLPTESDEDIAPAQ
                     QVDIVTEEMPENALPSDEDDKDPNDPYRALDIDLDKPLADSEKLPIQKHRNTETSKSP
                     EKDVPMVEKKSKKPKKKEKKHKEKERDKEKKKEKEKKKSPKPKKKKHRKEKEERTKGK
                     KKSKKQPPGSEEAAGEPVQNGAPEEEQLPPESSYSLLAENSYVKMTCDIRGSLQEDSQ
                     VTVAIVLENRSSSILKGMELSVLDSLNARMARPQGSSVHDGVPVPFQLPPGVSNEAQY
                     VFTIQSIVMAQKLKGTLSFIAKNDEGATHEKLDFRLHFSCSSYLITTPCYSDAFAKLL
                     ESGDLSMSSIKVDGIRMSFQNLLAKICFHHHFSVVERVDSCASMYSRSIQGHHVCLLV
                     KKGENSVSVDGKCSDSTLLSNLLEEMKATLAKC"
repeat_region        3435. .3567
                     /rpt_family="AluSq"
repeat_region        3570. .3866
                     /rpt_family="AluY"
repeat_region        3867. .3906
                     /rpt_family="(TAAA)n"
repeat_region        3908. .4090
                     /rpt_family="AluSq"
repeat_region        4101. .4266
                     /rpt_family="FRAM"
repeat_region        complement(5793. .6096)
                     /rpt_family="AluJo"
repeat_region        complement(6348. .6855)
                     /rpt_family="MER9"
repeat_region        6938. .7227
                     /rpt_family="AluSg"
repeat_region        7255. .7402
                     /rpt_family="L1MB5"
repeat_region        7403. .7689
                     /rpt_family="AluSx"
repeat_region        7693. .8049
                     /rpt_family="L1MB5"
repeat_region        8105. .8399
                     /rpt_family="AluY"
repeat_region        8582. .8882
                     /rpt_family="AluY"
misc_feature         complement(9810. .9889)
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 0, quality: excellent, score: 90.000--DDS
```



```
                     similarity to AA630776 ac14c10.s1 Stratagene HeLa cell s3
...
Note: remainder of annotations omitted.

  Query Match           32.8%;  Score 20;  DB 27;  Length 43514;
  Best Local Similarity 100.0%;  Pred. No. 1.22e+00;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  36322 TCAAAAACAAAATAAAACAA 36341
          ||||||||||||||||||||
Qy     15 TCAAAAACAAAATAAAACAA 34


RESULT    7
LOCUS       AC005355    78065 bp    DNA              PRI       31-JUL-1998
DEFINITION  Homo sapiens chromosome 5, P1 clone 1307e8 (LBNL H60), complete
            sequence.
ACCESSION   AC005355 AC001062 L81708 L81709 L81710 L81882 L81883
NID         g3366557
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 78065)
  AUTHORS   Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
            Kadner,K., Miguel,T., Miller,C., Pitluck,S., Pollard,M.,
            Rojeski,H., Subramanian,S. and Martin,C.H.
  TITLE     Sequencing of human chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 78065)
  AUTHORS   Ricke,D.O.
  TITLE     Large Scale Sequence Analysis and Annotation with the Sequence
            Comparison Analysis (SCAN) System
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 78065)
  AUTHORS   Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
            Davis,C.A., Kadner,K., Miguel,T., Pitluck,S., Pollard,M.,
            Rojeski,H., Subramanian,S. and Martin,C.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JUL-1998) Human Genome Center, DOE Joint Genome
            Institute, Lawrence Berkeley National Laboratory, MS 74-157,
            Berkeley, CA 94720, U.S.A.
COMMENT     Sequence submitted by:
            DOE Joint Genome Institute.
FEATURES             Location/Qualifiers
     source          1. .78065
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="5q"
                     /clone="1307e8"
```



```
                     /chromosome="5"
                     /note="LBNL H60"
repeat_region        3. .46
                     /rpt_family="Alu"
repeat_region        69. .96
                     /note="(AGGG)7"
                     /rpt_type=tandem
                     /rpt_unit=AGGG
repeat_region        963. .1232
                     /rpt_family="Alu"
misc_feature         1801. .2307
                     /note="100% identity"
repeat_region        3286. .3343
                     /rpt_family="MIR"
misc_feature         3576. .3688
                     /note="GRAIL 2 excellent exon, frame 0"
repeat_region        complement(4146. .4250)
                     /rpt_family="Alu"
repeat_region        4282. .4303
                     /note="(T)22"
                     /rpt_type=tandem
                     /rpt_unit=T
repeat_region        complement(4286. .4684)
                     /rpt_family="Alu"
repeat_region        5405. .5546
                     /rpt_family="MIR"
repeat_region        7167. .7443
                     /rpt_family="Alu"
repeat_region        8222. .8410
                     /rpt_family="Alu"
repeat_region        complement(9121. .9458)
                     /rpt_family="L1"
repeat_region        9629. .9912
                     /rpt_family="Alu"
repeat_region        complement(10068. .10346)
                     /rpt_family="Alu"
repeat_region        complement(10903. .11006)
                     /rpt_family="Alu"
repeat_region        complement(11273. .11564)
                     /rpt_family="Alu"
repeat_region        complement(11675. .12012)
                     /rpt_family="THE1"
repeat_region        complement(12115. .12401)
                     /rpt_family="Alu"
repeat_region        12544. .12823
                     /rpt_family="Alu"
repeat_region        12806. .12829
                     /note="(A)24"
                     /rpt_type=tandem
                     /rpt_unit=A
repeat_region        complement(13163. .13938)
```



```
                     /rpt_family="L1"
repeat_region        complement(14017. .14317)
                     /rpt_family="Alu"
repeat_region        complement(14452. .14739)
                     /rpt_family="Alu"
repeat_region        complement(14758. .14897)
                     /rpt_family="Alu"
repeat_region        15490. .15766
                     /rpt_family="Alu"
repeat_region        complement(15844. .15979)
                     /rpt_family="Alu"
repeat_region        16073. .16364
                     /rpt_family="Alu"
repeat_region        16347. .16367
                     /note="(A)21"
                     /rpt_type=tandem
                     /rpt_unit=A
repeat_region        16413. .16708
                     /rpt_family="Alu"
repeat_region        16691. .16710
                     /note="(A)20"
                     /rpt_type=tandem
                     /rpt_unit=A
repeat_region        complement(16982. .17275)
                     /rpt_family="Alu"
repeat_region        17771. .18227
                     /rpt_family="Alu"
repeat_region        18456. .18718
                     /rpt_family="Alu"
repeat_region        complement(18763. .19196)
                     /rpt_family="Alu"
repeat_region        19473. .19548
                     /rpt_family="MER44C"
repeat_region        19689. .19985
                     /rpt_family="MER44C"
repeat_region        complement(20234. .20537)
                     /rpt_family="Alu"
repeat_region        20825. .21104
                     /rpt_family="Alu"
repeat_region        21451. .21758
                     /rpt_family="Alu"
repeat_region        complement(21864. .22509)
                     /rpt_family="Alu"
repeat_region        22190. .22216
                     /note="(T)27"
                     /rpt_type=tandem
                     /rpt_unit=T
repeat_region        22670. .22693
                     /note="(T)24"
                     /rpt_type=tandem
                     /rpt_unit=T
```

```
     repeat_region    complement(22676. .22768)
                      /rpt_family="Alu"
     repeat_region    complement(22814. .22862)
                      /rpt_family="Alu"
     repeat_region    complement(23056. .23307)
                      /rpt_family="Alu"
     repeat_region    complement(23477. .23569)
                      /rpt_family="Alu"
     repeat_region    complement(23803. .24108)
                      /rpt_family="Alu"
     repeat_region    24197. .24502
                      /rpt_family="Alu"
     repeat_region    24485. .24506
                      /note="(A)22"
                      /rpt_type=tandem
                      /rpt_unit=A
     repeat_region    24904. .25177
                      /rpt_family="Alu"
     repeat_region    25370. .25510
                      /rpt_family="L1"
     misc_feature     complement(26535. .26624)
                      /note="GRAIL 2 excellent exon, frame 1"
     repeat_region    complement(26705. .26996)
                      /rpt_family="Alu"
     repeat_region    26958. .27071
                      /rpt_family="THE1"
     repeat_region    complement(28016. .28299)
                      /rpt_family="Alu"
     repeat_region    complement(28658. .28822)
                      /rpt_family="L1"
     repeat_region    complement(29152. .29437)
                      /rpt_family="Alu"
     misc_feature     31894. .32016
                      /note="GRAIL 2 excellent exon, frame 0"
     repeat_region    33793. .33935
                      /rpt_family="Alu"
     repeat_region    complement(35555. .35853)
                      /rpt_family="Alu"
     repeat_region    36311. .36882
                      /rpt_family="Alu"
     repeat_region    37120. .37382
                      /rpt_family="Alu"
     repeat_region    38098. .38833
                      /rpt_family="MER7"
     repeat_region    38352. .38639
                      /rpt_family="Alu"
     repeat_region    38823. .39113
                      /rpt_family="Alu"
     repeat_region    39195. .39433
                      /rpt_family="Alu"
     repeat_region    complement(39516. .39913)
```



```
                      /rpt_family="L1"
     repeat_region    40057. .40692
                      /rpt_family="MER7"
     repeat_region    complement(40128. .40441)
                      /rpt_family="Alu"
     repeat_region    40790. .41056
                      /rpt_family="Alu"
     repeat_region    42301. .42473
                      /rpt_family="MER58B"
     repeat_region    42481. .42760
...
Note: remainder of annotations omitted.

  Query Match           32.8%;  Score 20;  DB 26;  Length 78065;
  Best Local Similarity 100.0%;  Pred. No. 1.22e+00;
  Matches   20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  16992 TTGTTTTATTTTGTTTTTGA 17011
          ||||||||||||||||||||
Cp     34 TTGTTTTATTTTGTTTTTGA 15


RESULT    8
LOCUS       U82668      85132 bp    DNA             PRI       26-JUL-1997
DEFINITION  Homo sapiens shox gene, alternatively spliced products, complete
            cds.
ACCESSION   U82668
NID         g2276452
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 42151 to 74510)
  AUTHORS   Rao,E., Weiss,B., Fukami,M., Rump,A., Niesler,B., Mertz,A.,
            Muroya,K., Binder,G., Kirsch,S., Winkelmann,M., Nordsiek,G.,
            Heinrich,U., Breuning,M.H., Ranke,M.B., Rosenthal,A., Ogata,T. and
            Rappold,G.A.
  TITLE     Pseudoautosomal deletions encompassing a novel homeobox gene cause
            growth failure in idiopathic short stature and Turner syndrome
  JOURNAL   Nature Genet. 16 (1), 54-63 (1997)
  MEDLINE   97285122
REFERENCE   2  (bases 1 to 85132)
  AUTHORS   Rump,A., Rosenthal,A., Drescher,B. and Schattevoy,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-1996) Genome Analysis, Institute Of Molecular
            Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
REFERENCE   3  (bases 1 to 85132)
  AUTHORS   Rump,A., Rosenthal,A., Drescher,B. and Schattevoy,R.
  TITLE     Direct Submission
```



```
  JOURNAL   Submitted (07-MAY-1997) Genome Analysis, Institute Of Molecular
            Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
  REMARK    Sequence update by submitter
COMMENT     On Jul 25, 1997 this sequence version replaced gi:2102705.
FEATURES             Location/Qualifiers
     source          1. .85132
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="Xp22, Pseudoautosomal region 1 on human X
                     chromosome"
     misc_feature    47. .552
                     /note="GC score = 21.40 (506bp)"
                     /note="Region: GC content"
                     /evidence=not_experimental
     repeat_region   700. .739
                     /note="homology = 90.00%, score = 22, counts= 4"
                     /rpt_type=tandem
                     /rpt_unit=aaaaccaaaa
                     /evidence=not_experimental
     repeat_region   complement(922. .1167)
                     /rpt_family="ALU"
                     /evidence=not_experimental
     repeat_region   1537. .1643
                     /rpt_family="ALU"
                     /evidence=not_experimental
     repeat_region   1646. .1818
                     /rpt_family="ALU"
                     /evidence=not_experimental
     repeat_region   2085. .2374
                     /rpt_family="ALU"
                     /evidence=not_experimental
     repeat_region   complement(2522. .2581)
                     /rpt_family="MLT1C"
                     /evidence=not_experimental
     repeat_region   complement(2653. .2716)
                     /rpt_family="MLT1C"
                     /evidence=not_experimental
     repeat_region   complement(2794. .2849)
                     /rpt_family="MLT1C"
                     /evidence=not_experimental
     repeat_region   complement(3019. .3188)
                     /rpt_family="ALU"
                     /evidence=not_experimental
     repeat_region   complement(3189. .3310)
                     /rpt_family="ALU"
                     /evidence=not_experimental
     misc_feature    3411. .4320
                     /note="CpG island score = 0.84, GC = 54.70%,CpGs = 58"
                     /note="Region: CpG island"
                     /evidence=not_experimental
     repeat_region   3416. .3479
```



```
                     /rpt_family="ALU"
                     /evidence=not_experimental
     repeat_region   3480. .3542
                     /rpt_family="ALU"
                     /evidence=not_experimental
     repeat_region   3543. .3703
                     /rpt_family="ALU"
                     /evidence=not_experimental
     misc_feature    3551. .3676
                     /note="GC score = 12.40 (126bp)"
                     /note="Region: GC content"
                     /evidence=not_experimental
     repeat_region   complement(3910. .4133)
                     /rpt_family="ALU"
                     /evidence=not_experimental
     misc_feature    3938. .4063
                     /note="GC score = 8.40 (126bp)"
                     /note="Region: GC content"
                     /evidence=not_experimental
     repeat_region   complement(4134. .4197)
                     /rpt_family="ALU"
                     /evidence=not_experimental
     repeat_region   5196. .5221
                     /note="homology = 100.00%, score = 24, counts= 13"
                     /rpt_type=tandem
                     /rpt_unit=tg
                     /evidence=not_experimental
     repeat_region   6314. .6414
                     /rpt_family="ALU"
                     /evidence=not_experimental
     repeat_region   6438. .6621
                     /rpt_family="ALU"
                     /evidence=not_experimental
     repeat_region   6626. .6708
                     /rpt_family="ALU"
                     /evidence=not_experimental
     repeat_region   6709. .6812
                     /rpt_family="ALU"
                     /evidence=not_experimental
     repeat_region   6813. .6998
                     /rpt_family="ALU"
                     /evidence=not_experimental
     repeat_region   7281. .7688
                     /note="homology = 67.90%, score = 134, counts= 34"
                     /rpt_type=tandem
                     /rpt_unit=ctttttttttt
                     /evidence=not_experimental
     repeat_region   7715. .7780
                     /rpt_family="MER41"
                     /evidence=not_experimental
     repeat_region   complement(7804. .7972)
```

```
                     /rpt_family="ALU"
                     /evidence=not_experimental
     repeat_region   complement(7973. .8102)
                     /rpt_family="ALU"
                     /evidence=not_experimental
     repeat_region   9157. .9252
                     /rpt_family="L1MB7"
                     /evidence=not_experimental
     repeat_region   9483. .9592
                     /rpt_family="SVA"
                     /evidence=not_experimental
     repeat_region   9728. .9819
                     /rpt_family="ALU"
                     /evidence=not_experimental
     repeat_region   9820. .9858
                     /rpt_family="ALU"
                     /evidence=not_experimental
     repeat_region   9865. .10010
                     /rpt_family="ALU"
                     /evidence=not_experimental
     repeat_region   10559. .10583
                     /note="1I1 with 92% homology to 1I2"
                     /rpt_type=inverted
                     /rpt_unit=gcagagagagacagagacagagaga
                     /evidence=not_experimental
     repeat_region   10952. .10991
                     /rpt_family="MLT1C"
                     /evidence=not_experimental
     repeat_region   complement(10994. .11060)
                     /rpt_family="ALU"
                     /evidence=not_experimental
     repeat_region   11081. .11134
                     /rpt_family="MSTA"
                     /evidence=not_experimental
     repeat_region   11136. .11187
                     /rpt_family="THE1B"
                     /evidence=not_experimental
     repeat_region   11397. .11628
                     /note="homology = 70.70%, score = 94, counts= 116"
                     /rpt_type=tandem
                     /rpt_unit=tc
                     /evidence=not_experimental
     repeat_region   11607. .11631
                     /note="1I2 with 92% homology to 1I1"
                     /rpt_type=inverted
                     /rpt_unit=tctctctgtttctgtctgtctctgc
                     /evidence=not_experimental
     repeat_region   11747. .11801
                     /rpt_family="MLT1E"
                     /evidence=not_experimental
     repeat_region   11824. .11909
```



```
                     /rpt_family="MLT1E"
                     /evidence=not_experimental
     repeat_region   complement(12177. .12292)
                     /rpt_family="LTR7"
                     /evidence=not_experimental
     repeat_region   12453. .12588
                     /note="homology = 93.40%, score = 84, counts= 4"
                     /rpt_type=tandem
                     /rpt_unit=ggcgggggaggatattacaaagtaccttctcaag
                     /evidence=not_experimental
     repeat_region   12591. .12692
                     /note="homology = 94.10%, score = 56, counts= 3"
                     /rpt_type=tandem
                     /rpt_unit=gggggaggatattacaaagtaccttctcaagggt
                     /evidence=not_experimental
     repeat_region   12693. .12760
                     /note="homology = 95.60%, score = 28, counts= 2"
                     /rpt_type=tandem
                     /rpt_unit=ggggaggatattacaaagtaccttctcaagggtg
...
Note: remainder of annotations omitted.

  Query Match           32.8%;  Score 20;  DB 26;  Length 85132;
  Best Local Similarity 100.0%;  Pred. No. 1.22e+00;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  52555 TCAAAAACAAAATAAAACAA 52574
          ||||||||||||||||||||
Qy     15 TCAAAAACAAAATAAAACAA 34


RESULT    9
LOCUS       AP000049   100000 bp    DNA             PRI       13-MAY-1998
DEFINITION  Homo sapiens genomic DNA, chromosome 21q11.1, segment 20/28,
            complete sequence.
ACCESSION   AP000049
NID         g3132359
KEYWORDS    HTG.
SOURCE      Homo sapiens DNA, clone:130N6-149C3f_1.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 100000)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Shiba,T. and Sakaki,Y.
  TITLE     Homo sapiens genomic DNA, chromosome 21q
  JOURNAL   Published Only in DataBase (1998) In press
REFERENCE   2  (bases 1 to 100000)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Shiba,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-MAY-1998) to the DDBJ/EMBL/GenBank databases.
            Masahira Hattori, Kitasato University, Department of Science, JST
```



```
            Sequencing Laboratory; Kitasato 1-15-1, Sagamihara 228, Japan
            (E-mail:hattori@hgc.ims.u-tokyo.ac.jp, Tel:0427-78-9732,
            Fax:0427-78-9561)
FEATURES              Location/Qualifiers
     source           1. .100000
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="21"
                      /clone="130N6-149C3f_1"
                      /map="21q22.1"
BASE COUNT    29068 a  18797 c  20533 g  31602 t
ORIGIN

  Query Match           32.8%;  Score 20;  DB 26;  Length 100000;
  Best Local Similarity 100.0%;  Pred. No. 1.22e+00;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  61254 TTGTTTTATTTTGTTTTTGA 61273
          ||||||||||||||||||||
Cp     34 TTGTTTTATTTTGTTTTTGA 15


RESULT   10
LOCUS       AC005087   105984 bp    DNA             HTG       12-JUN-1998
DEFINITION  *** SEQUENCING IN PROGRESS *** Homo sapiens clone RG306D19; HTGS
            phase 1, 17 unordered pieces.
ACCESSION   AC005087
NID         g3212904
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 105984)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 105984)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-1998) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     ***                                                        ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                        ***
            * This sequence is unfinished. It consists of 17 contigs for
            * which the order is not known; their order in this record is
            * arbitrary. In some cases, the exact lengths of the gaps
            * between the contigs are also unknown; these gaps are  presented
            * as runs of N as a convenience only. When sequencing is complete,
```



```
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
            *       1     1848: contig of 1848 bp in length
            *    1849     1865: gap of unknown length
            *    1866     5003: contig of 3138 bp in length
            *    5004     5020: gap of unknown length
            *    5021     7305: contig of 2285 bp in length
            *    7306     7322: gap of unknown length
            *    7323    10299: contig of 2977 bp in length
            *   10300    10316: gap of unknown length
            *   10317    12889: contig of 2573 bp in length
            *   12890    12906: gap of unknown length
            *   12907    15650: contig of 2744 bp in length
            *   15651    15667: gap of unknown length
            *   15668    19509: contig of 3842 bp in length
            *   19510    19526: gap of unknown length
            *   19527    22955: contig of 3429 bp in length
            *   22956    22972: gap of unknown length
            *   22973    26374: contig of 3402 bp in length
            *   26375    26391: gap of unknown length
            *   26392    30938: contig of 4547 bp in length
            *   30939    30955: gap of unknown length
            *   30956    37078: contig of 6123 bp in length
            *   37079    37095: gap of unknown length
            *   37096    45603: contig of 8508 bp in length
            *   45604    45620: gap of unknown length
            *   45621    53337: contig of 7717 bp in length
            *   53338    53354: gap of unknown length
            *   53355    60507: contig of 7153 bp in length
            *   60508    60524: gap of unknown length
            *   60525    69631: contig of 9107 bp in length
            *   69632    69648: gap of unknown length
            *   69649    85278: contig of 15630 bp in length
            *   85279    85295: gap of unknown length
            *   85296   105984: contig of 20689 bp in length.
FEATURES              Location/Qualifiers
     source           1. .105984
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="RG306D19"
BASE COUNT    30851 a  22089 c  23772 g  29000 t    272 others
ORIGIN

  Query Match           32.8%;  Score 20;  DB 17;  Length 105984;
  Best Local Similarity 100.0%;  Pred. No. 1.22e+00;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  80159 TCAAAAACAAAATAAAACAA 80178
          ||||||||||||||||||||
Qy     15 TCAAAAACAAAATAAAACAA 34
```

```
RESULT   11
LOCUS       AC004659   129577 bp    DNA             PRI       02-MAY-1998
DEFINITION  Homo sapiens chromosome 19, CIT-HSP-87m17 BAC clone, complete
            sequence.
ACCESSION   AC004659
NID         g3108020
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 129577)
  AUTHORS   Lamerdin,J.E., McCready,P.M., Skowronski,E., Adamson,A.W.,
            Burkhart-Schultz,K., Gordon,L., Kyle,A., Ramirez,M., Stilwagen,S.,
            Phan,H., Velasco,N., Garnes,J., Danganan,L., Poundstone,P.,
            Christensen,M., Georgescu,A., Avila,J., Liu,S., Attix,C.,
            Andreise,T., Trankheim,M., Amico-Keller,G., Coefield,J., Duarte,S.,
            Lucas,S., Bruce,R., Thomas,P., Quan,G., Kronmiller,B., Arellano,A.,
            Montgomery,M., Ow,D., Nolan,M., Trong,S., Kobayashi,A., Olsen,A.S.
            and Carrano,A.V.
  TITLE     Sequence analysis of an ~1.5 Mb olfactory receptor (OLFR) cluster
            in 19p13.1
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 129577)
  AUTHORS   Lamerdin,J.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-MAY-1998) Joint Genome Institute, Lawrence Livermore
            National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
COMMENT     Map and sequence oriented from p telomere to centromere.  BAC 87m17
            (BC62940) overlaps with cosmid R27726 to the left and there is
            currently a sequence gap of approx. 7.5 kb to the right between BAC
            87m17 and cosmid R31973.  Additional chr 19 map and sequence
            information is available at:
            http://www-bio.llnl.gov/bbrp/genome/genome.html.
FEATURES             Location/Qualifiers
     source          1. .129577
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="87m17"
                     /chromosome="19"
                     /map="19p13.1 between OLFR and MEL"
                     /tissue_type="Sperm"
                     /clone_lib="CIT-HSP"
                     /note="LLNL clone name: BC62940"
     misc_feature    complement(1. .276)
                     /note="BLASTN similarity to U86225 (1. .276); match: 1,
                     score: 1.4e-106; database searched: nt; Homo sapiens
                     olfactory receptor (OR14-12) pseudogene, partial sequence"
     CDS             complement(<1. .476)
                     /note="Hypothetical OLF4-like protein; 81% identical to
```



```
                     (AC002988) OLF4 [Homo sapiens]"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="BC62940_1"
                     /db_xref="PID:g3108021"
                     /translation="MEPENGTRILGFLLLGLSEEPELQPVMFGLFLSMYLTTVFGNLL
                     IILAICSGSHLHTPMYFFLSNLSFVDICVTSTTVPKTLSNIRTQSKVITYAGCITQMY
                     FPVLFIVLDSLLLTVMAYDQFVAICHPLHYTVIVNPRLCGLLVLASWIMSALNSLI"
     repeat_region   complement(731. .1153)
                     /rpt_family="LINE2"
     repeat_region   complement(1914. .2181)
                     /rpt_family="L1MA6"
     repeat_region   2185. .2499
                     /rpt_family="L1"
     repeat_region   complement(2539. .2770)
                     /rpt_family="L1MA2"
     repeat_region   2771. .3071
                     /rpt_family="AluSx"
     repeat_region   complement(3120. .3419)
                     /rpt_family="AluJo"
     repeat_region   3422. .3601
                     /rpt_family="L1MA8"
     repeat_region   complement(3608. .3905)
                     /rpt_family="AluSx"
     repeat_region   3916. .4332
                     /rpt_family="L1MA8"
     repeat_region   complement(4334. .4631)
                     /rpt_family="AluSq"
     repeat_region   4632. .4820
                     /rpt_family="L1MA8"
     repeat_region   5317. .5457
                     /rpt_family="(TA)n"
     misc_feature    5696. .5779
                     /note="predicted exon, program: 0, frame: 2, quality:
                     excellent, score: 100.000"
     repeat_region   complement(5881. .5917)
                     /rpt_family="L1MA8"
     repeat_region   5918. .6111
                     /rpt_family="AluSx"
     repeat_region   complement(6812. .7112)
                     /rpt_family="AluSx"
     repeat_region   complement(7276. .7461)
                     /rpt_family="AluY"
     repeat_region   complement(7907. .8200)
                     /rpt_family="AluSg"
     repeat_region   8476. .8576
                     /rpt_family="MER20"
     repeat_region   complement(9039. .9582)
                     /rpt_family="LTR10A"
     repeat_region   complement(9869. .9983)
                     /rpt_family="(TAAAA)n"
```



```
     misc_feature    10280. .10347
                     /note="predicted exon, program: 0, frame: 1, quality:
                     good, score: 64.000"
     repeat_region   10983. .11048
                     /rpt_family="(CATG)n"
     repeat_region   11304. .11678
                     /rpt_family="LINE2"
     repeat_region   complement(11692. .12035)
                     /rpt_family="AluJb"
     repeat_region   12267. .12594
                     /rpt_family="L1MA8"
     misc_feature    12719. .13683
                     /note="BLASTX similarity to 2447219 (1. .306); match: 0.87,
                     score: 1.2e-173; database searched: nr; (AC002988) OLF4
                     [Homo sapiens]."
     misc_feature    13040. .13186
                     /note="predicted exon, program: 0, frame: 0, quality:
                     excellent, score: 76.000"
     misc_feature    13577. .13683
                     /note="predicted exon, program: 0, frame: 0, quality:
                     good, score: 52.000"
     repeat_region   complement(13880. .13909)
                     /rpt_family="(CAAAA)n"
     repeat_region   complement(14051. .14346)
                     /rpt_family="AluSq"
     repeat_region   14587. .14773
                     /rpt_family="AluSg"
     repeat_region   14814. .15172
                     /rpt_family="THE1B"
     repeat_region   complement(15242. .15364)
                     /rpt_family="AluSp"
     repeat_region   complement(15755. .15985)
                     /rpt_family="L1PA8"
     repeat_region   complement(16725. .17020)
                     /rpt_family="AluY"
     repeat_region   complement(17025. .17083)
                     /rpt_family="LINE2"
     repeat_region   17401. .17702
                     /rpt_family="AluJb"
     repeat_region   complement(18211. .18513)
                     /rpt_family="AluSx"
     repeat_region   complement(18804. .18905)
                     /rpt_family="(TAAAA)n"
     repeat_region   19258. .19534
                     /rpt_family="AluSq"
     repeat_region   complement(19859. .19895)
                     /rpt_family="POLY_A"
     repeat_region   complement(19896. .20207)
                     /rpt_family="L1PB1"
     repeat_region   20208. .20509
                     /rpt_family="AluSq"
```



```
     repeat_region   complement(20512. .20556)
                     /rpt_family="L1PB1"
     repeat_region   complement(20669. .20698)
                     /rpt_family="AT_rich"
     repeat_region   complement(20852. .20900)
                     /rpt_family="(CA)n"
     repeat_region   21097. .21394
                     /rpt_family="AluY"
     repeat_region   21555. .21854
                     /rpt_family="AluSx"
     repeat_region   21929. .22103
                     /rpt_family="LINE2"
     repeat_region   complement(22121. .22535)
                     /rpt_family="L1PA13"
     repeat_region   22554. .22779
                     /rpt_family="LINE2"
     repeat_region   complement(22814. .23116)
                     /rpt_family="AluSp"
     misc_feature    23355. .24339
                     /note="BLASTX similarity to 2447219 (1. .308); match: 0.67,
                     score: 5.6e-168; database searched: nr; (AC002988) OLF4
                     [Homo sapiens]"
     repeat_region   complement(23434. .23484)
                     /rpt_family="U5"
     misc_feature    23615. .24264
                     /note="BLASTN similarity to U86253 (1. .650); match: 0.99,
                     score: 3.3e-262; database searched: nt; Homo sapiens
                     olfactory receptor (OR19-134) pseudogene, partial
                     sequence"
     repeat_region   complement(24458. .24757)
                     /rpt_family="AluY"
     repeat_region   complement(25723. .26220)
                     /rpt_family="L1M3/4"
     repeat_region   26234. .26308
                     /rpt_family="(TAAA)n"
     repeat_region   complement(26314. .26515)
                     /rpt_family="MER2"
     repeat_region   26765. .26852
                     /rpt_family="AluJ/FLAM"
     repeat_region   complement(26858. .27503)
                     /rpt_family="PABL_A"
     repeat_region   27504. .27563
                     /rpt_family="FLAM_A"
     repeat_region   27932. .28232
                     /rpt_family="AluSx"
     misc_feature    28545. .28766
                     /note="predicted exon, program: 0, frame: 2, quality:
                     good, score: 64.000"
     repeat_region   28996. .29041
...
Note: remainder of annotations omitted.
```

```
  Query Match          32.8%;  Score 20;  DB 26;  Length 129577;
  Best Local Similarity 100.0%;  Pred. No. 1.22e+00;
  Matches   20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  98391 TTATTGTTTTATTTTGTTTT 98410
          ||||||||||||||||||||
Cp     37 TTATTGTTTTATTTTGTTTT 18


RESULT   12
LOCUS       AC005043   193869 bp    DNA              HTG       12-JUN-1998
DEFINITION  *** SEQUENCING IN PROGRESS *** Homo sapiens clone NH0576N21; HTGS
            phase 1, 5 unordered pieces.
ACCESSION   AC005043
NID         g3212948
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 193869)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 193869)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-1998) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     ***                                                          ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                          ***
            * This sequence is unfinished. It consists of 5 contigs for
            * which the order is not known; their order in this record is
            * arbitrary. In some cases, the exact lengths of the gaps
            * between the contigs are also unknown; these gaps are  presented
            * as runs of N as a convenience only. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
            *      1     2805: contig of 2805 bp in length
            *   2806     2823: gap of unknown length
            *   2824    57368: contig of 54545 bp in length
            *  57369    57386: gap of unknown length
            *  57387   190422: contig of 133036 bp in length
            *  190423  190439: gap of unknown length
            *  190440  192126: contig of 1687 bp in length
            *  192127  192143: gap of unknown length
            *  192144  193869: contig of 1726 bp in length.
FEATURES            Location/Qualifiers
```



```
     source          1. .193869
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="NH0576N21"
BASE COUNT    50036 a  42256 c  46413 g  54305 t    859 others
ORIGIN

  Query Match          32.8%;  Score 20;  DB 17;  Length 193869;
  Best Local Similarity 100.0%;  Pred. No. 1.22e+00;
  Matches   20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   9531 TTGTTTTATTTTGTTTTTGA 9550
          ||||||||||||||||||||
Cp     34 TTGTTTTATTTTGTTTTTGA 15


RESULT   13
LOCUS       AC003101   208396 bp    DNA              PRI       05-JUN-1998
DEFINITION  Homo sapiens chromosome 17, clone HRPC41C23, complete sequence.
ACCESSION   AC003101
NID         g3184508
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 208396)
  AUTHORS   Birren,B., Fasman,K., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone HRPC41C23
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 208396)
  AUTHORS   Birren,B., Fasman,K., McKernan,K., Munro,C., Nusbaum,C.,
            Richardson,P., Lander,E., Baldwin,J., Barna,N., Cantu,C., Chang,A.,
            Cooke,P., Daly,M.J., Depayre,E., Devon,K., Dewar,K., DuRette,B.,
            Forrest,C., Gage,D., Gensheimer,S., Geraigery,K., Gilmartin,T.,
            Hagos,B., Halphen,I., Harris,K., Howland,J.C., Huang,J., Hui,L.,
            Jacotot,L., Linton,L., MacKenzie,J., Marquis,N., McDermott,J.,
            McGurk,A., Molla,M., Morris,W., Morrow,J., Nachman,A., Naylor,J.,
            O'Connor,T., Olotu,A., Peterson,K., Roberts,D., Rollins,G.,
            Sarnaik,A., Shiu,P., Shyam,R., Stilwell,J., Stone,C.,
            Strickland,C., Sydney,K., Tang,L., Zemtseva,I., Zhao,J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-NOV-1997) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 208396)
  AUTHORS   Birren,B., Fasman,K., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boatin,C.,
            Boutwell,C., Brown,A., Castle,A., Cerny,J., Cooke,P., Depayre,E.,
            Devon,K., Dewar,K., Donelan,L., Etemadi,S., Ferreira,P.,
            FitzHugh,W., Forrest,C., Funke,R., Gage,D., Gardyna,S.,
            Gensheimer,S., Geraigery,K., Gilmartin,T., Grant,G., Hagos,B.,
```



```
            Harris,K., Horton,L., Howland,J.C., Hui,L., Jacotot,L., Kann,L.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Nachman,A., Nahf,R., Naylor,J., O'Connor,T., Pavlin,B.,
            Peterson,K., Riley,R., Roberts,D., Rossello,R., Roy,A., Shyam,R.,
            Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C.,
            Strickland,C., Subramanian,A., Torruella-Miller,I., Vassiliev,H.,
            Vo,A., Wagner,A., Wang,B., Wheeler,J., Wu,Y., Ye,W.J., Zhao,J. and
            Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JUN-1998) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3168631.
            All repeats were identified using RepeatMasker: Smit, A.F.A. &
            Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html.
FEATURES            Location/Qualifiers
     source         1. .208396
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="HRPC41C23"
                    /clone_lib="Peter de Jong/ human PAC library"
                    /map="17"
                    /chromosome="17"
     repeat_region  27. .209
                    /rpt_family="MIR"
     repeat_region  complement(378. .403)
                    /rpt_family="(CAT)n"
     repeat_region  489. .734
                    /rpt_family="AluSx"
     repeat_region  1139. .1440
                    /rpt_family="AluY"
     repeat_region  2566. .2634
                    /rpt_family="MIR"
     repeat_region  complement(2905. .2964)
                    /rpt_family="MIR"
     repeat_region  complement(4148. .4257)
                    /rpt_family="MIR"
     repeat_region  complement(5399. .5473)
                    /rpt_family="MIR"
     repeat_region  complement(5674. .5787)
                    /rpt_family="MIR"
     repeat_region  complement(6063. .6120)
                    /rpt_family="MIR"
     repeat_region  complement(6143. .6229)
                    /rpt_family="MER45"
     repeat_region  7875. .8160
                    /rpt_family="AluJb"
     repeat_region  8163. .8465
                    /rpt_family="AluSq"
     repeat_region  complement(8521. .8826)
```



```
                    /rpt_family="AluJb"
     repeat_region  complement(8869. .9295)
                    /rpt_family="L1ME1"
     repeat_region  complement(9303. .9607)
                    /rpt_family="L1M4"
     repeat_region  complement(9608. .9907)
                    /rpt_family="AluY"
     repeat_region  complement(9909. .10262)
                    /rpt_family="L1M4"
     repeat_region  complement(10288. .10591)
                    /rpt_family="AluJo"
     repeat_region  complement(10592. .10820)
                    /rpt_family="L1M4"
     repeat_region  10861. .11163
                    /rpt_family="AluSp"
     repeat_region  complement(11168. .11260)
                    /rpt_family="L1M4"
     repeat_region  complement(11280. .11573)
                    /rpt_family="AluSx"
     repeat_region  complement(11615. .11917)
                    /rpt_family="AluSq"
     repeat_region  11957. .12251
                    /rpt_family="AluJo"
     repeat_region  complement(12253. .12681)
                    /rpt_family="L1M4"
     repeat_region  complement(12701. .12956)
                    /rpt_family="L1ME"
     repeat_region  complement(12976. .13276)
                    /rpt_family="AluY"
     repeat_region  complement(13277. .13417)
                    /rpt_family="L1ME"
     repeat_region  complement(13424. .13606)
                    /rpt_family="AluJb"
     repeat_region  complement(13657. .13911)
                    /rpt_family="L1ME"
     repeat_region  complement(14096. .14403)
                    /rpt_family="L1MD"
     repeat_region  complement(14527. .14634)
                    /rpt_family="AluSc"
     repeat_region  complement(14635. .14933)
                    /rpt_family="AluSx"
     repeat_region  15106. .15347
                    /rpt_family="MIR"
     repeat_region  complement(16322. .16354)
                    /rpt_family="AT_rich"
     repeat_region  16374. .16492
                    /rpt_family="MIR"
     repeat_region  complement(16657. .16957)
                    /rpt_family="AluSq"
     repeat_region  complement(17013. .17317)
                    /rpt_family="AluSx"
```

```
     repeat_region   18703. .18766
                     /rpt_family="L2"
     repeat_region   18855. .19009
                     /rpt_family="FRAM"
     repeat_region   complement(19035. .19146)
                     /rpt_family="AluSp/q"
     repeat_region   complement(19177. .19309)
                     /rpt_family="L2"
     repeat_region   19603. .19906
                     /rpt_family="AluSx"
     repeat_region   complement(20121. .20414)
                     /rpt_family="L2"
     repeat_region   complement(20415. .20717)
                     /rpt_family="AluY"
     repeat_region   complement(20719. .20808)
                     /rpt_family="L2"
     repeat_region   complement(21025. .21151)
                     /rpt_family="L2"
     repeat_region   22253. .22356
                     /rpt_family="MER81"
     repeat_region   complement(22414. .22718)
                     /rpt_family="AluJb"
     repeat_region   22901. .23218
                     /rpt_family="AluJb"
     repeat_region   24532. .24942
                     /rpt_family="MER21B"
     repeat_region   24963. .25259
                     /rpt_family="AluJo"
     repeat_region   25269. .25384
                     /rpt_family="MER21B"
     repeat_region   26217. .26321
                     /rpt_family="MIR"
     repeat_region   complement(26787. .26834)
                     /rpt_family="(CAAAA)n"
     repeat_region   complement(26835. .27117)
                     /rpt_family="AluY"
     repeat_region   28880. .29005
                     /rpt_family="FLAM_C"
     repeat_region   complement(29274. .29658)
                     /rpt_family="L1M4"
     repeat_region   complement(29660. .29977)
                     /rpt_family="AluSx"
     repeat_region   complement(30027. .30526)
                     /rpt_family="L1M4"
     repeat_region   30536. .30837
                     /rpt_family="AluSx"
     repeat_region   complement(30938. .31065)
                     /rpt_family="L1M4"
     repeat_region   complement(31068. .31374)
                     /rpt_family="AluSq"
     repeat_region   complement(31576. .31876)
```



```
                     /rpt_family="AluSq"
     repeat_region   complement(32053. .32347)
                     /rpt_family="AluY"
     repeat_region   complement(32460. .32677)
                     /rpt_family="MER58A"
     repeat_region   32727. .33028
                     /rpt_family="AluSx"
     repeat_region   33679. .33985
                     /rpt_family="AluSx"
     repeat_region   complement(35256. .35582)
                     /rpt_family="L1MC4"
     repeat_region   complement(35602. .35906)
                     /rpt_family="AluSg"
...
Note: remainder of annotations omitted.

  Query Match           32.8%;  Score 20;  DB 26;  Length 208396;
  Best Local Similarity 100.0%;  Pred. No. 1.22e+00;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db 185423 TTGTTTTATTTTGTTTTTGA 185442
          ||||||||||||||||||||
Cp     34 TTGTTTTATTTTGTTTTTGA 15


RESULT   14
LOCUS       SAU25696      882 bp    mRNA            PLN       29-JAN-1997
DEFINITION  Sinapis alba transcription factor SaMADS A mRNA, complete cds.
ACCESSION   U25696
NID         gl199574
KEYWORDS    .
SOURCE      white mustard.
  ORGANISM  Sinapis alba
            Eukaryotae; mitochondrial eukaryotes; Viridiplantae;
            Charophyta/Embryophyta group; Embryophyta; Magnoliophyta;
            Magnoliopsida; Capparales; Brassicaceae; Sinapis.
REFERENCE   1  (bases 1 to 882)
  AUTHORS   Menzel,G., Apel,K. and Melzer,S.
  TITLE     Identification of two MADS box genes that are expressed in the
            apical meristem of the long-day plant Sinapis alba in transition to
            flowering
  JOURNAL   Plant J. 9 (3), 399-408 (1996)
  MEDLINE   97077349
REFERENCE   2  (bases 1 to 882)
  AUTHORS   Menzel,G., Apel,K. and Melzer,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-APR-1995) Siegbert Melzer, Institute for Plant
            Sciences, Federal Institute of Technology, Universitaetstr. 2,
            Zuerich 8092, Switzerland
COMMENT     On Feb 22, 1996 this sequence version replaced gi:1049021.
FEATURES             Location/Qualifiers
```



```
     source          1. .882
                     /organism="Sinapis alba"
                     /note="gene expressed in apical meristems in transition to
                     flowering"
                     /db_xref="taxon:3728"
                     /tissue_type="inflorescence meristems"
     CDS             28. .669
                     /note="MADS box protein"
                     /codon_start=1
                     /product="transcription factor SaMADS A"
                     /db_xref="PID:gl049022"
                     /translation="MVRGKTQMKRIENATSRQVTFSKRRNGLLKKAFELSVLCDAEVS
                     LIIFSPKGKLYEFASSNMQDTVDRYLRHTKDRVSSKPVSEENMQHFKHEAANMKKIE
                     QLEASKRKLLGEGIGSCSIEELQQIEQQLEKSVKCVRARKTQVFKEQIEQLKQKEKAL
                     AAENEKLAEKWGSHEIEVWSNKKQESGRGDEESSPSSEVETQLFIGLPCSSRK"
     polyA_site      882
                     /note="20 A nucleotides"
BASE COUNT      315 a    159 c    195 g    213 t
ORIGIN

  Query Match           31.1%;  Score 19;  DB 23;  Length 882;
  Best Local Similarity 100.0%;  Pred. No. 4.66e+00;
  Matches    19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    693 CAAAAACAAAATAAAACAA 711
          |||||||||||||||||||
Qy     16 CAAAAACAAAATAAAACAA 34


RESULT   15
LOCUS       MCAAX        1256 bp    DNA             BCT       19-AUG-1994
DEFINITION  M.capricolum DNA for CONTIG MCAAX.
ACCESSION   Z33360
NID         g534021
KEYWORDS    repetitive element.
SOURCE      Mycoplasma capricolum.
  ORGANISM  Mycoplasma capricolum
            Eubacteria; Firmicutes; Low G+C gram-positive bacteria; Mycoplasmas
            and walled relatives; Mycoplasmatales; capricolum group.
REFERENCE   1  (bases 1 to 1256)
  AUTHORS   Bork,P., Ouzounis,C., Casari,G., Sander,C., Dolan,M. and
            Gillevet,P.
  TITLE     Exploring the Mycoplasma capricolum genome: A Parasite reveals it's
            physiology
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1256)
  AUTHORS   Gillevet,P., Ally,A., Barton,F., Brenner,S.E., Clark-Whitehead,R.,
            Dolan,M., Douglas,N., Hsu,E., Purzycki,M.S., Richter,B., Russo,S.,
            Sartell,J., Smith,S.W., Wang,C., Williams,J. and Gilbert,W.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-AUG-1994) Mycoplasma capricolum Genome Project,
```



```
            Harvard University, Cambridge, MA 02138 Prokaryotic Genomes Project
            Institute for Computational Sciences and Informatics George Mason
            University, Fairfax, Virginia, USA, 22030-4444 E-mail:
            gillevet@uranus.nchgr.nih.gov
REFERENCE   3  (bases 1 to 1256)
  AUTHORS   Bork,P., Ouzounis,C., Casari,G., Schneider,R., Sander,C., Dolan,M.,
            Gilbert,W. and Gillevet,P.M.
  TITLE     Exploring the Mycoplasma capricolum genome: a minimal cell reveals
            its physiology
  JOURNAL   Mol. Microbiol. 16 (5), 955-967 (1995)
  MEDLINE   96059641
COMMENT     NOTES:
            The tga codon is read as Tryptophan in Mycoplasma capricolum Coding
            sequences below were putatively identified by Peer Bork using
            the program Genequiz at the EMBL.
            EMBL, Meyerhofstr.1, 69012 Heidelberg, Germany
            E-mail bork@embl-heidelberg.de
            More than 870,477 bases have been sequenced by the genomic walking
            technique and assembled into 372 non-overlapping contigs (Accession
            numbers Z33005 to Z33376) covering 214,528 base pairs. The length
            of the contigs varies from 63 base pairs to 2049 base pairs with an
            average length of 658 base pairs. The data consists of
            13,091 bases (6%) from one pass fluorescent sequencing and 201,437
            bases (94%) produced by genomic walking. Of the latter, 155,805
            bases (77%) have multiple coverage on at least one strand and
            112,621 bases (56%) are covered on both strands. There is a total
            of
            968 ambiguous nucleotide assignments in the data set (
            0.5%). We have compared 8,868 bases of our data with Mycoplasma
            capricolum sequences already stored in the public databases and
            note less than 0.7% difference between the two data sets (
            including ambiguous calls, insertions, deletions and mismatches ).
FEATURES             Location/Qualifiers
     source          1. .1256
                     /organism="Mycoplasma capricolum"
                     /strain="ATCC 27343(Kid)"
                     /db_xref="taxon:2095"
     repeat_region   <1. .>1256
BASE COUNT      394 a    198 c    136 g    518 t     10 others
ORIGIN

  Query Match           31.1%;  Score 19;  DB 15;  Length 1256;
  Best Local Similarity 100.0%;  Pred. No. 4.66e+00;
  Matches    19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    177 CATTCTTTTCAAAAACAAA 195
          |||||||||||||||||||
Qy      7 CATTCTTTTCAAAAACAAA 25


RESULT   16
```

```
LOCUS       HSA010069    2746 bp    mRNA            PRI       26-AUG-1998
DEFINITION  Homo sapiens mRNA for HMGBCG protein.
ACCESSION   AJ010069
NID         g3483012
KEYWORDS    HMG box; HMGBCG protein.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2746)
  AUTHORS   Seroussi,E., Kedra,D., Kost-Alimova,M., Sandberg-Nordqvist,A.C.,
            Fransson,I., Fu,Y., Pan,H.Q., Roe,B.A., Imreh,S. and Dumanski,J.P.
  TITLE     The human TOM1 gene and its mouse ortholog belong to a novel
            syntenic group between human chromosome 22q13.1 and mouse
            chromosome 8C1 and encode proteins with N-terminal domain similar
            to that of HRS and STAM endosomal proteins
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2746)
  AUTHORS   Kedra,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-AUG-1998) Kedra D., Department of Molecular Medicine,
            Karolinska Hospital, CMM building, L8:00, Stockholm, S-171 76,
            SWEDEN
FEATURES             Location/Qualifiers
     source          1. .2746
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="22"
                     /map="q13.1"
     gene            1. .2746
                     /gene="HMGBCG"
     CDS             <1. .591
                     /gene="HMGBCG"
                     /codon_start=1
                     /product="HMGBCG protein"
                     /db_xref="PID:e1317752"
                     /db_xref="PID:g3483013"
                     /translation="PKKKNMSAYQVFCKEYRVTIVADHPGIDFGELSKKLAEVWKQLP
                     EKDKLIWKQKACYLQHKQNKAEATTVKRKASSSEGSMKVKASSVGVLSPQKKSPPTTM
                     LLPASPAKAPETEPIDVAAHLQLLGESLSLIGHRLQETEGMVAVSGSLSVLLDSIICA
                     LGPLACLTTQLPELNGCPKQVLSNTLDNIAYIMPGL"
     exon            1. .82
                     /gene="HMGBCG"
                     /number=1
                     /product="HMGBCG protein"
     exon            83. .147
                     /gene="HMGBCG"
                     /number=2
                     /evidence=experimental
                     /product="HMGBCG protein"
     exon            148. .253
```



```
                     /gene="HMGBCG"
                     /number=3
                     /evidence=experimental
                     /product="HMGBCG protein"
     exon            254. .423
                     /gene="HMGBCG"
                     /number=4
                     /evidence=experimental
                     /product="HMGBCG protein"
     exon            424. .546
                     /gene="HMGBCG"
                     /number=5
                     /evidence=experimental
                     /product="HMGBCG protein"
     exon            547. .2746
                     /gene="HMGBCG"
                     /number=6
                     /product="HMGBCG protein"
     polyA_signal    2698. .2703
                     /gene="HMGBCG"
                     /note="AATAAA"
     polyA_signal    2720. .2725
                     /gene="HMGBCG"
                     /note="ATTAAA"
BASE COUNT      767 a    558 c    551 g    870 t
ORIGIN
```

```
  Query Match           31.1%;  Score 19;  DB 27;  Length 2746;
  Best Local Similarity 100.0%;  Pred. No. 4.66e+00;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2629 ATTGTTTTATTTTGTTTTT 2647
           ||||||||||||||||||
Cp     35 ATTGTTTTATTTTGTTTTT 17
```

```
RESULT   17
LOCUS       TTU46561     3365 bp    DNA             INV       16-JAN-1997
DEFINITION  Tetrahymena thermophila polyubiquitin (TTU3) gene, complete cds,
            and RNA polymerase II subunit 2 (RPB2) gene, partial cds.
ACCESSION   U46561
NID         g1778711
KEYWORDS    .
SOURCE      Tetrahymena thermophila strain=SB1917.
  ORGANISM  Tetrahymena thermophila
            Eukaryotae; mitochondrial eukaryotes; Alveolata; Ciliophora;
            Oligohymenophorea; Hymenostomatida; Tetrahymenina; Tetrahymena.
REFERENCE   1  (bases 1 to 3365)
  AUTHORS   Guerreiro,P. and Rodrigues-Pousada,C.
  TITLE     Characterization of a polyubiquitin gene in T. thermophila and of
            ubiquitin gene expression during sexual reproduction and under
```



```
            stress conditions
  JOURNAL   Gene 182 (1-2), 183-188 (1996)
  MEDLINE   97136709
REFERENCE   2  (bases 1 to 3365)
  AUTHORS   Guerreiro,P. and Rodrigues-Pousada,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JAN-1996) Paulo Guerreiro, Laboratorio de Genetica
            Molecular, Instituto Gulbenkian de Ciencia, Apartado 14, Oeiras
            2781, Portugal
FEATURES             Location/Qualifiers
     source          1. .3365
                     /organism="Tetrahymena thermophila"
                     /strain="SB1917"
                     /db_xref="taxon:5911"
     promoter        1. .8
                     /note="heat shock element"
     misc_feature    100. .111
                     /note="AGAGA motif"
     gene            253. .1398
                     /gene="TTU3"
     CDS             253. .1398
                     /gene="TTU3"
                     /codon_start=1
                     /transl_table=6
                     /product="polyubiquitin"
                     /db_xref="PID:g1778712"
                     /translation="MQIFVKTLTGKTITLDVEASDTIENVKAKIQDKEGIPPDQQRLI
                     FAGKQLEDGRTLSDYNIQKESTLHLVLRLRGGMQIFVKTLTGKTITLDVEASDTIENV
                     KAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLRLRGGMQIFVKTL
                     TGKTITLDVEASDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYNIQKEST
                     LHLVLRLRGGMQIFVKTLTGKTITLDVEASDTIENVKAKIQDKEGIPPDQQRLIFAGK
                     QLEDGRTLSDYNIQKESTLHLVLRLRGGMQIFVKTLTGKTITLDVEASDTIENVKAKI
                     QDKEGIPPDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLRLRGGF"
     gene            complement(2926. .>3365)
                     /gene="RPB2"
     CDS             complement(join(2927. .3192,3260. .>3365))
                     /gene="RPB2"
                     /codon_start=1
                     /transl_table=6
                     /product="RNA polymerase II subunit 2"
                     /db_xref="PID:g1778713"
                     /translation="SRGPVVAIVRQPTHGRSRGGGLRFGEMERDCIIAHGTSNFLKER
                     TFGVSDQFRTHVCSRCGLFAKADVTNHKYDCTGCSQGDKKHTKIVQVYLPYAAKQLIQ
                     ELAAMHIVPRIKVDVTSKKYV"
     intron          complement(3193. .3259)
                     /gene="RPB2"
BASE COUNT     1295 a    529 c    402 g   1139 t
ORIGIN
```

```
  Query Match           31.1%;  Score 19;  DB 18;  Length 3365;
  Best Local Similarity 100.0%;  Pred. No. 4.66e+00;
```



```
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2712 TTTTCAAAAACAAAATAAA 2730
          ||||||||||||||||||
Qy     12 TTTTCAAAAACAAAATAAA 30
```

```
RESULT   18
LOCUS       CEZK1058     29101 bp    DNA             INV       22-SEP-1998
DEFINITION  Caenorhabditis elegans cosmid ZK1058, complete sequence.
ACCESSION   Z35604
NID         g1067078
KEYWORDS    HTG; Beta integrin; Galactose-1-phosphate uridylyl transferase;
            Methylmalonyl-coA mutase; Nitrilase.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 29101)
  AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
  TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
  JOURNAL   Nature 368 (6466), 32-38 (1994)
  MEDLINE   94150718
REFERENCE   2  (bases 1 to 29101)
  AUTHORS   Mortimore,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-NOV-1996) Nematode Sequencing Project, Sanger Centre,
            Hinxton, Cambridge CB10 1RQ, England and Department of Genetics,
            Washington University, St. Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     On Nov 21, 1995 this sequence version replaced gi:521060.
            Coding sequences below are predicted from computer analysis, using
            predictions from Genefinder (P. Green, U. Washington), and other
            available information.
            For a graphical representation of this sequence and its analysis
            see:-
            http://webace.sanger.ac.uk/cgi-
            bin/display?db=wormace&class=Sequence &object=ZK1058
            Current sequence finishing criteria for the C. elegans genome
            sequencing consortium are that all bases are either sequenced
```

unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.
IMPORTANT: This sequence is not the entire insert of clone ZK1058. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.
The true left end of clone ZK1058 is at 21854 in sequence Z49936.
The true right end of clone ZK1058 is at 219 in sequence Z35600.
The true left end of clone F37A8 is at 21942 in this sequence. The true right end of clone F13B10 is at 7902 in this sequence. The start of this sequence (1. .7902) overlaps with the end of sequence Z49936.
The end of this sequence (28767. .29101) overlaps with the start of sequence Z35600.

FEATURES Location/Qualifiers

source 1. .29101
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="III"
/clone="ZK1058"

gene complement(5854. .10115)
/gene="ZK1058.1"

CDS complement(join(5854. .6031,6081. .6367,7057. .7509, 8042. .8299,8344. .8992,9706. .10115))
/gene="ZK1058.1"
/note="similar to Methylmalonyl-coA mutase; cDNA EST EMBL:M89407 comes from this gene; cDNA EST CEMSB12F comes from this gene; cDNA EST EMBL:D32709 comes from this gene; cDNA EST EMBL:D32887 comes from this gene; cDNA EST EMBL:D33969 comes from this gene; cDNA EST EMBL:D33036 comes from this gene; cDNA EST EMBL:D35424 comes from this gene; cDNA EST EMBL:D35591 comes from this gene; cDNA EST EMBL:D36948 comes from this gene; cDNA EST EMBL:D35766 comes from this gene; cDNA EST EMBL:C10217 comes from this gene; cDNA EST EMBL:C11775 comes from this gene; cDNA EST EMBL:C10279 comes from this gene; cDNA EST EMBL:C11863 comes from this gene; cDNA EST yk208e8.3 comes from this gene; cDNA EST yk208e8.5 comes from this gene; cDNA EST yk193a9.3 comes from this gene; cDNA EST yk193a9.5 comes from this gene"
/codon_start=1
/db_xref="PID:g521061"
/db_xref="SWISS-PROT:Q23381"
/translation="MYLQLLKPTLLRCSTRPSSSGAYTRSPIDQKWAAMAKKAMKGRE

ATTLTWNTPEGIPIKPLYLRSDRDCDAQRSVELPGQFPPTRGPYPTNYTQRPWTIRQI
AGFSTVEESNKFYKENIKAGQQGLSVAPDLATHRGYDSDNPRVPGDVGNAGVAVDSVE
DMRQLFDGINLEKMSVSMTMNGAVVPVLAMYVVAAEEAGVSRKLLAGTIQNDILKEFM
VRNTYIYPPEPSMRIIGDIFAYTSREMPKFNSISISGYHMQEAGADAVLEMAFTIADG
IQYCETGLNAGLTIDAFAPRLSFFWGISMNFYMEIAKMRAARRLWANLIKERFSPRSD
KSMMLRTHSQTSGWSLTPQDPYNNIGTTIEAMASVFGGTQSLHTNSFDEALGLPTKF
SARIARNTQIIIQEESGICNVADPWGGSYMMESLTDEIYEKALAVIKEIDELGGMAKA
VASGMTKLKIEEAAAKKQARIDAGKDVIVGVNKYRLDHEQQVEVLKIDNAKVREEQCA
KLNHIRATRDAEKAQKALDAITEGARGNGNLMELAIEAARARCTVGEISDAMEKVFNR
HAAVNRLVSGAYKSEFGETSEMSQVLERVKSFADRDGRQPRIMVAKMGQDGHDRGAKV
IATGFADLGFDVDVGPLFQTPLEAAQQAVDADVHVIGASSLAAGHLTLIPQLIGELKK
LGRPDILVVAGGVIPPQDYKELYDAGVALVPGPGTRLPACANQILEKLEANLPEAPGK
AASR"

gene complement(11785. .16585)
/gene="pat-3"

CDS complement(join(11785. .11856,11910. .12196,12630. .13117, 13677. .14015,14572. .15008,15107. .15510,16128. .16332, 16388. .16585))
/gene="pat-3"
/note="similar to beta integrin; cDNA EST EMBL:D64780 comes from this gene; cDNA EST EMBL:D32673 comes from this gene; cDNA EST EMBL:D33876 comes from this gene; cDNA EST EMBL:D33734 comes from this gene; cDNA EST EMBL:D36820 comes from this gene; cDNA EST EMBL:D35368 comes from this gene; cDNA EST EMBL:D36628 comes from this gene; cDNA EST EMBL:D67312 comes from this gene; cDNA EST EMBL:D74646 comes from this gene; cDNA EST EMBL:D67906 comes from this gene; cDNA EST EMBL:D64771 comes from this gene; cDNA EST EMBL:D67922 comes from this gene; cDNA EST EMBL:D68258 comes from this gene; cDNA EST EMBL:D64978 comes from this gene; cDNA EST EMBL:D68350 comes from this gene; cDNA EST EMBL:C10814 comes from this gene; cDNA EST EMBL:C12687 comes from this gene; cDNA EST EMBL:C12596 comes from this gene"
/codon_start=1
/db_xref="PID:g521062"
/db_xref="SPTREMBL:Q27874"
/translation="MPPSTSLLLLAALLPFALPASDWKTGEVTGKVVEKSEFPCYSLS RDNYTCSACIQYHESCAWCGAPMFDEKKPYARCDSRAKLMEHGCPNSYIEDPATKLDI TEDSKLSDQGQVESEEEAVQIKPQEMYVEIRPKSRVRFNVTYRQAVDYPVDLYYLMDL SYSMKDDKQKLSELGDLLAERMRTVTKNFRLGFGSFIDKKLMPFIDPRIEKQLSPCPT PCAEPYGFKHQMSLTTNTAKFKAEVDKAEISGNLDAPEGGFDAVVQALACNKTIGWRE RARRMIVFSTDAGFHFAGDGRLAGVVEPNDGTCHLDREGYYTETLNQDIYPSIALLHQM IKDRKANVIFAVTKNNQDLYTQLSNALPDVSSSVGVLANDSRNIVDLIEKEYLRISEK IIMVDNANASDGLKLTYRSMCLDGTTLKDTNVCEGIRVGDEVQFEVTLENTHCIDKRD PVLRIGPSGLDETLIVEVKVLCDCDCERQDRIVTNSADCNGGDMVCGVCRCKGGNVGK YCECNNRPGMSTAALNEKCRRTNESAICEGRGVCNCGRCECNPRANPEEQISGEPCECD NFNCPRHDRKICAEHGECNCGKCICAPGWTGRACECPISTDSCLSANGKICNGKGECI CGRCRCFDSPDGNRYSGAKCEICPTCPTKCVEYKNCVMCQQWQTGPLNETACDQCEFK VIPVEELPNLNETTPCQFVDPADDCTFYYLIYYDEATDNATVWVRKHKDCPPPVPVLA IVLGVIAGIVILGILLLLLWKLLTVLHDRSEYATFNNERLMAKWDTNENPIYKQATTT

FKNPVYAGKAN"

gene 19452. .20982
/gene="ZK1058.4"

CDS join(19452. .19607,19661. .19951,20000. .20119,20165. .20763, 20820. .20982)
/gene="ZK1058.4"
/note="cDNA EST EMBL:Z14593 comes from this gene; cDNA EST EMBL:T01764 comes from this gene; cDNA EST EMBL:D37668 comes from this gene; cDNA EST EMBL:D73892 comes from this gene; cDNA EST EMBL:D71243 comes from this gene; cDNA EST EMBL:D71512 comes from this gene; cDNA EST EMBL:C07506 comes from this gene; cDNA EST EMBL:C08880 comes from this gene; cDNA EST EMBL:C08452 comes from this gene; cDNA EST yk505b9.3 comes from this gene; cDNA EST yk462g11.3 comes from this gene; cDNA EST yk462g11.5 comes from this gene; cDNA EST yk460h1.3 comes from this gene; cDNA EST yk460h1.5 comes from this gene; cDNA EST yk467b7.3 comes from this gene; cDNA EST yk467b7.5 comes from this gene; cDNA EST yk452a3.3 comes from this gene; cDNA EST yk452a3.5 comes from this gene; cDNA EST yk427b5.3 comes from this gene; cDNA EST yk427b5.5 comes from this gene; cDNA EST yk404h4.3 comes from this gene; cDNA EST yk404h4.5 comes from this gene; cDNA EST yk340c3.3 comes from this gene; cDNA EST yk340c3.5 comes from this gene; cDNA EST yk339c8.3 comes from this gene; cDNA EST yk339c8.5 comes from this gene; cDNA EST yk336c7.3 comes from this gene; cDNA EST yk336c7.5 comes from this gene; cDNA EST yk289c8.3 comes from this gene; cDNA EST yk289c8.5 comes from this gene; cDNA EST yk289c1.3 comes from this gene; cDNA EST yk289c1.5 comes from this gene; cDNA EST yk236a11.3 comes from this gene; cDNA EST yk213h3.3 comes from this gene; cDNA EST yk290h3.5 comes from this gene"
/codon_start=1
/db_xref="PID:g1067079"
/db_xref="SPTREMBL:Q23382"
/translation="MKIVWIFLIFPIGFAISTDDNEFAEFEDEFVGSSATQAPEIQRE GEPPVLKQKDPEEEDPGVVEEEPEEAEKVREADSDDAAPAQPLKFADVPAHFRSNWA SYQVEGIVVLIILIIYMTNYLIGKTTNASIAQTIFDMCRPTLEEQFAVVGDDGTTDLDK MIPSLKHDTDSTFSAWCTGRVNVNSLFLQMKHVKRQDVVSRIKEMFTPSGDKMTIKAS LETTNDTDPLIFAVGEKKIASKYFKEMLDLNSFASERKQAAQQFNLPASWQVYADQNE VVFSILDPGVVSLLKKHEDAIEFIHISDQFTGPKPAEGESYTRLPEAQRYMPVSLNLQ YLGQDEESVMEILNLVFTLIDKARKMKLSKDAKVKAERRRKEPEDAFLKQTHQFRQEA AQARREEKTRERKQKLMDESDPERQKRLEAKELKREAKAKSPKMKQLKVK"

gene 21329. .23094
/gene="ZK1058.3"

CDS join(21329. .21605,21651. .21754,22107. .22289,22547. .22732, 22786. .23094)
/gene="ZK1058.3"
/note="similar to galactose-1-phosphate uridylyltransferase"

/codon_start=1
/db_xref="PID:g521063"
/db_xref="SPTREMBL:Q27536"
/translation="MSKQNNFRRYNPLIDEWVIVAVNRINRPWQGAKTEKSTTISTSS

...
Note: remainder of annotations omitted.

Query Match 31.1%; Score 19; DB 18; Length 29101;
Best Local Similarity 100.0%; Pred. No. 4.66e+00;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 10603 AAAATAAAACAATAAAGAC 10621
||||||||||||||||||
Qy 23 AAAATAAAACAATAAAGAC 41

RESULT 19
LOCUS CELD2092 37558 bp DNA INV 07-FEB-1997
DEFINITION Caenorhabditis elegans cosmid D2092.
ACCESSION U88167
NID g1825585
KEYWORDS .
SOURCE Caenorhabditis elegans strain=Bristol N2.
ORGANISM Caenorhabditis elegans
Eukaryotae; mitochondrial eukaryotes; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 37558)
AUTHORS Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M., Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J., Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A., Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M., Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N., Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B., O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A., Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E., Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M., Vaughan,K., Waterston,R., Watson,A., Weinstock,L., Wilkinson-Sproat,J. and Wohldman,P.
TITLE 2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans
JOURNAL Nature 368 (6466), 32-38 (1994)
MEDLINE 94150718
REFERENCE 2 (bases 1 to 37558)
AUTHORS Gattung,S. and Maggi,L.
TITLE The sequence of C. elegans cosmid D2092
JOURNAL Unpublished (1997)
REFERENCE 3 (bases 1 to 37558)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (03-FEB-1997)

```
COMMENT     Submitted by:
            Genome Sequencing Center
            Department of Genetics, Washington University,
            St. Louis, MO  63110, USA, and
            Sanger Centre, Hinxton Hall
            Cambridge CB10 1RQ, England
            e-mail: rw@nematode.wustl.edu and jes@sanger.ac.uk

            NOTICE:  This sequence may not be the entire insert of this clone.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we provide a small overlap between
            neighboring submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded or sequenced with an alternate
            chemistry; an attempt was made to resolve all sequencing problems,
            such as compressions and repeats; all regions were covered by
            sequence from more than one subclone

                        NEIGHBORING COSMID INFORMATION:

            The 5' cosmid is F48A9, 200 bp overlap;3' cosmid is ZC581, 2000 bp
            overlap.  Actual start of this cosmid is at base position 1 of
            CELD2092; actual end is at 10992 of CELZC581

            NOTES:

            Coding sequences below are predicted from computer analysis, using
            the program Genefinder(P. Green and L. Hillier, ms in preparation).
FEATURES              Location/Qualifiers
     source           1. .37558
                      /organism="Caenorhabditis elegans"
                      /strain="Bristol N2"
                      /db_xref="taxon:6239"
                      /chromosome="I"
                      /clone="D2092"
     gene             1557. .2840
                      /gene="D2092.6"
     CDS              join(1557. .1777,2024. .2636,2715. .2840)
                      /gene="D2092.6"
                      /codon_start=1
                      /evidence=not_experimental
                      /db_xref="PID:g1825590"
                      /translation="MERVIDELEMISSPLAESEDSLISEMSTKNDVARSSWLEYRGIR
                      RRIEDGNESDPEDDSEVIGPPPIPERDYSRLRPLAYRRRGEEFPMRRSKSYERPNVTR
                      LPEKKWIRNTRSLVDKSRKINPIYATPIRKKISKASVEPPIIEEPSEEVPRYVRPLK
                      TPKSSEEIIEEKEKENKKEEMKVIPVIDMSTEVEPSSKVERSIGKLAREDSLTSLLK
                      RQLSSKNRRKNREELPEYVEPKDPIMKELQEFLKIEPKSSEENSNSQNSTLTSIEKER
```



```
                      KEESEEKEKAVQADEHRLIVRGWQELLKSSSSSSSSSTTSTNV"
     gene             complement(3972. .11856)
                      /gene="D2092.5"
     CDS              complement(join(3972. .4143,4496. .4764,4838. .5083,
                      5173. .5329,5482. .6131,6519. .6898,7072. .7122,8999. .9132,
                      9595. .9712,10137. .10204,10295. .10373,10747. .10909,
                      11167. .11153,11401. .11507,11553. .11625,11804. .11856))
                      /gene="D2092.5"
                      /codon_start=1
                      /evidence=not_experimental
                      /db_xref="PID:g1825589"
                      /translation="MHVDQLGYTYKCAKLKLIQIIFFFKSLFIENCFRFLFNVFFIM
                      MQQQKPGKPKKINRIDKIKRLQINRSRRPDINQTVPSPLFYVRIVVTWLGMVSLDAMT
                      GFRFELLWPTWLMIRAAAESIQMRNQHCVTTIANPTAARFSVLFICVTATSDLICYLF
                      IPIRMLIPLATTYVWISLYYHTQGGFLRSLATVYGGERLQSWPIVFITCFIVIFELFL
                      RIRITKCDYIYLLFRSHPILISFFPNVAETAGVSPVWPRSLNAFPGAHSIGYPVILITV
                      SMHTYFNEWKLRRKQCDVSNRNEQLFRILVTGLPAETPGPEDYTSQQCLEDGFRIEKM
                      LLKTRMYKKKNLYYLDPPVQTLQPMQAIQAASATPPTSSKKNGIHKRNGDVTSSTTTS
                      SRKKKHNGNSGFNSTPPNDKKKGKSIRDVDMDDGDDSDDDYSYRDTSSSTIEDQRRGG
                      GISIIRFIPSSAAWLFSFVFESSTPSENSLSNQQIDDDEDYEDGDGDEKNGRTDSMTS
                      TTKGRANTMPSTTRSQNNNNSQKQQKQSNGKSHHQHSSHQNNHQKSNGNSNGHARGPA
                      AVRDSSHDTNASNETDIRSMSRELESLRSEISSRRSQEEDFKLQVSMHESNETRLSQQ
                      LSNMRLKVEQMEIKCSSIERHRESDKHQLEQAERKYADLLGKKAEIEATLSAERKARM
                      EVTSKKYDVAEHQRERERQLESEIDKLRIELKSKDESNMRMESELHGLRNYKEEMDID
                      SLNMELRFVRDKSHQMEESLAGENKLKQSLFKCLGDARDTIKSLERRVQEFQIKNGSS
                      IGGGSSETLMNGRSSTEANNENDTTASDQSSPHQHSAMGSPVPPAKMPLSVNVSNRNG
                      SPFNGKVSPIASIGSVLAAAGGPAPPDYMMAVGANVTATTGPVPQKQPRAGFHGISRY
                      NEFTNIASGGEHRLFDTPASAISASAINGSNPEDDFLMNKGKFGAPSQPAARLA"
     gene             complement(16427. .19870)
                      /gene="D2092.4"
     CDS              complement(join(16427. .16543,16615. .16684,16885. .16986,
                      17312. .17519,17566. .17757,19188. .19386,19769. .19870))
                      /gene="D2092.4"
                      /note="coded for by C. elegans cDNA CEESO80F; coded for by
                      C. elegans cDNA CESAA53F; coded for by C. elegans cDNA
                      CEESX68F"
                      /codon_start=1
                      /db_xref="PID:g1825588"
                      /translation="MKTSSFLTLITVSVIFLVFLSTHVDAKKKDKNLKVTVADDVAPP
                      APAKKSKKSAVKEAPAPPPKALSSKHKDDDNSIDDVDEDRIDEILRDASKNLVIFLYD
                      GKVPCPTCTEALSEVEEIDDDIEATGYVQVVKTNDRSVARELGINVFPSLVYYRRKNP
                      ILYDGDFKDSETLLRWLRAHEEVATWDLTDDTFESRTDSHSPDEGSIDWFVMFYDADE
                      GNSNAFVPLWETVAHKLRGLVNVGKIEISVNDDVTERFHIEERDCPVFLLVPEPPTAI
                      EQVTEFAKEKIMDIMDDNQTLSVLGVGGLIVIVAVTLIIKAYRIRAENKNKVA"
     gene             18419. .22529
                      /gene="D2092.7"
     CDS              join(18419. .18579,20302. .20380,20523. .20668,20713. .20969,
                      21111. .21280,21492. .21721,21895. .21996,22113. .22349,
                      22403. .22529)
                      /gene="D2092.7"
                      /codon_start=1
```



```
                      /evidence=not_experimental
                      /db_xref="PID:g1825592"
                      /translation="MQSISSLLCVLALAMCSIVQETTAQYYGYSSYYPSYYGGYGLGA
                      YNGYYGGSYGRRRGHAGTDEEIVKTEDAEPEPFEEMYSVPRIVALYEMLIHGQWKYMA
                      NYLSKKRTQLVLGMYIGAAFELVRFVCALVIAIVLILLRMSVLLPLYSDSIIGNLSWI
                      CEMNYLAFWIALIVFGLIVIQPTVYQCRNPTYVVISTSINFIFSIILITALIHTVILK
                      DNKEAFLKEGLHKKTLFSGTMSEMRTTYQARMKCCGVEGADDTSKSHIMTYVNPCSTNP
                      NSLTLSNPKTPPSKLQTPTITGGKISVFYKGLSYLLENTIRVPVTREPSSIYEYHFY
                      SIPMSCRKFDGVVCSQRNLTSEETQEVRKYAEVGKWDLMLKYFVMPELGLYWSRGCL
                      TEYVDSIETYITVLISTLVIFALPTALTIGFVIVLLLYEDGIGHIVSDTKMFGVERNT
                      GVHFLMSLDMMNSDFQDDFSLIEFLRNTDIDRFVAGDKLGDTINGRKDAEKA"
     gene             24729. .25153
                      /gene="D2092.8"
     CDS              join(24729. .24731,24783. .24943,24994. .25153)
                      /gene="D2092.8"
                      /codon_start=1
                      /evidence=not_experimental
                      /db_xref="PID:g1825591"
                      /translation="MIARLAIAALLVSSSLACLPGLGGLGGGCGGAAASPCAAPPMPP
                      PVSCGGGCGAAPYPAGPMAAPYFAPAAIAPPPPPQGAVIGVGAQPAVYSAPLPAGNAY
                      VGQGK"
     gene             complement(25674. .28881)
                      /gene="D2092.3"
     CDS              complement(join(25674. .25914,25960. .26197,26756. .26842,
                      26910. .27002,27524. .27679,27740. .27954,28274. .28389,
                      28595. .28728,28776. .28881))
                      /gene="D2092.3"
                      /note="similar to ligand-gated ionic channel protein"
                      /codon_start=1
                      /evidence=not_experimental
                      /db_xref="PID:g1825587"
                      /translation="MIFNLINSILSLVIIHSNLCDGSVAETKLFTDLLKGYNPLERPV
                      QNSSQPLEVKIKLFLQQILDVDEKNQIVSVNAWLSYTWFDHKLQWEPKKYGGIQDIRF
                      PGSSDHIWKPDVLLYNSAAEDFDSTFKSNLLTYHTGTVVWIPPGVLKPVCQLDVTWFP
                      FDDQVCEMKFGSWTFHGYAIDLQIDDDTNGTQSMDLSTYLVNGEWQVISTNAKRVVSY
                      YKCCPEPYPTVNYYLHIRRRTLYYGFNLIIPSLLISLMAILGFMFPPDAGEKITLEVT
                      ILLAIVFFLSMVSEMTPPTSEAVPLIGVFFSCCMLVVSASVVFTIVVLNLHFRSADSH
                      EMNPLVRRVLLEFLPWLLFMSRPGYKFVKANVIDSTDKHPKFKPKNPLDCNLPSNHAGY
                      EAQILLLHSVHTELRHVVAFYNKEENDERIQTDWRFAAMVVDRACLLLFTVFIVISIL
                      AIMMSAPHIIA"
     gene             complement(29384. .30767)
                      /gene="D2092.2"
     CDS              complement(join(29384. .29797,29854. .29964,30011. .30195,
                      30245. .30374,30426. .30569,30615. .30767))
                      /gene="D2092.2"
                      /note="coded for by C. elegans cDNA cm18h7"
                      /codon_start=1
                      /db_xref="PID:g1825593"
                      /translation="MVKSPGASPERKKVKIDVYPMVVKDPQAREINETLLFDYTTALS
                      SFETPTEEQLLVSEGDAFRHHPNEDVEQYFRASAQWGVPAIAWNIVRPAPLWKLEFCI
                      TEFMNVEKKKKEAAANSAQANSEVEKIAESPVKEKLSILGHKVNLIKTTPVVFSTEES
```



```
                      MEFVLSKAKSFDGFPFTWQRLCELLIEPMRHYNTIDKFLRAVDKVINVVTTINENGGR
                      SFGDWDLPNSQQPHLENTFPGVVDEVEMMELEKVKHDFQSSSSPSEEPLDMSQKSMPA
                      PRVSTASPDPPSFVLPTARSPINTSPSSSPKASSPKAASPKAASPVSSPKSPVIQDQS
                      PKKEDAKEGEGEIVHEEEMPEETLEVETKNEVNKEDADMEIGAE"
     misc_feature     30775. .30776
                      /note="SL1 trans-splice site, see EST cm18h7"
     gene             complement(33893. .37478)
                      /gene="D2092.1"
     CDS              complement(join(33893. .33937,34097. .34408,34516. .34680,
                      34732. .34857,34900. .35173,35222. .35363,35455. .35903,
                      36019. .36268,36329. .36364,36612. .36778,36824. .37046,
                      37169. .37292,37377. .37478))
                      /gene="D2092.1"
...
Note: remainder of annotations omitted.

  Query Match           31.1%;  Score 19;  DB 18;  Length 37558;
  Best Local Similarity 100.0%;  Pred. No. 4.66e+00;
  Matches    19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  18941 TTTTATTTTGTTTTTGAAA 18959
          ||||||||||||||||||||
Cp     31 TTTTATTTTGTTTTTGAAA 13


RESULT  20
LOCUS       AC004650    79635 bp    DNA             PRI       01-MAY-1998
DEFINITION  Homo sapiens chromosome 5, P1 clone 207c2 (LBNL H33), complete
            sequence.
ACCESSION   AC004650 AC000986 AC000987 AC000988 AC000989 AC001493 L81384 L81385
            L81630 L81631 L81632 L81633 L81634 L81635 L81805 L81806 L81807
            L78761 L77061 L77062 L77063 L77064
NID         g3097834
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 79635)
  AUTHORS   Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
            Kadner,K., Miguel,T., Miller,C., Pitluck,S., Pollard,M.,
            Rojeski,H., Subramanian,S. and Martin,C.H.
  TITLE     Sequencing of human chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 79635)
  AUTHORS   Ricke,D.O.
  TITLE     Large Scale Sequence Analysis and Annotation with the Sequence
            Comparison Analysis (SCAN) System
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 79635)
  AUTHORS   Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
```

```
            Kadner,K., Miguel,T., Miller,C., Pitluck,S., Pollard,M.,
            Rojeski,H., Subramanian,S. and Martin,C.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAY-1998) Human Genome Center, DOE Joint Genome
            Institute, Lawrence Berkeley National Laboratory, MS 74-157,
            Berkeley, CA 94720, U.S.A.
COMMENT     Sequence submitted by:
            DOE Joint Genome Institute.
FEATURES             Location/Qualifiers
     source          1. .79635
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="5q"
                     /clone="207c2"
                     /chromosome="5"
                     /note="LBNL H33"
     repeat_region   1. .266
                     /rpt_family="Alu"
     repeat_region   complement(542. .835)
                     /rpt_family="Alu"
     repeat_region   complement(1893. .6170)
                     /rpt_family="L1"
     repeat_region   6166. .6495
                     /rpt_family="Alu"
     repeat_region   complement(6479. .8265)
                     /rpt_family="L1"
     repeat_region   8250. .9367
                     /rpt_family="MER11A"
     repeat_region   complement(9368. .9467)
                     /rpt_family="L1"
     repeat_region   complement(9538. .9863)
                     /rpt_family="Alu"
     repeat_region   9558. .9580
                     /note="(T)23"
                     /rpt_type=tandem
                     /rpt_unit=T
     repeat_region   9967. .11154
                     /rpt_family="L1"
     repeat_region   11154. .11502
                     /rpt_family="THE1b"
     repeat_region   11503. .11657
                     /rpt_family="THE1BR"
     repeat_region   complement(11654. .11800)
                     /rpt_family="L1"
     repeat_region   11813. .12573
                     /rpt_family="THE1BR"
     repeat_region   12567. .15182
                     /rpt_family="L1"
     repeat_region   12998. .13023
                     /note="(A)26"
                     /rpt_type=tandem
```



```
                     /rpt_unit=A
     repeat_region   15252. .15291
                     /note="(AC)20"
                     /rpt_type=tandem
                     /rpt_unit=AC
     repeat_region   15292. .15319
                     /note="(AT)14"
                     /rpt_type=tandem
                     /rpt_unit=AT
     repeat_region   complement(15329. .15655)
                     /rpt_family="Alu"
     repeat_region   15653. .15956
                     /rpt_family="Alu"
     repeat_region   17261. .17593
                     /rpt_family="Alu"
     repeat_region   17623. .17918
                     /rpt_family="L1MB3"
     repeat_region   complement(17935. .24040)
                     /rpt_family="L1"
     repeat_region   24042. .24102
                     /rpt_family="L1MB3"
     repeat_region   24296. .24585
                     /rpt_family="Alu"
     repeat_region   24764. .25085
                     /rpt_family="MER65A"
     repeat_region   complement(25207. .25317)
                     /rpt_family="MER63C"
     repeat_region   complement(25572. .25792)
                     /rpt_family="L1"
     misc_feature    25860. .25959
                     /note="80% identity n122a09.s1"
                     /db_xref="dbEST:AA570228"
     repeat_region   26348. .26646
                     /rpt_family="Alu"
     repeat_region   26667. .26972
                     /rpt_family="Alu"
     repeat_region   26949. .26974
                     /note="(A)26"
                     /rpt_type=tandem
                     /rpt_unit=A
     repeat_region   26986. .27594
                     /rpt_family="L1"
     repeat_region   28380. .29099
                     /rpt_family="L1"
     repeat_region   complement(29155. .29312)
                     /rpt_family="MER42"
     repeat_region   complement(29757. .29862)
                     /rpt_family="MER42"
     repeat_region   29891. .30627
                     /rpt_family="L1"
     repeat_region   complement(30718. .31050)
```



```
                     /rpt_family="Alu"
     repeat_region   complement(31086. .31293)
                     /rpt_family="MER42"
     misc_feature    complement(31691. .31791)
                     /note="GRAIL 2 excellent exon, frame 0"
     repeat_region   31809. .32168
                     /rpt_family="Alu"
     repeat_region   complement(32733. .33005)
                     /rpt_family="Alu"
     repeat_region   33161. .33354
                     /rpt_family="Alu"
     repeat_region   33333. .33360
                     /note="(A)28"
                     /rpt_type=tandem
                     /rpt_unit=A
     repeat_region   33612. .33940
                     /rpt_family="Alu"
     repeat_region   complement(34394. .34714)
                     /rpt_family="AluSc"
     repeat_region   complement(35229. .35579)
                     /rpt_family="Alu"
     repeat_region   complement(36610. .36936)
                     /rpt_family="Alu"
     repeat_region   36623. .36649
                     /note="(T)27"
                     /rpt_type=tandem
                     /rpt_unit=T
     repeat_region   38175. .38451
                     /rpt_family="Alu"
     repeat_region   39112. .39470
                     /rpt_family="Alu"
     repeat_region   39393. .39416
                     /note="(A)24"
                     /rpt_type=tandem
                     /rpt_unit=A
     repeat_region   complement(39537. .39866)
                     /rpt_family="Alu"
     repeat_region   40073. .40378
                     /rpt_family="Alu"
     repeat_region   40198. .40217
                     /note="(A)20"
                     /rpt_type=tandem
                     /rpt_unit=A
     prim_transcript complement(40285. .50149)
                     /standard_name="protein kinase C inhibitor"
     repeat_region   complement(40962. .41277)
                     /rpt_family="MLT1"
     repeat_region   41312. .41523
                     /rpt_family="Alu"
     repeat_region   complement(41519. .41856)
                     /rpt_family="MLT1a"
```



```
     repeat_region   complement(41975. .42038)
                     /rpt_family="MLT1"
     repeat_region   42270. .42613
                     /rpt_family="Alu"
     repeat_region   42908. .43218
                     /rpt_family="Alu"
     repeat_region   43346. .43667
                     /rpt_family="Alu"
     prim_transcript complement(join(44314. .44642,47603. .47728,50114. .50276))
                     /function="protein kinase C inhibitor"
     misc_feature    complement(44529. .44642)
                     /note="GRAIL 2 excellent exon, frame 1"
     repeat_region   complement(44881. .45183)
                     /rpt_family="Alu"
     misc_feature    complement(45311. .45841)
                     /note="99% identity zt50d04.r1"
                     /db_xref="dbEST:AA292196"
     misc_feature    complement(45530. .46000)
                     /note="95% identity yl98d04.r1"
                     /db_xref="dbEST:H09202"
     repeat_region   complement(46012. .46080)
                     /rpt_family="Tigger3"
...
Note: remainder of annotations omitted.
```

```
  Query Match           31.1%;  Score 19;  DB 26;  Length 79635;
  Best Local Similarity 100.0%;  Pred. No. 4.66e+00;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  35432 TTGTTTTATTTTGTTTTTG 35450
          |||||||||||||||||||
Cp     34 TTGTTTTATTTTGTTTTTG 16


RESULT   21
LOCUS       AT9A21     82697 bp    DNA             PLN       03-FEB-1998
DEFINITION  Arabidopsis thaliana DNA chromosome 4, BAC clone T9A21 (ESSAII
            project).
ACCESSION   AL021713
NID         g2832689
KEYWORDS    .
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryotae; Viridiplantae; Charophyta/Embryophyta group;
            Embryophyta; Tracheophyta; seed plants; Magnoliophyta;
            eudicotyledons; Rosidae; Capparales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 82697)
  AUTHORS   Bevan,M., Murphy,G., Drost,L., Hall,C., Hudson,S., Ridley,P.,
            Bancroft,I., Mewes,H.W., Mayer,K. and Schueller,C.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 82697)
```

```
  AUTHORS   EU Arabidopsis sequencing,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-FEB-1998) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, Project
            Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
            Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
            E-mail: michael.bevan@bbsrc.ac.uk
FEATURES             Location/Qualifiers
     source          1. .82697
                     /organism="Arabidopsis thaliana"
                     /variety="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="4"
     gene            402. .1585
                     /gene="T9A21.10"
     CDS             join(402. .522,577. .832,946. .1090,1178. .1585)
                     /gene="T9A21.10"
                     /note="similarity to SPF1 protein, sweet protein,
                     PIR2:S51529 and WRKY protein family, Petroselinum crispum,
                     MNOS:S72443, MNOS:S72444, MNOS:S72445"
                     /codon_start=1
                     /product="DNA binding-like protein"
                     /db_xref="PID:e1249681"
                     /db_xref="PID:g2832690"
                     /translation="MSNETRDLYNYQYPSSFSLHEMMNLPTSNPSSYGNLPSQNAYES
                     LLQKTFGLSPSSSEVFNSSIDQEPNRDVTNDVINGGACNETETRVSPSNSSSSEADHP
                     GEDSGKSRRKRELVGEEDQISKKVGKTKKTEVKKQREPRVSFMTKSEVDHLEDGYRWR
                     KYGQKAVKNSPYPRIIANGNENRSYYRCTTQKCNVKKRVERSPQDPYVVITTYEGQHN
                     HPIPTNLRGSSAAAAAMFSADLMTPRSFAHDMFRTAAYTNGGSVAAALDYGYGQSGYGS
                     VNSNPSSHQVYHQGGEYELLREIPPSIFFKQEP"
     exon            402. .522
                     /gene="T9A21.10"
                     /number=1
     intron          523. .576
                     /gene="T9A21.10"
                     /number=1
     exon            577. .832
                     /gene="T9A21.10"
                     /number=2
     intron          833. .945
                     /gene="T9A21.10"
                     /number=2
     exon            946. .1090
                     /gene="T9A21.10"
                     /number=3
     intron          1091. .1177
                     /gene="T9A21.10"
                     /number=3
     exon            1178. .1585
                     /gene="T9A21.10"
                     /number=4
```



```
     exon            4778. .4921
                     /gene="T9A21.20"
                     /number=1
     gene            4778. .6211
                     /gene="T9A21.20"
     CDS             join(4778. .4921,5023. .5187,5263. .5552,5638. .5866,
                     5951. .6211)
                     /gene="T9A21.20"
                     /note="Protein sequence is in conflict with the conceptual
                     translation; strong similarity to exopolygalacturonase,
                     Arabidopsis thaliana, PIR2:S34266"
                     /codon_start=1
                     /product="polygalacturonase-like protein"
                     /db_xref="PID:e1249682"
                     /db_xref="PID:g2832691"
                     /translation="AFVAAWDKACKSSSSSSVNLIIPRGEFSVGSLRFSGPCTNVSNLT
                     VRVTASTDLSKYRSGGGWIQFGWINGLTLTGGGTFDGQGALAWPFNNCTSDSNCKLLP
                     TSLKFVGMNRTVVRRISSVNSKFFHIALVECRDFKGTRLNITAPSDSPNTDGIHIERS
                     SNVYFSRSHIATGDDCVSIGQGNSQITITSIKCGPGHGISVGSLGRYPNEKDVNGLVV
                     KDCKISGTTNGIRIKTWANSPGLSAATNMTFENIIMNNVTNPIIIDQSYCPFSSCISN
                     VPSKVELSEIYFKNIRGTSSSLVAVQLHCSRGMPCKKVYLENVHLDLSSSDGGRKQSS
                     NRGNEAVSSSCRNVRANYIGTQIPPPCH"
     intron          4922. .5022
                     /gene="T9A21.20"
                     /number=1
     exon            5023. .5187
                     /gene="T9A21.20"
                     /number=2
     intron          5188. .5262
                     /gene="T9A21.20"
                     /number=2
     exon            5263. .5552
                     /gene="T9A21.20"
                     /number=3
     intron          5553. .5637
                     /gene="T9A21.20"
                     /number=3
     exon            5638. .5866
                     /gene="T9A21.20"
                     /number=4
     intron          5867. .5950
                     /gene="T9A21.20"
                     /number=4
     exon            5951. .6211
                     /gene="T9A21.20"
                     /number=5
     CDS             join(6969. .7148,7191. .8087)
                     /gene="T9A21.30"
                     /note="similarity to W30DMY32F, W25DMY32F"
                     /codon_start=1
                     /product="putative protein"
```



```
                     /db_xref="PID:e1249683"
                     /db_xref="PID:g2832692"
                     /translation="MVEPEGKFSTEERSHKYSWRLRVSLYVTLLLAGETIATLLGRLY
                     YEKGGKSTWLETLVQLPEPSKTKTIITKKTTSSFLTLSLVYIGLGLLVAGHCILYSPGL
                     LYLPVSTFSLISASQLAFNAVFSYFLHSQKITPPILNSLVLLTIGSTLLVIQHEPESP
                     SSTSKSAAKSKYVIGYICAVGSSAGYSLVLSLTDYAFEKILKKTTFKAILDMATYPSM
                     VATCVVVVGLPGSGGWKKLSTEMEEPQLGKSSYILINIGSTISWQACLIGSVGLIIEV
                     SSLFSNVISTLCLPVVPVLAVVFFRDEMSGIKLVAMFLAIWGFVSYGYQHYVNDRKPE
                     EDQELPQSKEEEEQKQVDTIHVQA"
     exon            6969. .7148
                     /gene="T9A21.30"
                     /number=1
     gene            6969. .8087
                     /gene="T9A21.30"
     intron          7149. .7190
                     /gene="T9A21.30"
                     /number=1
     exon            7191. .8087
                     /gene="T9A21.30"
                     /number=2
     gene            8612. .13959
                     /gene="T9A21.50"
     exon            8612. .8838
                     /gene="T9A21.50"
                     /number=1
     CDS             join(8612. .8838,8884. .9860,10782. .11155,11210. .11915,
                     12857. .13959)
                     /gene="T9A21.50"
                     /note="similarity to W30DMY32F, W15DMY32F; contains EST
                     gb:T42237"
                     /codon_start=1
                     /product="putative protein"
                     /db_xref="PID:e1249685"
                     /db_xref="PID:g2832694"
                     /translation="MEIIPQVIVFMGKQDASRSVDYLILFANLLFLMPSDHNNIEANLT
                     GQEEMNTTMEIESSSSVPQSKNYKKWLRISIYVVYYENGGKSTWMGTLVQLIGFPVLFL
                     FRFFSQTKNPKPTEADFRKFSSFTILGSVYIVTGLLVSANSYMSSVGLLYLPVSTFSL
                     ILASQLAFTAFFSYFLNSQKFTPPIVNSLFLLTISSALLVVNTDSENTAKVSRVKYVI
                     GIICTIGASAGIGLLLSLVQLILRKVLKKQTFSTVTDLVAYQSLVASCVVLIGLFASG
                     EWKTLTSEMENYKLGKVPYVMTLASIAISWQVYTIGVVGLIFESSSVFSNSITAVGLP
                     IVPVVAVIVFHDKMNASKIFSIILAIWGFISFVYQHYLDEKKLKTSHTSPVGDPHLLP
                     AEEGHTNIHSDQNLEANLIDHEVVTESSSSAVPQTENYKRWLRVSIYVIFVLFCQPLA
                     TILGRLYIENGGNSTYVVTLLQLIGFPVLVLFRFFSRIRQPKSTDTNFSQSPSFTTLA
                     SVTLCTGLLVSAYAYLSALAFTAFFSYFLNSQKFTPLIVSSLLLLTVSSALLVVNTDS
                     ENSTNVSRVQYVIGPICTIGASAGIGLLLSLIQMLFREVFTKHTSSAVTDLAIIQSLV
                     ASCVVLIGLFASGEWETLPSEMRNYKLGKVSYVLTLASAAISWQVTTLGLVGLIFESS
                     SVFSNSITAVGLPIVPVAAVIVFHDRMDASKIFSIILAICGFLSFVYQHYLDEKKLNT
                     SHTSAVGDLHLPVEEGHTNIQSDQNLEANLLDHEETESFSVPQTKNCKRWLRVSIYAI
                     FVIFCQPLATVLGRLYYENGGKSTYVVTLLQLIGFPVLILFRFFSRIRQPKSTDTNFS
                     QSPSFTTLASVYLCTGLLVSAYAYLSAVGLLYLPVSTFSLILASQLAPTAFFSYFLNS
                     QKFTPLIVNSLFLLTVSSALLVVNTDSENTTNVSRVQYVIGFICTIGASAGIGLVLSL
```



```
                     IQLLFRKVFTKHTSSAVLDLANYQSLVATCVVLIGLFASGEWRTLPSEMRNYKLGKVS
                     YILTLASAAIFWQVYTVGCVGLIFESSSVFSNSITAVGLPIVPVVAVIVFHDKMDASK
                     IFSIILAIWGFLSFVYQHYLDEKKLKTCQTKPVEEETQTL"
     intron          8839. .8883
                     /gene="T9A21.50"
                     /number=1
     exon            8884. .9860
                     /gene="T9A21.50"
                     /number=2
     exon            complement(9085. .9263)
                     /gene="T9A21.40"
                     /number=1
     gene            9085. .11487
                     /gene="T9A21.40"
     CDS             complement(join(9085. .9263,10105. .10153,10203. .10267,
                     11175. .11487))
                     /gene="T9A21.40"
                     /codon_start=1
                     /product="hypothetical protein"
                     /db_xref="PID:e1249684"
                     /db_xref="PID:g2832693"
                     /translation="MAKSVTAEDVCFVKTFLKSICIRDSNSPIPAEAPMVHMNPITYC
                     TLDTLVEFSESVLTTRRAEDTVRRSKELTIKGVNFCELRKYEKKAVKANCEAKMREKV
                     ETGVPLQGTVQVTTLVICYVSGFQLWFSKDALAYLPLCYRLLTTRRAEEIVRRNREFT
                     MKGVNFCEFRKYEKKAVKANCEAKMREKVETGKYSKPTEDI"
     intron          complement(9264. .10104)
                     /gene="T9A21.40"
                     /number=1
     intron          9861. .10781
                     /gene="T9A21.40"
                     /number=2
...
Note: remainder of annotations omitted.

  Query Match           31.1%;  Score 19;  DB 23;  Length 82697;
  Best Local Similarity 100.0%;  Pred. No. 4.66e+00;
  Matches    19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  45305 TTTCAAAAACAAAATAAAA 45323
          |||||||||||||||||||
Qy     13 TTTCAAAAACAAAATAAAA 31


RESULT   22
LOCUS       HS833B7     86574 bp    DNA             PRI       25-AUG-1998
DEFINITION  Human DNA sequence from clone 833B7 on chromosome 22q12.3-13.2
            Contains genes for NCF4 (P40PHOX) protein,cytokine receptor common
            beta chain precursor CSF2RB (partial), ESTs, CA repeat, STS, GSS,
            complete sequence.
ACCESSION   AL008637
NID         g3136000
```

```
KEYWORDS    HTG; CSF2RB; cytokine receptor; growth factor receptor; NCF4;
            P40PHOX; repeat polymorphism.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 86574)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-MAY-1998) E-mail enquiries: humquery@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On May 18, 1998 this sequence version replaced gi:2578146.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            IMPORTANT: This sequence is not the entire insert of clone 833B7.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we arrange for a small overlap between
            neighbouring submissions.
            The true left end of clone fF45C1 is at 86475 in this sequence. The
            true right end of clone fF24E5 is at 100 in this sequence. The
            start of this sequence overlaps with the end of sequence Z82185.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            This sequence was generated from part of bacterial clone contigs of
            human chromosome 22, constructed by the Sanger Centre Chromosome 22
            Mapping Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr22
            833B7 is from the human BAC library described in U-J. Kim et al.
            (1996) Genomics 34, 213-218. VECTOR: pBeloBAC11.
FEATURES            Location/Qualifiers
     source         1. .86574
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="22"
                    /clone="833B7"
                    /map="q12.3-13.2"
     repeat_region  1. .166
                    /note="AluSp repeat: matches 138. .303 of consensus"
     repeat_region  complement(870. .903)
                    /note="MIR repeat: matches 244. .212 of consensus"
     repeat_region  1791. .2090
                    /note="AluSx repeat: matches 1. .302 of consensus"
     repeat_region  3533. .3832
```



```
                    /note="AluSx repeat: matches 2. .302 of consensus"
     repeat_region  5291. .5347
                    /note="L1MC1 repeat: matches 273. .328 of consensus"
     repeat_region  complement(5379. .5676)
                    /note="AluSx repeat: matches 302. .1 of consensus"
     repeat_region  5677. .6255
                    /note="L1MB8 repeat: matches 316. .918 of consensus"
     repeat_region  7347. .7532
                    /note="MIR repeat: matches 2. .191 of consensus"
     repeat_region  complement(7537. .7640)
                    /note="MIR repeat: matches 262. .158 of consensus"
     repeat_region  complement(8163. .8215)
                    /note="MIR2 repeat: matches 127. .75 of consensus"
     repeat_region  8278. .8341
                    /note="MIR2 repeat: matches 84. .144 of consensus"
     repeat_region  complement(11386. .11474)
                    /note="MIR2 repeat: matches 145. .54 of consensus"
     mRNA           join(11913. .12573,15415. .15499,16289. .16442,18762. .18832,
                    21785. .21912,23022. .23079,23697. .23795,27023. .27153,
                    27399. .27464,28998. .29288)
                    /gene="NCF4"
                    /note="match: U50720 U50721 U50722 U50723 U50724; match:
                    U50725 U50726 U50727 U50728 U50729"
                    /evidence=not_experimental
     gene           11913. .29288
                    /gene="NCF4"
     CDS            join(12542. .12573,15415. .15499,16289. .16442,18762. .18832,
                    21785. .21912,23022. .23079,23697. .23795,27023. .27153,
                    27399. .27464,28998. .29193)
                    /gene="NCF4"
                    /codon_start=1
                    /evidence=not_experimental
                    /db_xref="PID:e1293191"
                    /db_xref="PID:g3136001"
                    /translation="MAVAQQLRAESDFPQLPDDVAISANIADIEERGFTSHFVFVIE
                    VKTKGSSYLIYRRPQFHALQSKLEERFGPDSKSSALACTLPTLPAVTVGVKQEIA
                    EMRIPALNATMKSLLSLPWWLMDEDVRIFFTQSPYDSEQVPQALRRLRPRTRKVKSV
                    SPQGNSVDRMAAPRAEALFDFTGNSKLELNFKAGDVIFLLSRINKDWLEGTVRGATGI
                    FPLSFVKILKDFPEEDDPTNWLRCTYYEDTISTIKDIAVEEDLSSTPLLKDLLELTRR
                    EFQREDIALNYRDAEGDLVRLLSDEDVALMVRQARGLPSQKRLFPWKLHITQKDNYRV
                    YNTMP"
     repeat_region  complement(12766. .12882)
                    /note="MIR repeat: matches 136. .20 of consensus"
     repeat_region  13103. .13247
                    /note="MIR2 repeat: matches 1. .146 of consensus"
     repeat_region  complement(13315. .13403)
                    /note="MIR repeat: matches 143. .53 of consensus"
     repeat_region  complement(14202. .14239)
                    /note="MIR2 repeat: matches 137. .100 of consensus"
     repeat_region  complement(14545. .14625)
                    /note="MIR2 repeat: matches 124. .33 of consensus"
```



```
     repeat_region  complement(14656. .14743)
                    /note="MIR repeat: matches 179. .89 of consensus"
     repeat_region  14983. .15050
                    /note="MIR2 repeat: matches 34. .106 of consensus"
     repeat_region  15616. .15815
                    /note="MER20 repeat: matches 1. .218 of consensus"
     repeat_region  complement(18243. .18412)
                    /note="MER5B repeat: matches 175. .1 of consensus"
     repeat_region  complement(18413. .18485)
                    /note="MIR2 repeat: matches 94. .22 of consensus"
     misc_feature   complement(18431. .18744)
                    /gene="NCF4"
                    /note="match: Z51609 STS containing (CA) repeat"
     repeat_region  18486. .18513
                    /note="14 copies of TG 100 % conserved; differs from
                    Z51609"
     repeat_region  19642. .19730
                    /note="MIR repeat: matches 57. .145 of consensus"
     repeat_region  complement(20609. .20820)
                    /note="MIR repeat: matches 261. .35 of consensus"
     repeat_region  21117. .21212
                    /note="MIR repeat: matches 50. .149 of consensus"
     repeat_region  21325. .21412
                    /note="MIR repeat: matches 31. .120 of consensus"
     repeat_region  22584. .22764
                    /note="MIR repeat: matches 57. .239 of consensus"
     repeat_region  22785. .22855
                    /note="MIR repeat: matches 109. .185 of consensus"
     repeat_region  23491. .23616
                    /note="MIR repeat: matches 20. .144 of consensus"
     prim_transcript <23695. .>29387
                    /note="match: multiple ESTs; match: AA177839 AA465462
                    AA485518 AA648472; match: AA465389 W95229 AA744805
                    AA702857; match: AA688071 AA714514 AA909156 D20144; match:
                    AA034506 W92604"
     repeat_region  complement(25055. .25240)
                    /note="MER20 repeat: matches 216. .11 of consensus"
     repeat_region  25290. .25405
                    /note="MIR repeat: matches 15. .139 of consensus"
     repeat_region  26032. .26111
                    /note="MIR repeat: matches 107. .187 of consensus"
     repeat_region  complement(26397. .26699)
                    /note="AluSq repeat: matches 303. .1 of consensus"
     repeat_region  26727. .26778
                    /note="MIR2 repeat: matches 95. .146 of consensus"
     repeat_region  complement(27576. .27701)
                    /note="MIR repeat: matches 192. .61 of consensus"
     repeat_region  complement(28508. .28806)
                    /note="AluSx repeat: matches 299. .1 of consensus"
     polyA_site     29361. .29366
     repeat_region  complement(30298. .30750)
```



```
                    /note="MLT1C repeat: matches 460. .1 of consensus"
     repeat_region  31938. .32322
                    /note="MSTA repeat: matches 1. .401 of consensus"
     repeat_region  complement(32350. .32650)
                    /note="AluSq repeat: matches 302. .2 of consensus"
     repeat_region  32652. .32689
                    /note="MSTA repeat: matches 389. .426 of consensus"
     repeat_region  complement(33615. .33661)
                    /note="MLT1B repeat: matches 132. .86 of consensus"
     repeat_region  complement(33662. .33961)
                    /note="AluSp repeat: matches 302. .2 of consensus"
     repeat_region  complement(33963. .34057)
                    /note="MLT1C repeat: matches 94. .1 of consensus"
     repeat_region  34907. .35205
                    /note="AluSq repeat: matches 1. .299 of consensus"
     repeat_region  complement(37404. .37645)
                    /note="MLT2_internal repeat: matches 5379. .5124 of
                    consensus"
     repeat_region  complement(37765. .38033)
                    /note="AluSq repeat: matches 303. .43 of consensus"
     repeat_region  complement(38243. .38549)
                    /note="AluSx repeat: matches 302. .1 of consensus"
     repeat_region  complement(40046. .40413)
                    /note="MLT1A1 repeat: matches 365. .1 of consensus"
     repeat_region  40607. .40931
                    /note="AluY repeat: matches 2. .294 of consensus"
     repeat_region  41986. .42285
                    /note="AluSq repeat: matches 1. .300 of consensus"
     repeat_region  42597. .42876
                    /note="AluJo repeat: matches 6. .301 of consensus"
     repeat_region  42969. .43106
                    /note="L1ME1 repeat: matches 4. .144 of consensus"
     repeat_region  43119. .43209
                    /note="L1MC2 repeat: matches 291. .383 of consensus"
     repeat_region  43198. .43375
                    /note="L1ME3 repeat: matches 402. .573 of consensus"
     repeat_region  44217. .44298
                    /note="MIR repeat: matches 59. .157 of consensus"
     repeat_region  46084. .46509
                    /note="MSTA repeat: matches 1. .426 of consensus"
...
Note: remainder of annotations omitted.

  Query Match           31.1%;  Score 19;  DB 27;  Length 86574;
  Best Local Similarity 100.0%;  Pred. No. 4.66e+00;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  41735 TTGTTTTATTTTGTTTTTG 41753
          |||||||||||||||||||
Cp      34 TTGTTTTATTTTGTTTTTG 16
```

```
RESULT   23
LOCUS       HS510H16   102484 bp    DNA             PRI       24-JUN-1998
DEFINITION  Homo sapiens DNA sequence from PAC 510H16 on chromosome
            22q12.1-12.3. Contains TOM protein, ESTs, CA repeat, STS, GSS,
            complete sequence.
ACCESSION   AL008635
NID         g3183882
KEYWORDS    HTG; 22q12.1-12.3; repeat polymorphism; TOM.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 102484)
  AUTHORS   Bridgeman,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUN-1998) sanger.ac.uk/HGP/Chr22/) Sanger Centre,
            Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquires:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jun 4, 1998 this sequence version replaced gi:2578121.
            IMPORTANT: This sequence is not the entire insert of clone 510H16.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we arrange for a small overlap between
            neighbouring submissions.
            During sequence assembly data are compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variations annotated may not be found in the sequence submission
            corresponding to the overlapping clone as we submit sequences with
            only a small overlap as described above.
            This sequence was generated from part of bacterial clone contigs of
            human chromosome 22, constructed by the Sanger Centre chromosome 22
            mapping group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr22/
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            The true left end of clone 510H16 is at 1 in this sequence. The
            true left end of clone 77D12 is at 56548 in this sequence. The true
            right end of clone 77D12 is at 98164 in this sequence. The true
            left end of clone 286B10 is at 102372 in this sequence. The start
            of this sequence overlaps with the end of sequence Z99755.
            510H16 is from the library RPCI3 constructed at the Roswell Park
            Cancer Institute by the group of Pieter de Jong.
            For further details see http://bacpac.med.buffalo.edu/.
FEATURES             Location/Qualifiers
     source          1. .102484
                     /organism="Homo sapiens"
```



```
                     /db_xref="taxon:9606"
                     /chromosome="22"
                     /map="q12.1-12.3"
                     /clone="510H16"
                     /clone_lib="RPCI3"
     repeat_region   8. .79
                     /note="MER25 repeat: matches 1030. .959 of consensus"
     repeat_region   84. .127
                     /note="MER25 repeat: matches 911. .868 of consensus"
     repeat_region   157. .794
                     /note="pTR5 repeat: matches 1850. .2438 of consensus"
     repeat_region   799. .1082
                     /note="MER25 repeat: matches 870. .584 of consensus"
     repeat_region   1211. .1738
                     /note="MER25 repeat: matches 554. .16 of consensus"
     repeat_region   1750. .2110
                     /note="L1 repeat: matches 3744. .3362 of consensus"
     repeat_region   2325. .2350
                     /note="13 copies of 2 mer 100 % conserved"
     repeat_region   2352. .2880
                     /note="L1 repeat: matches 3212. .2674 of consensus"
     repeat_region   2883. .3183
                     /note="AluSq repeat: matches 1. .302 of consensus"
     repeat_region   3184. .3288
                     /note="MLT2D repeat: matches 111. .1 of consensus"
     repeat_region   3329. .3637
                     /note="AluSx repeat: matches 302. .1 of consensus"
     repeat_region   4086. .4179
                     /note="AluSx repeat: matches 302. .209 of consensus;
                     incomplete repeat"
     repeat_region   4515. .4743
                     /note="LTR10 repeat: matches 242. .20 of consensus"
     repeat_region   4766. .4965
                     /note="AluSq repeat: matches 201. .1 of consensus;
                     incomplete repeat"
     repeat_region   6068. .6366
                     /note="AluY repeat: matches 300. .1 of consensus"
     repeat_region   7520. .7559
                     /note="10 copies of 4 mer 90 % conserved"
     repeat_region   7668. .7987
                     /note="MER7B repeat: matches 2. .297 of consensus"
     repeat_region   7987. .8170
                     /note="MSTA repeat: matches 426. .243 of consensus"
     repeat_region   8199. .8497
                     /note="AluSx repeat: matches 302. .2 of consensus"
     repeat_region   8499. .8694
                     /note="MSTB repeat: matches 200. .1 of consensus"
     repeat_region   8687. .8843
                     /note="MER7B repeat: matches 284. .440 of consensus"
     repeat_region   8852. .9153
                     /note="AluSx repeat: matches 301. .1 of consensus"
```



```
     repeat_region   9164. .9245
                     /note="MER7B repeat: matches 454. .533 of consensus"
     repeat_region   9569. .9627
                     /note="U7 repeat: matches 5. .64 of consensus"
     repeat_region   9988. .10420
                     /note="MSTB repeat: matches 426. .1 of consensus"
     repeat_region   10867. .11168
                     /note="AluY repeat: matches 301. .1 of consensus"
     repeat_region   11493. .11656
                     /note="L1MA10 repeat: matches 1069. .909 of consensus"
     repeat_region   11723. .12028
                     /note="AluSx repeat: matches 1. .302 of consensus"
     repeat_region   12044. .12396
                     /note="L1MA9 repeat: matches 885. .502 of consensus"
     repeat_region   12516. .12649
                     /note="AluSp repeat: matches 1. .136 of consensus;
                     incomplete repeat"
     repeat_region   12650. .12946
                     /note="AluSq repeat: matches 2. .303 of consensus"
     repeat_region   14416. .14555
                     /note="MIR2 repeat: matches 4. .145 of consensus"
     repeat_region   15010. .15325
                     /note="AluJb repeat: matches 1. .299 of consensus"
     repeat_region   16106. .16593
                     /note="AluSx repeat: matches 288. .2 of consensus"
     repeat_region   16928. .17149
                     /note="MIR repeat: matches 1. .249 of consensus"
     repeat_region   20950. .21246
                     /note="AluSg repeat: matches 1. .297 of consensus"
     prim_transcript <22115. .>25353
                     /note="match: multiple ESTs; match: AA134281 W29270
                     AA595272 AA192830 AA053700; match: AA223380 AA223568
                     AA247285 AA375743 T56497; match: AA053680 AA595227
                     AA743543 AA223263 T56347; match: AA223571 AA373511
                     AA192754 AA134282; match: AA857342"
     repeat_region   23086. .23374
                     /note="AluSq repeat: matches 303. .3 of consensus"
     repeat_region   25582. .25882
                     /note="AluSx repeat: matches 299. .1 of consensus"
     repeat_region   26495. .26799
                     /note="AluJo repeat: matches 1. .301 of consensus"
     repeat_region   27497. .27718
                     /note="MER42c repeat: matches 1537. .1317 of consensus"
     repeat_region   27771. .27895
                     /note="AluSg repeat: matches 174. .298 of consensus;
                     incomplete repeat"
     repeat_region   27896. .28093
                     /note="L1ME2 repeat: matches 890. .689 of consensus"
     misc_feature    28290. .28300
                     /note="single clone region"
     repeat_region   28375. .28487
```



```
                     /note="MIR2 repeat: matches 25. .146 of consensus"
     repeat_region   28890. .29186
                     /note="AluSx repeat: matches 2. .299 of consensus"
     repeat_region   29567. .29795
                     /note="AluJo repeat: matches 49. .302 of consensus;
                     incomplete repeat"
     repeat_region   29874. .30374
                     /note="MER1A repeat: matches 1. .527 of consensus"
     repeat_region   30778. .30987
                     /note="AluJo repeat: matches 4. .215 of consensus;
                     incomplete repeat"
     repeat_region   30992. .31040
                     /note="AluJ repeat: matches 251. .300 of consensus;
                     incomplete repeat"
     repeat_region   31265. .31456
                     /note="MIR repeat: matches 2. .214 of consensus"
     repeat_region   32450. .32743
                     /note="AluSq repeat: matches 2. .298 of consensus"
     repeat_region   33645. .33749
                     /note="MIR repeat: matches 105. .206 of consensus"
     repeat_region   34254. .34479
                     /note="MIR repeat: matches 260. .22 of consensus"
     repeat_region   34690. .34886
                     /note="MIR repeat: matches 13. .235 of consensus"
     repeat_region   34911. .35213
                     /note="AluSq repeat: matches 303. .1 of consensus"
     repeat_region   35392. .35542
                     /note="AluJb repeat: matches 137. .290 of consensus;
                     incomplete repeat"
     repeat_region   36790. .37023
                     /note="MLT1E repeat: matches 566. .312 of consensus"
     repeat_region   37142. .37441
                     /note="AluSq repeat: matches 303. .1 of consensus"
     repeat_region   37487. .38247
                     /note="L1MC1 repeat: matches 1071. .301 of consensus"
     repeat_region   38437. .38645
                     /note="AluJo repeat: matches 302. .98 of consensus;
                     incomplete repeat"
     repeat_region   38645. .38859
                     /note="AluJo repeat: matches 93. .301 of consensus;
                     incomplete repeat"
     repeat_region   38938. .39230
                     /note="AluSp repeat: matches 303. .3 of consensus"
     repeat_region   39249. .39798
                     /note="L1 repeat: matches 5172. .4624 of consensus"
     repeat_region   39962. .40040
                     /note="MLT1E repeat: matches 78. .1 of consensus"
     repeat_region   40824. .40872
                     /note="MIR2 repeat: matches 95. .146 of consensus"
...
Note: remainder of annotations omitted.
```

```
  Query Match           31.1%;  Score 19;  DB 26;  Length 102484;
  Best Local Similarity 100.0%;  Pred. No. 4.66e+00;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  55483 ATTGTTTTATTTTGTTTTT 55501
          ||||||||||||||||||| 
Cp     35 ATTGTTTTATTTTGTTTTT 17


RESULT   24
LOCUS       HUAC002551 114411 bp    DNA             PRI       17-MAR-1998
DEFINITION  Human Chromosome 16 BAC clone CIT987SK-A-951C11, complete sequence.
ACCESSION   AC002551
NID         g2809276
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 114411)
  AUTHORS   Adams,M.D., Loftus,B.J., Zhou,L., Crosby,M., Fuhrmann,J.,
            Brandon,R., Kim,U.J., Kerlavage,A.R. and Venter,J.C.
  TITLE     Human Chromosome 16 BAC clone CIT987SK-A-951C11
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 114411)
  AUTHORS   Adams,M.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-SEP-1997) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
REFERENCE   3  (bases 1 to 114411)
  AUTHORS   Adams,M.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JAN-1998) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
REFERENCE   4  (bases 1 to 114411)
  AUTHORS   Adams,M.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JAN-1998) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
REFERENCE   5  (bases 1 to 114411)
  AUTHORS   Adams,M.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-FEB-1998) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
REFERENCE   6  (bases 1 to 114411)
  AUTHORS   Adams,M.D. and Loftus,B.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAR-1998) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
COMMENT     On Jan 26, 1998 this sequence version replaced gi:2642422.
```



```
            Address all correspondence to: Mark Adams The Institute for Genomic
            Research 9712 Medical Center Dr, Rockville, MD 20850, USA e-mail
            address: mdadams@tigr.org.   The orientation of the sequence is
            from SP6 end to T7 end. Genes were identified by a combination of
            five methods including: XGRAIL (available by anonymous ftp from
            arthur.epm.ornl.gov), Genefinder (Phil Green, University of
            Washington), Genscan (Chris Burge, http://genomic.stanford.Edu/~
            chris/GENSCANW.html)searches of the complete sequence against a
            peptide database,and the Human gene Index database at TIGR
            (http://www.tigr.org/tdb/hg i/hgi.html). Genes without pepetide
            homolgy having spliced EST hits are termed 'Unknown gene product'.
            Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy,
            http://genome.wustl.edu/eddy/tRNAscan-SE/).
FEATURES                Location/Qualifiers
     source             1. .114411
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="16"
                        /map="16p12.1"
                        /clone="A-951C11"
     mRNA               complement(<3210. .3493)
                        /gene="A-951C1A1."
     gene               complement(<3210. .3493)
                        /gene="A-951C1A1."
     CDS                complement(<3210. .3430)
                        /gene="A-951C1A1."
                        /codon_start=1
                        /product="Transcription factor (TFIIIC) alpha chain, (3'
                        partial)"
                        /db_xref="PID:g2833636"
                        /translation="MDALESLLDEVALEGLDGLCLPALWSRLETRVPPFPLPLEPCTQ
                        EFLWRALATHPGISFYEEPRERPDLQLQDR"
     STS                6327. .6444
                        /db_xref="dbSTS:G02122"
     STS                1011138. .101255
                        /db_xref="dbSTS:G02122"
BASE COUNT    29730 a  26699 c  27271 g  30711 t
ORIGIN

  Query Match           31.1%;  Score 19;  DB 26;  Length 114411;
  Best Local Similarity 100.0%;  Pred. No. 4.66e+00;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  48156 TTGTTTTATTTTGTTTTTG 48174
          ||||||||||||||||||| 
Cp     34 TTGTTTTATTTTGTTTTTG 16


RESULT   25
LOCUS       CHMPXX     121024 bp    DNA   circular  PLN       28-OCT-1994
DEFINITION  Liverwort Marchantia polymorpha chloroplast genome DNA.
```



```
ACCESSION   X04465 Y00686
NID         g11640
KEYWORDS    16S ribosomal RNA; 23S ribosomal RNA; 4.5S ribosomal RNA; 5S
            ribosomal RNA; ATPase; chlorophyll a apoprotein; cytochrome;
            cytochrome b-559; cytochrome b6; cytochrome f; cytochrome subunit
            IV; F-0-ATPase; F1-ATPase; initiation factor; photosynthesis;
            photosystem II; ribosomal protein; ribosomal protein L14; ribosomal
            protein L16; ribosomal protein L20; ribosomal protein L21;
            ribosomal protein L22; ribosomal protein L23; ribosomal protein
            L33; ribosomal protein S11; ribosomal protein S12; ribosomal
            protein S14; ribosomal protein S15; ribosomal protein S18;
            ribosomal protein S19; ribosomal RNA; ribulose bisphosphate
            carboxylase; RNA polymerase; transfer RNA; transfer RNA-Ala;
            transfer RNA-Arg; transfer RNA-Asp; transfer RNA-Cys; transfer
            RNA-Gln; transfer RNA-Glu; transfer RNA-Gly; transfer RNA-His;
            transfer RNA-Ile; transfer RNA-Leu; transfer RNA-Lys; transfer
            RNA-Met; transfer RNA-Phe; transfer RNA-Pro; transfer RNA-Ser;
            transfer RNA-Thr; transfer RNA-Trp; transfer RNA-Tyr; transfer
            RNA-Val; unidentified reading frame.
SOURCE      liverwort.
  ORGANISM  Chloroplast Marchantia polymorpha
            Eukaryotae; mitochondrial eukaryotes; Viridiplantae;
            Charophyta/Embryophyta group; Embryophyta; Bryophyta;
            Hepaticopsida; Marchantiales; Marchantiaceae; Marchantia.
REFERENCE   1  (bases 1 to 121024)
  AUTHORS   Ohyama,K., Fukuzawa,H., Kohchi,T., Shirai,H., Sano,T., Sano,S.,
            Umesono,K., Shiki,Y., Takeuchi,M., Chang,Z., Aota,S., Inokuchi,H.
            and Ozeki,H.
  TITLE     Chloroplast gene organization deduced from complete sequence of
            liverwort Marchantia polymorpha chloroplast DNA
  JOURNAL   Nature 322, 572-574 (1986)
REFERENCE   2  (bases 1 to 121024)
  AUTHORS   Ohyama,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-OCT-1986) K. Ohyama, Research Centre for Cell and
            Tissue culture, Faculty of Agriculture, Kyoto University, Kyoto
            606, Japan
REFERENCE   3  (bases 1 to 121024)
  AUTHORS   Yamano,Y., Ohyama,K. and Komano,T.
  TITLE     Nucleotide sequences of chloroplast 5S ribosomal RNA from cell
            suspension cultures of the liverworts Marchantia polymorpha and
            Jungermannia subulata
  JOURNAL   Nucleic Acids Res. 12 (11), 4621-4624 (1984)
  MEDLINE   84247325
REFERENCE   4  (bases 1 to 121024)
  AUTHORS   Umesono,K., Inokuchi,H., Ohyama,K. and Ozeki,H.
  TITLE     Nucleotide sequence of Marchantia polymorpha chloroplast DNA: a
            region possibly encoding three tRNAs and three proteins including a
            homologue of E. coli ribosomal protein S14
  JOURNAL   Nucleic Acids Res. 12 (24), 9551-9565 (1984)
  MEDLINE   85087956
```



```
REFERENCE   5  (bases 1 to 121024)
  AUTHORS   Fukuzawa,H., Uchida,Y., Yamano,Y., Ohyama,K. and Komano,T.
  TITLE     Molecular cloning of promoters functional in Escherichia coli from
            chloroplast DNA of a liverwort, Marchantia polymorpha
  JOURNAL   Agric. Biol. Chem. 49, 2725-2731 (1985)
REFERENCE   6  (bases 1 to 121024)
  AUTHORS   Yamano,Y., Kohchi,T., Fukuzawa,H., Ohyama,K. and Komano,T.
  TITLE     Nucleotide sequences of chloroplast 4.5 S ribosomal RNA from a
            leafy liverwort, Jungermannia subulata, and a thalloid liverwort,
            Marchantia polymorpha
  JOURNAL   FEBS Lett. 185, 203-207 (1985)
REFERENCE   7  (bases 1 to 121024)
  AUTHORS   Fukuzawa,H., Kohchi,T., Shirai,H., Ohyama,K., Umesono,K.,
            Inokuchi,H. and Ozeki,H.
  TITLE     Coding sequences for chloroplast ribosomal protein S12 from the
            liverwort, Marchantia polymorpha, are separated far apart on the
            different DNA strands
  JOURNAL   FEBS Lett. 198, 11-15 (1986)
REFERENCE   8  (bases 1 to 121024)
  AUTHORS   Fukuzawa,H., Yoshida,T., Kohchi,T., Okumura,T., Sawano,Y. and
            Ohyama,K.
  TITLE     Splicing of group II introns in mRNAs coding for cytochrome b6 and
            subunit IV in liverwort Marchantia polymorpha chloroplast genome:
            Exon specifying a region coding for two genes with the spacer
            region
  JOURNAL   FEBS Lett. 220, 61-66 (1987)
REFERENCE   9  (bases 1 to 121024)
  AUTHORS   Kohchi,T., Ogura,Y., Umesono,K., Yamada,Y., Komano,T., Ozeki,H. and
            Ohyama,K.
  TITLE     Ordered processing and splicing in a polycistronic transcript in
            liverwort chloroplasts
  JOURNAL   Curr. Genet. 14 (2), 147-154 (1988)
  MEDLINE   89028845
REFERENCE   10 (bases 1 to 121024)
  AUTHORS   Fukuzawa,H., Kohchi,T., Sano,T., Shirai,H., Umesono,K.,
            Inokuchi,H., Ozeki,H. and Ohyama,K.
  TITLE     Structure and organization of Marchantia polymorpha chloroplast
            genome. III. Gene organization of the large single copy region from
            rbcL to trnI(CAU)
  JOURNAL   J. Mol. Biol. 203 (2), 333-351 (1988)
  MEDLINE   89068687
REFERENCE   11 (bases 1 to 121024)
  AUTHORS   Shimada,H. and Sugiura,M.
  TITLE     Fine structural features of the chloroplast genome: comparison of
            the sequenced chloroplast genomes
  JOURNAL   Nucleic Acids Res. 19 (5), 983-995 (1991)
  MEDLINE   91212240
REFERENCE   12 (bases 1 to 121024)
  AUTHORS   Ohyama,K., Fukuzawa,H., Kohchi,T., Sano,T., Sano,S., Shirai,H.,
            Umesono,K., Shiki,Y., Takeuchi,M., Chang,Z., Aota,S.i., Inokuchi,H.
            and Ozeki,H.
```

```
  TITLE     Structure and organization of Marchantia polymorpha chloroplast
            genome. I. Cloning and gene identification
  JOURNAL   J. Mol. Biol. 203 (2), 281-298 (1988)
  MEDLINE   89068685
REFERENCE   13 (bases 1 to 121024)
  AUTHORS   Umesono,K., Inokuchi,H., Shiki,Y., Takeuchi,M., Chang,Z.,
            Fukuzawa,H., Kohchi,T., Shirai,H., Ohyama,K. and Ozeki,H.
  TITLE     Structure and organization of Marchantia polymorpha chloroplast
            genome. II. Gene organization of the large single copy region from
            rps'12 to atpB
  JOURNAL   J. Mol. Biol. 203 (2), 299-331 (1988)
  MEDLINE   89068686
REFERENCE   14 (bases 1 to 121024)
  AUTHORS   Kohchi,T., Shirai,H., Fukuzawa,H., Sano,T., Komano,T., Umesono,K.,
            Inokuchi,H., Ozeki,H. and Ohyama,K.
  TITLE     Structure and organization of Marchantia polymorpha chloroplast
            genome. IV. Inverted repeat and small single copy regions
  JOURNAL   J. Mol. Biol. 203 (2), 353-372 (1988)
  MEDLINE   89068688
COMMENT     GENERAL DESCRIPTION OF SEQUENCE

                                                gene name   product name

            Ribosomal RNA genes                               16S rRNA
                                                              23S rRNA
                                                            4.5S rRNA
                                                              5S rRNA


            RNA polymerase genes

            alpha subunit                          rpoA: homologous to E.coli beta
            subunit                              rpoB: homologous to E.coli beta'
            subunit                            rpoC1: homologous to E.coli beta'
            subunit                            rpoC2: homologous to E.coli


            Ribosomal protein genes and related genes

            50S subunit
                                                 rpl2:  homologous to E. coli L2
            L14                                 rpl14:  homologous to E. coli L16
                                              rpl16:  homologous to E. coli L20
                                           rpl20:  homologous to E. coli L21
                                        rpl21:  homologous to E. coli L22
                                     rpl22:  homologous to E. coli L23
                                  rpl23:  homologous to E. coli L33
                               rpl33:  homologous to E. coli
```



```
            30S subunit
                                                 rps2:  homologous to E. coli S2
                                                 rps3:  homologous to E. coli S3
                                                 rps4:  homologous to E. coli S4
                                                 rps7:  homologous to E. coli S7
                                                 rps8:  homologous to E. coli S8
            S11                                 rpsl1:  homologous to E. coli S12
                                             rpsl2:  homologous to E. coli S14
                                          rps14:  homologous to E. coli S15
                                       rps15:  homologous to E. coli S18
                                    rps18:  homologous to E. coli S19
                                 rps19:  homologous to E. coli


            other genes                          infA:  initiation factor 1
                                                 secX:  homologous to E. coli


            Genes for photosynthesis
                                                 rbcL:  large subunit of RubisCO
                                                 psaA:  photosystem I P700
                                                        chlorophyll
                                                        (an apoprotein)
                                                 psaB:  same as above
                                                 psbA:  photosystem II 32kd
                                                        protein
                                                 psbB:  photosystem II P680
                                                        chlorophyll
                                                        (an apoprotein)
                                                 psbC:  same as above
                                                 psbD:  photosystem II D2 protein
                                                 psbE:  cytochrome b559
                                                 psbF:  same as above
                                                 psbG:  photosystem II G protein
                                                 atpA:  ATPase F1 subunit alpha
                                                 atpB:  ATPase F1 subunit beta
                                                 atpE:  ATPase F1 subunit epsilon
                                                 atpF:  ATPase F0 subunit I
                                                 atpH:  ATPase F0 subunit III
                                                 atpI:  ATPase F0 subunit IV
                                                 petA:  cytochrome f
                                                 petB:  cytochrome b6
                                                 petD:  subunit 4 of cytochrome
...
Note: remainder of annotations omitted.

  Query Match           31.1%;  Score 19;  DB 23;  Length 121024;
  Best Local Similarity 100.0%;  Pred. No. 4.66e+00;
  Matches    19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```



```
Db  49555 TTTTGTTTTTGAAAAGAAT 49573
          ||||||||||||||||||
Cp     26 TTTTGTTTTTGAAAAGAAT 8


RESULT   26
LOCUS       AC004027   133191 bp    DNA             PRI       22-JAN-1998
DEFINITION  Human PAC clone DJ0791I21 from 7q11.23-q21, complete sequence.
ACCESSION   AC004027
NID         g2804349
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 133191)
  AUTHORS   Graves,T., Tin-Wollam,A., Sutterer,C. and Keppler,D.
  TITLE     The sequence of H. sapiens PAC clone DJ0791I21
  JOURNAL   Unpublished (1998)
REFERENCE   2  (bases 1 to 133191)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JAN-1998) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     SUBMITTED BY:
            Genome Sequencing Center
            Department of Genetics
            Washington University
            St. Louis MO 63108, USA
            http://genome.wustl.edu/gsc
            mailto:sapiens@watson.wustl.edu

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded or sequenced with an alternate
            chemistry; an attempt was made to resolve all sequencing problems,
            such as compressions and repeats; all regions were covered by
            sequence from more than one subclone; and the assembly was
            confirmed by restriction digest.

            MAPPING INFORMATION:
            The sequence of this clone was established as part of a mapping and
            sequencing collaboration between the NHGRI Chromosome 7 Mapping
            Project (Eric D. Green, Director), John D. McPherson in the
            Department of Genetics (Washington University), and the Washington
            University Genome Sequencing Center.  For additional information
```



```
            about the map position of this sequence, see
            http://www.nhgri.nih.gov/DIR/GTB/CHR7 or send
            mailto:egreen@nhgri.nih.gov

            SOURCE INFORMATION:
            This clone was derived from human PAC library RPCI-4, prepared by
            Pieter de Jong and coworkers at Roswell Park Cancer Institute,
            using the method described by Ioannou et al., Nature Genetics
            6:84-9 (1994).  The library is from one male donor.  For further
            details, see http://bacpac.med.buffalo.edu/
            The clone is available from Genome Systems, Inc.
            (http://www.genomesystems.com).
            VECTOR:  pCYPAC2
            NEIGHBORING SEQUENCE INFORMATION:
            The actual start of this clone is at base position 1 of DJ0791I21;
            actual end is at 133191 of DJ0791I21. The orientation of this clone
            is unknown.

            This clone contains STS sWSS2762 (NID:g1113569).
FEATURES              Location/Qualifiers
     source           1. .133191
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="7"
                      /clone="DJ0791I21"
                      /clone_lib="RPCI-4"
                      /map="7q11.23-q21"
     repeat_region    997. .1304
                      /rpt_family="Alu"
     repeat_region    2244. .2803
                      /rpt_family="MER1_type"
     repeat_region    2849. .2946
                      /rpt_family="MER1_type"
     repeat_region    3053. .3207
                      /rpt_family="L2"
     repeat_region    6268. .6521
                      /rpt_family="Alu"
     repeat_region    6922. .7110
                      /rpt_family="Alu"
     repeat_region    7146. .7274
                      /rpt_family="Alu"
     repeat_region    9896. .9944
                      /rpt_family="L2"
     repeat_region    10008. .10228
                      /rpt_family="MaLR"
     repeat_region    10229. .10810
                      /rpt_family="Retroviral"
     repeat_region    10827. .10945
                      /rpt_family="MaLR"
     repeat_region    11065. .11102
                      /rpt_family="L2"
```

```
     repeat_region   11631. .11775
                     /rpt_family="MIR"
     repeat_region   12827. .13118
                     /rpt_family="Alu"
     repeat_region   13200. .13247
                     /rpt_family="L1"
     repeat_region   13251. .13532
                     /rpt_family="Alu"
     repeat_region   13560. .14045
                     /rpt_family="L1"
     repeat_region   14088. .14366
                     /rpt_family="L1"
     repeat_region   16194. .16258
                     /rpt_family="L2"
     repeat_region   16717. .16860
                     /rpt_family="L1"
     repeat_region   16919. .17319
                     /rpt_family="L1"
     repeat_region   18449. .18988
                     /rpt_family="L1"
     repeat_region   19035. .19448
                     /rpt_family="L1"
     repeat_region   19465. .20087
                     /rpt_family="L1"
     repeat_region   20088. .20830
                     /rpt_family="L1"
     repeat_region   20822. .20908
                     /rpt_family="L1"
     repeat_region   21964. .22023
                     /rpt_family="L2"
     repeat_region   22057. .22133
                     /rpt_family="L1"
     repeat_region   22154. .22225
                     /rpt_family="L1"
     repeat_region   25888. .27561
                     /rpt_family="L1"
     repeat_region   27556. .27810
                     /rpt_family="L1"
     repeat_region   29512. .29575
                     /rpt_family="MER1_type"
     repeat_region   30701. .30865
                     /rpt_family="MER1_type"
     repeat_region   32473. .32700
                     /rpt_family="MaLR"
     repeat_region   34090. .34364
                     /rpt_family="MER2_type"
     repeat_region   34803. .35379
                     /rpt_family="MER2_type"
     repeat_region   35374. .35462
                     /rpt_family="MER2_type"
     repeat_region   36408. .36501
```



```
                     /rpt_family="L2"
     repeat_region   38227. .38333
                     /rpt_family="L2"
     repeat_region   38549. .40154
                     /rpt_family="L1"
     repeat_region   40155. .40989
                     /rpt_family="L1"
     repeat_region   41029. .45671
                     /rpt_family="L1"
     repeat_region   45745. .46242
                     /rpt_family="L1"
     repeat_region   46843. .47018
                     /rpt_family="Alu"
     repeat_region   47988. .48050
                     /rpt_family="L2"
     repeat_region   48438. .48491
                     /rpt_family="L2"
     repeat_region   49735. .49840
                     /rpt_family="L2"
     repeat_region   50453. .50573
                     /rpt_family="MER1_type"
     repeat_region   53115. .53407
                     /rpt_family="L1"
     repeat_region   53395. .53447
                     /rpt_family="L1"
     repeat_region   53725. .54024
                     /rpt_family="Alu"
     repeat_region   54843. .55007
                     /rpt_family="Retroviral"
     repeat_region   56476. .56650
                     /rpt_family="L2"
     repeat_region   57586. .57704
                     /rpt_family="L1"
     repeat_region   58320. .58463
                     /rpt_family="Alu"
     repeat_region   59759. .59807
                     /rpt_family="MaLR"
     repeat_region   59925. .60047
                     /rpt_family="(TA)n"
     repeat_region   60501. .60565
                     /rpt_family="MaLR"
     repeat_region   61594. .61664
                     /rpt_family="MER2_type"
     repeat_region   61653. .62009
                     /rpt_family="MER2_type"
     repeat_region   62767. .62887
                     /rpt_family="(CATA)n"
     repeat_region   62870. .62988
                     /rpt_family="(CATA)n"
     repeat_region   63826. .64126
                     /rpt_family="Alu"
```



```
     repeat_region   64233. .64347
                     /rpt_family="MER1_type"
     repeat_region   64793. .65054
...
Note: remainder of annotations omitted.

  Query Match           31.1%;  Score 19;  DB 26;  Length 133191;
  Best Local Similarity 100.0%;  Pred. No. 4.66e+00;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  60968 TTGTTTTTGAAAAGAATGG 60986
          |||||||||||||||||||
Cp     24 TTGTTTTTGAAAAGAATGG 6


RESULT   27
LOCUS       HS359N14   136299 bp    DNA             PRI       29-OCT-1997
DEFINITION  Human DNA sequence from PAC 359N14 on chromosome 6q16.3-6q21.
            Contains CAF-I pseudogene, polymorphic CA repeat, STS.
ACCESSION   Z97206
NID         g2578094
KEYWORDS    6q16.3-6q21; CAF-I pseudogene; repeat polymorphism.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 136299)
  AUTHORS   Phillips,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-OCT-1997) sanger.ac.uk/HGP/Chr6/) Sanger Centre,
            Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquires:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 31, 1997 this sequence version replaced gi:2226410.
            IMPORTANT: This sequence is the entire insert of clone 359N14.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variations annotated may not be found in the sequence submission
            corresponding to the overlapping clone as we submit sequences with
            only a small overlap as described above.
            This sequence was generated from part of bacterial clone contigs of
            human chromosome 6, constructed by the Sanger Centre chromosome 6
            mapping group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6/
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            The true left end of clone 359N14 is at 1 in this sequence. The
```



```
            true right end of clone 359N14 is at 136299.
            359N14 is from the library RPCI3 constructed at the Roswell Park
            Cancer Institute by the group of Pieter de Jong.
            For further details see http://bacpac.med.buffalo.edu/.
FEATURES            Location/Qualifiers
     source         1. .136299
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="6"
                    /map="6q16.3-6q21"
                    /clone="359N14"
                    /clone_lib="RPCI3"
     repeat_region  569. .867
                    /note="AluSg repeat: matches 300. .1 of consensus"
     repeat_region  890. .1361
                    /note="L1MD2 repeat: matches 654. .178 of consensus"
     repeat_region  2458. .2501
                    /note="11 copies of 4 mer 91 % conserved"
     repeat_region  2728. .2874
                    /note="MIR repeat: matches 20. .168 of consensus"
     repeat_region  5738. .5952
                    /note="MIR repeat: matches 28. .253 of consensus"
     repeat_region  6476. .6589
                    /note="MIR repeat: matches 202. .89 of consensus"
     repeat_region  7266. .7638
                    /note="MSTC repeat: matches 404. .1 of consensus"
     repeat_region  8044. .8200
                    /note="MIR repeat: matches 15. .191 of consensus"
     repeat_region  9464. .9587
                    /note="MIR2 repeat: matches 25. .146 of consensus"
     repeat_region  9776. .9986
                    /note="AluSq repeat: matches 1. .211 of consensus;
                    incomplete repeat"
     repeat_region  13063. .13391
                    /note="L1MD2 repeat: matches 1088. .747 of consensus"
     repeat_region  13406. .13471
                    /note="33 copies of 2 mer 86 % conserved"
     repeat_region  13934. .14282
                    /note="THE1B repeat: matches 364. .1 of consensus"
     repeat_region  14527. .14664
                    /note="MIR repeat: matches 67. .208 of consensus"
     repeat_region  15028. .15389
                    /note="MSTA repeat: matches 426. .1 of consensus"
     repeat_region  17600. .17671
                    /note="AluYa5 repeat: matches 230. .301 of consensus;
                    incomplete repeat"
     repeat_region  19156. .19501
                    /note="MLT1F repeat: matches 199. .541 of consensus"
     repeat_region  19817. .20361
                    /note="MLT1F repeat: matches 5. .536 of consensus"
     repeat_region  21112. .21410
```

```
                     /note="AluY repeat: matches 1. .299 of consensus"
     misc_feature    complement(21486. .21866)
                     /note="match: Z24099 STS containing (CA) repeat"
     repeat_region   21737. .21775
                     /note="19 copies of GT 100% conserved; differs from
                     Z24099"
     repeat_region   21822. .22091
                     /note="AluJo repeat: matches 272. .1 of consensus;
                     incomplete repeat"
     repeat_region   22411. .22434
                     /note="12 copies of 2 mer 100 % conserved"
     misc_feature    complement(<23041. .23380)
                     /note="match: STS Z77962"
     repeat_region   24684. .24828
                     /note="AluSp repeat: matches 303. .159 of consensus;
                     incomplete repeat"
     repeat_region   24835. .24880
                     /note="MADE1 repeat: matches 35. .80 of consensus"
     repeat_region   25966. .26266
                     /note="AluSx repeat: matches 1. .302 of consensus"
     repeat_region   28121. .28355
                     /note="MIR repeat: matches 262. .6 of consensus"
     repeat_region   32306. .32363
                     /note="29 copies of 2 mer 95 % conserved"
     repeat_region   32364. .32662
                     /note="AluSp repeat: matches 301. .1 of consensus"
     repeat_region   32668. .33149
                     /note="L1MA9 repeat: matches 989. .527 of consensus"
     repeat_region   33180. .33368
                     /note="L1MA2 repeat: matches 545. .358 of consensus"
     repeat_region   33369. .33671
                     /note="AluJo repeat: matches 300. .1 of consensus"
     repeat_region   33674. .33980
                     /note="L1MA10 repeat: matches 365. .59 of consensus"
     repeat_region   34074. .34237
                     /note="MIR repeat: matches 208. .48 of consensus"
     misc_feature    34782. .34918
                     /note="match: STS Z78925"
     repeat_region   35348. .35389
                     /note="21 copies of 2 mer 100 % conserved"
     repeat_region   37666. .37707
                     /note="21 copies of 2 mer 83 % conserved"
     repeat_region   37976. .38043
                     /note="17 copies of 4 mer 97 % conserved"
     misc_feature    <38634. .39029
                     /note="match: STS G13699"
     repeat_region   41091. .41421
                     /note="MLT1A1 repeat: matches 1. .365 of consensus"
     repeat_region   41477. .41563
                     /note="MIR2 repeat: matches 63. .146 of consensus"
     repeat_region   41782. .41839
```



```
                     /note="MIR repeat: matches 140. .81 of consensus"
     repeat_region   43866. .43936
                     /note="MIR2 repeat: matches 74. .146 of consensus"
     repeat_region   45541. .45830
                     /note="AluJb repeat: matches 1. .290 of consensus"
     repeat_region   45867. .46113
                     /note="L1MA1 repeat: matches 1051. .780 of consensus"
     repeat_region   46470. .46770
                     /note="AluSq repeat: matches 1. .303 of consensus"
     repeat_region   49159. .49458
                     /note="AluSp repeat: matches 302. .2 of consensus"
     repeat_region   50480. .51090
                     /note="MER11B repeat: matches 632. .2 of consensus"
     repeat_region   50775. .51397
                     /note="MER11A repeat: matches 738. .60 of consensus"
     repeat_region   51396. .51544
                     /note="MER11A repeat: matches 151. .4 of consensus"
     repeat_region   52221. .52614
                     /note="MLT1B repeat: matches 1. .390 of consensus"
     repeat_region   52670. .52946
                     /note="AluSx repeat: matches 301. .28 of consensus;
                     incomplete repeat"
     repeat_region   52970. .53039
                     /note="MIR2 repeat: matches 69. .146 of consensus"
     repeat_region   54783. .54928
                     /note="MIR repeat: matches 85. .235 of consensus"
     repeat_region   54937. .55224
                     /note="AluY repeat: matches 301. .1 of consensus"
     repeat_region   55281. .55419
                     /note="MIR2 repeat: matches 4. .145 of consensus"
     repeat_region   60071. .60204
                     /note="MIR repeat: matches 257. .118 of consensus"
     repeat_region   65123. .65257
                     /note="MIR2 repeat: matches 142. .4 of consensus"
     repeat_region   67872. .68136
                     /note="L1 repeat: matches 4620. .4884 of consensus"
     repeat_region   68138. .68412
                     /note="AluSc repeat: matches 25. .299 of consensus;
                     incomplete repeat"
     repeat_region   68414. .68873
                     /note="L1 repeat: matches 4933. .5381 of consensus"
     repeat_region   68731. .69193
                     /note="L1MB4 repeat: matches 4. .466 of consensus"
     repeat_region   71931. .72230
                     /note="AluY repeat: matches 301. .1 of consensus"
     repeat_region   73816. .73927
                     /note="MIR2 repeat: matches 24. .139 of consensus"
     repeat_region   76548. .76654
                     /note="MIR repeat: matches 186. .73 of consensus"
     repeat_region   77216. .77287
                     /note="2 copies of 36 mer 93 % conserved"
```



```
     repeat_region   77375. .77400
                     /note="13 copies of 2 mer 92 % conserved"
     repeat_region   79357. .79624
                     /note="AluJb repeat: matches 1. .295 of consensus"
     repeat_region   82707. .82954
                     /note="MIR repeat: matches 262. .13 of consensus"
     repeat_region   85153. .85278
                     /note="MIR repeat: matches 75. .212 of consensus"
     repeat_region   85385. .85684
                     /note="AluSx repeat: matches 299. .1 of consensus"
     repeat_region   87813. .88111
                     /note="AluSg repeat: matches 1. .299 of consensus"
...
Note: remainder of annotations omitted.

  Query Match           31.1%;  Score 19;  DB 26;  Length 136299;
  Best Local Similarity 100.0%;  Pred. No. 4.66e+00;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db 119022 TTGTTTTATTTTGTTTTTG 119040
          |||||||||||||||||||
Cp     34 TTGTTTTATTTTGTTTTTG 16


RESULT   28
LOCUS       AC004987   144136 bp    DNA             HTG       12-JUN-1998
DEFINITION  *** SEQUENCING IN PROGRESS *** Homo sapiens clone DJ1173I20; HTGS
            phase 1, 5 unordered pieces.
ACCESSION   AC004987
NID         g3213010
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 144136)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 144136)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-1998) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     ***                                                          ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                          ***
            * This sequence is unfinished. It consists of 5 contigs for
            * which the order is not known; their order in this record is
            * arbitrary. In some cases, the exact lengths of the gaps
```



```
            * between the contigs are also unknown; these gaps are  presented
            * as runs of N as a convenience only. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
            *       1     1706: contig of 1706 bp in length
            *    1707     1724: gap of unknown length
            *    1725     3182: contig of 1458 bp in length
            *    3183     3200: gap of unknown length
            *    3201     5080: contig of 1880 bp in length
            *    5081     5098: gap of unknown length
            *    5099    82140: contig of 77042 bp in length
            *   82141    82158: gap of unknown length
            *   82159   144136: contig of 61978 bp in length.
FEATURES              Location/Qualifiers
     source           1. .144136
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="DJ1173I20"
BASE COUNT    40691 a  32186 c  30937 g  40250 t     72 others
ORIGIN

  Query Match           31.1%;  Score 19;  DB 17;  Length 144136;
  Best Local Similarity 100.0%;  Pred. No. 4.66e+00;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  53390 TTGCAGTCTTTATTGTTTT 53408
          |||||||||||||||||||
Cp     46 TTGCAGTCTTTATTGTTTT 28


RESULT   29
LOCUS       HS441J1    153687 bp    DNA             HTG       01-OCT-1998
DEFINITION  Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone 441J1;
            HTGS phase 1.
ACCESSION   Z99495
NID         g3688364
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 153687)
  AUTHORS   Tubby,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-1998) Wellcome Trust Genome Campus, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquires:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 2, 1998 this sequence version replaced gi:3367610.
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and the release of this data is based on the understanding that the
```

```
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc. Order of segments is not known; 800 n's separate
            segments. Unfinished: dJ441J1  Contig_ID: 01209  acc=Z99495
            Length: 151213 bp Unfinished: dJ441J1  Contig_ID: 02978  acc=Z99495
             Length: 1674 bp.
            ***                                                    ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                    ***
            * This sequence is unfinished. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
FEATURES             Location/Qualifiers
     source          1. .153687
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="441J1"
BASE COUNT    45338 a  29440 c  30240 g  47830 t    839 others
ORIGIN

  Query Match           31.1%;  Score 19;  DB 17;  Length 153687;
  Best Local Similarity 100.0%;  Pred. No. 4.66e+00;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  13808 CAAAAACAAAATAAAACAA 13826
          ||||||||||||||||||
Qy     16 CAAAAACAAAATAAAACAA 34


RESULT   30
LOCUS       HS500L14   160218 bp    DNA             HTG       02-OCT-1998
DEFINITION  Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone
            500L14; HTGS phase 1.
ACCESSION   AL023583
NID         g3702135
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 160218)
  AUTHORS   Mashreghi-Mohammadi,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-OCT-1998) Wellcome Trust Genome Campus, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquires:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 6, 1998 this sequence version replaced gi:3676210.
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and the release of this data is based on the understanding that the
```



```
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc. Order of segments is not known; 800 n's separate
            segments. Unfinished: dJ500L14  Contig_ID: 02706  acc=AL023583
            Length: 19258 bp Unfinished: dJ500L14  Contig_ID: 02779
            acc=AL023583  Length: 11687 bp Unfinished: dJ500L14  Contig_ID:
            00319  acc=AL023583  Length: 28901 bp Unfinished: dJ500L14
            Contig_ID: 02287  acc=AL023583  Length: 1050 bp Unfinished:
            dJ500L14  Contig_ID: 01618  acc=AL023583  Length: 1585 bp
            Unfinished: dJ500L14  Contig_ID: 00728  acc=AL023583  Length: 1424
            bp Unfinished: dJ500L14  Contig_ID: 02540  acc=AL023583  Length:
            1014 bp Unfinished: dJ500L14  Contig_ID: 01799  acc=AL023583
            Length: 1244 bp Unfinished: dJ500L14  Contig_ID: 01123
            acc=AL023583  Length: 4745 bp Unfinished: dJ500L14  Contig_ID:
            03094  acc=AL023583  Length: 59868 bp Unfinished: dJ500L14
            Contig_ID: 02910  acc=AL023583  Length: 18173 bp Unfinished:
            dJ500L14  Contig_ID: 03010  acc=AL023583  Length: 2469 bp.
            ***                                                    ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                    ***
            * This sequence is unfinished. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
FEATURES             Location/Qualifiers
     source          1. .160218
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="500L14"
BASE COUNT    42111 a  33002 c  32850 g  43453 t   8802 others
ORIGIN

  Query Match           31.1%;  Score 19;  DB 17;  Length 160218;
  Best Local Similarity 100.0%;  Pred. No. 4.66e+00;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  67436 TTGTTTTATTTTGTTTTTG 67454
          ||||||||||||||||||
Cp     34 TTGTTTTATTTTGTTTTTG 16


RESULT   31
LOCUS       HS323A16   161146 bp    DNA             HTG       08-OCT-1998
DEFINITION  Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone
            323A16; HTGS phase 1.
ACCESSION   AL031300
NID         g3724228
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
```



```
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 161146)
  AUTHORS   Blakey,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-OCT-1998) Wellcome Trust Genome Campus, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquires:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 9, 1998 this sequence version replaced gi:3702436.
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and the release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc. Order of segments is not known; 800 n's separate
            segments. Unfinished: dJ323A16  Contig_ID: 02281  acc=AL031300
            Length: 16010 bp Unfinished: dJ323A16  Contig_ID: 01438
            acc=AL031300  Length: 144336 bp.
            ***                                                    ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                    ***
            * This sequence is unfinished. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
FEATURES             Location/Qualifiers
     source          1. .161146
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="22"
                     /clone="323A16"
BASE COUNT    45582 a  35940 c  34782 g  43983 t    859 others
ORIGIN

  Query Match           31.1%;  Score 19;  DB 17;  Length 161146;
  Best Local Similarity 100.0%;  Pred. No. 4.66e+00;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db 150128 ATTGTTTTATTTTGTTTTT 150146
          ||||||||||||||||||
Cp     35 ATTGTTTTATTTTGTTTTT 17


RESULT   32
LOCUS       AC004129   162701 bp    DNA             PRI       06-FEB-1998
DEFINITION  Homo sapiens BAC clone RG136N17 from 7p15-p21, complete sequence.
ACCESSION   AC004129
NID         g2842787
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
```



```
REFERENCE   1  (bases 1 to 162701)
  AUTHORS   Stoneking,T., Langston,Y. and Courtney,L.
  TITLE     The sequence of Homo sapiens BAC clone RG136N17
  JOURNAL   Unpublished (1998)
REFERENCE   2  (bases 1 to 162701)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-FEB-1998) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     SUBMITTED BY:  WUGSC
            Genome Sequencing Center
            Department of Genetics
            Washington University
            St. Louis MO 63108, USA
            http://genome.wustl.edu/gsc
            mailto:sapiens@watson.wustl.edu

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded or sequenced with an alternate
            chemistry; an attempt was made to resolve all sequencing problems,
            such as compressions and repeats; all regions were covered by
            sequence from more than one subclone; and the assembly was
            confirmed by restriction digest.

            MAPPING INFORMATION:
            The sequence of this clone was established as part of a mapping and
            sequencing collaboration between the NHGRI Chromosome 7 Mapping
            Project (Eric D. Green, Director), John D. McPherson in the
            Department of Genetics (Washington University), and the Washington
            University Genome Sequencing Center.  For additional information
            about the map position of this sequence, see
            http://www.nhgri.nih.gov/DIR/GTB/CHR7 or send
            mailto:egreen@nhgri.nih.gov

            SOURCE INFORMATION:
            Clone RG136N17 is from the first release of the human BAC library
            CITB-978SK-B.  The library contains cloned DNA from the male
            fibroblast cell line 978SK.  See:  Shizuya et al., Proc. Natl.
            Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8
            (1996).  This clone is available from Research Genetics, Inc.
            (http://www.resgen.com).
            VECTOR:  pBeloBAC11
            Selection:  chloramphenicol

            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is DJ0978I12, 200 bp overlap.
```

```
            Actual start of this clone is at base position 98702 of DJ0978I12;
            actual end is at 62701 of RG136N17.

            This clone contains STS sWSS3049 (NID:g1234137).
  FEATURES             Location/Qualifiers
       source          1. .162701
                       /organism="Homo sapiens"
                       /db_xref="taxon:9606"
                       /chromosome="7"
                       /clone="RG136N17"
                       /clone_lib="CITB-978SK-B"
                       /map="7p15-p21"
       repeat_region   6. .344
                       /rpt_family="MER4-group"
       repeat_region   390. .638
                       /rpt_family="MER21_g"
       repeat_region   1778. .1819
                       /rpt_family="MER1_type"
       repeat_region   2946. .3071
                       /rpt_family="Alu"
       repeat_region   4799. .4975
                       /rpt_family="MER1_type"
       repeat_region   5593. .6071
                       /rpt_family="MaLR"
       repeat_region   7918. .8385
                       /rpt_family="Retroviral"
       repeat_region   9110. .9170
                       /rpt_family="MIR"
       repeat_region   9455. .9745
                       /rpt_family="Alu"
       repeat_region   10643. .10701
                       /rpt_family="L2"
       repeat_region   10703. .10786
                       /rpt_family="MER1_type"
       repeat_region   11651. .12121
                       /rpt_family="MaLR"
       repeat_region   13509. .13635
                       /rpt_family="MIR"
       repeat_region   14128. .14361
                       /rpt_family="MER1_type"
       repeat_region   16497. .17204
                       /rpt_family="L1"
       repeat_region   17205. .17272
                       /rpt_family="Alu"
       repeat_region   17274. .17335
                       /rpt_family="L1"
       repeat_region   17348. .17645
                       /rpt_family="Alu"
       repeat_region   20806. .20958
                       /rpt_family="L2"
       repeat_region   20966. .21150
```



```
                       /rpt_family="L2"
       repeat_region   21715. .21886
                       /rpt_family="MER1_type"
       repeat_region   22352. .22475
                       /rpt_family="L2"
       repeat_region   22686. .22911
                       /rpt_family="U3"
       repeat_region   22957. .23244
                       /rpt_family="L2"
       repeat_region   23387. .23693
                       /rpt_family="L2"
       repeat_region   28559. .28610
                       /rpt_family="(GA)n"
       repeat_region   28610. .28682
                       /rpt_family="(CA)n"
       repeat_region   29493. .29622
                       /rpt_family="MER1_type"
       repeat_region   32042. .32168
                       /rpt_family="L2"
       repeat_region   32751. .33427
                       /rpt_family="L1"
       repeat_region   33471. .34261
                       /rpt_family="MaLR"
       repeat_region   34262. .34720
                       /rpt_family="MaLR"
       repeat_region   36535. .36700
                       /rpt_family="Retroviral"
       repeat_region   37988. .38272
                       /rpt_family="Alu"
       repeat_region   38803. .39494
                       /rpt_family="Retroviral"
       repeat_region   39522. .39625
                       /rpt_family="MER1_type?"
       repeat_region   40089. .40300
                       /rpt_family="L2"
       repeat_region   40485. .40529
                       /rpt_family="MIR"
       repeat_region   40796. .40865
                       /rpt_family="MIR"
       repeat_region   40868. .41054
                       /rpt_family="MER1_type"
       repeat_region   41527. .41539
                       /rpt_family="MIR"
       repeat_region   41540. .41579
                       /rpt_family="L2"
       repeat_region   41580. .41589
                       /rpt_family="MIR"
       repeat_region   43328. .43365
                       /rpt_family="MER1_type"
       repeat_region   44891. .44960
                       /rpt_family="L2"
```



```
       repeat_region   45321. .45608
                       /rpt_family="Alu"
       repeat_region   47651. .47862
                       /rpt_family="Alu"
       repeat_region   50102. .50235
                       /rpt_family="Alu"
       repeat_region   50236. .50294
                       /rpt_family="(TAGA)n"
       repeat_region   50299. .50461
                       /rpt_family="Alu"
       repeat_region   51076. .51458
                       /rpt_family="MER4-group"
       repeat_region   52311. .52604
                       /rpt_family="Alu"
       repeat_region   54812. .55067
                       /rpt_family="L2"
       repeat_region   55139. .55399
                       /rpt_family="Alu"
       repeat_region   55580. .55776
                       /rpt_family="MIR"
       repeat_region   56471. .56658
                       /rpt_family="L2"
       repeat_region   56950. .57241
                       /rpt_family="Alu"
       repeat_region   57563. .57613
                       /rpt_family="(CA)n"
       misc_feature    59102. .59266
                       /note="match to EST AA017622 (NID:g1479776) ze37g05.r1"
       misc_feature    60535. .60738
                       /note="match to EST AA017622 (NID:g1479776) ze37g05.r1"
       repeat_region   61627. .61802
                       /rpt_family="L2"
       repeat_region   64712. .64860
                       /rpt_family="MER1_type"
       repeat_region   67356. .67525
                       /rpt_family="L1"
       repeat_region   67756. .67936
                       /rpt_family="MER1_type"
...
Note: remainder of annotations omitted.

  Query Match           31.1%;  Score 19;  DB 26;  Length 162701;
  Best Local Similarity 100.0%;  Pred. No. 4.66e+00;
  Matches    19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db 112299 AAAACAAAATAAAACAATA 112317
          |||||||||||||||||||
Qy     18 AAAACAAAATAAAACAATA 36


RESULT   33
```



```
LOCUS       AC005089   177137 bp    DNA          HTG       12-JUN-1998
DEFINITION  *** SEQUENCING IN PROGRESS *** Homo sapiens clone RG315H11; HTGS
            phase 1, 5 unordered pieces.
ACCESSION   AC005089
NID         g3212902
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 177137)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 177137)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-1998) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     ***                                                          ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                          ***
            * This sequence is unfinished. It consists of 5 contigs for
            * which the order is not known; their order in this record is
            * arbitrary. In some cases, the exact lengths of the gaps
            * between the contigs are also unknown; these gaps are  presented
            * as runs of N as a convenience only. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
            *        1     1727: contig of 1727 bp in length
            *     1728     1744: gap of unknown length
            *     1745    29208: contig of 27464 bp in length
            *    29209    29225: gap of unknown length
            *    29226    58247: contig of 29022 bp in length
            *    58248    58264: gap of unknown length
            *    58265    95900: contig of 37636 bp in length
            *    95901    95917: gap of unknown length
            *    95918   177137: contig of 81220 bp in length.
FEATURES             Location/Qualifiers
     source          1. .177137
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RG315H11"
BASE COUNT    44550 a  42732 c  43015 g  46747 t     93 others
ORIGIN

  Query Match           31.1%;  Score 19;  DB 17;  Length 177137;
  Best Local Similarity 100.0%;  Pred. No. 4.66e+00;
  Matches    19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```

```
Db  75716 TTTCAAAAACAAAATAAAA 75734
          |||||||||||||||||||
Qy     13 TTTCAAAAACAAAATAAAA 31


RESULT   34
LOCUS       AC005639   183039 bp    DNA             HTG       26-SEP-1998
DEFINITION  *** SEQUENCING IN PROGRESS *** R48M01; D397; Drosophila
            melanogaster; Chromosome 2R; Region 59E3-59F4; HTGS phase 1, 31
            unordered pieces.
ACCESSION   AC005639
NID         g3659427
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea;
            Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 183039)
  AUTHORS   Celniker,S.E., George,R.A., Galle,R., Svirskas,R.R., Hoskins,R.A.,
            Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Chavez,C.,
            Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A.,
            Hummasti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B.,
            Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S., Moshrefi,A.R.,
            Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,E.,
            Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R.,
            Zieran,L.L. and Kimmel,B.E.
  TITLE     Sequencing of Drosophila melanogaster
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 183039)
  AUTHORS   Celniker,S.E., George,R.A., Galle,R., Svirskas,R.R., Hoskins,R.A.,
            Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Chavez,C.,
            Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A.,
            Hummasti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B.,
            Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S., Moshrefi,A.R.,
            Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,E.,
            Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R.,
            Zieran,L.L. and Kimmel,B.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-SEP-1998) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Sep 26, 1998 this sequence version replaced gi:3581660.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://fruitfly.berkeley.edu/sequence/) or send
            email to drosophila@mhgc.lbl.gov. All contigs in this submission
            meet the following cutoffs: length >= 400 bases, phrap computed
            error rate <= 1/10.
            ***                                                          ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                          ***
```



```
            * This sequence is unfinished. It consists of 31 contigs for
            * which the order is not known; their order in this record is
            * arbitrary. In some cases, the exact lengths of the gaps
            * between the contigs are also unknown; these gaps are  presented
            * as runs of N as a convenience only. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
            *       1      743: contig of 743 bp in length
            *     744      823: gap of unknown length
            *     824     1617: contig of 794 bp in length
            *    1618     1697: gap of unknown length
            *    1698     3510: contig of 1813 bp in length
            *    3511     3590: gap of unknown length
            *    3591     4628: contig of 1038 bp in length
            *    4629     4708: gap of unknown length
            *    4709     6233: contig of 1525 bp in length
            *    6234     6313: gap of unknown length
            *    6314     7601: contig of 1288 bp in length
            *    7602     7681: gap of unknown length
            *    7682    10044: contig of 2363 bp in length
            *   10045    10124: gap of unknown length
            *   10125    13551: contig of 3427 bp in length
            *   13552    13631: gap of unknown length
            *   13632    15386: contig of 1755 bp in length
            *   15387    15466: gap of unknown length
            *   15467    18221: contig of 2755 bp in length
            *   18222    18301: gap of unknown length
            *   18302    20884: contig of 2583 bp in length
            *   20885    20964: gap of unknown length
            *   20965    23410: contig of 2446 bp in length
            *   23411    23490: gap of unknown length
            *   23491    26709: contig of 3219 bp in length
            *   26710    26789: gap of unknown length
            *   26790    29297: contig of 2508 bp in length
            *   29298    29377: gap of unknown length
            *   29378    31273: contig of 1896 bp in length
            *   31274    31353: gap of unknown length
            *   31354    33520: contig of 2167 bp in length
            *   33521    33600: gap of unknown length
            *   33601    37290: contig of 3690 bp in length
            *   37291    37370: gap of unknown length
            *   37371    41172: contig of 3802 bp in length
            *   41173    41252: gap of unknown length
            *   41253    45420: contig of 4168 bp in length
            *   45421    45500: gap of unknown length
            *   45501    51505: contig of 6005 bp in length
            *   51506    51585: gap of unknown length
            *   51586    59111: contig of 7526 bp in length
            *   59112    59191: gap of unknown length
            *   59192    66424: contig of 7233 bp in length
            *   66425    66504: gap of unknown length
```



```
            *   66505    74174: contig of 7670 bp in length
            *   74175    74254: gap of unknown length
            *   74255    81093: contig of 6839 bp in length
            *   81094    81173: gap of unknown length
            *   81174    92799: contig of 11626 bp in length
            *   92800    92879: gap of unknown length
            *   92880   103090: contig of 10211 bp in length
            *  103091   103170: gap of unknown length
            *  103171   115910: contig of 12740 bp in length
            *  115911   115990: gap of unknown length
            *  115991   127733: contig of 11743 bp in length
            *  127734   127813: gap of unknown length
            *  127814   140390: contig of 12577 bp in length
            *  140391   140470: gap of unknown length
            *  140471   159016: contig of 18546 bp in length
            *  159017   159096: gap of unknown length
            *  159097   183039: contig of 23943 bp in length.
FEATURES             Location/Qualifiers
     source          1. .183039
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="R48M01; D397"
                     /chromosome="2"
                     /map="59E3-59F4"
BASE COUNT    49233 a  41493 c  41552 g  48347 t   2414 others
ORIGIN

  Query Match          31.1%;  Score 19;  DB 17;  Length 183039;
  Best Local Similarity 100.0%;  Pred. No. 4.66e+00;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db 154953 ATTGTTTTATTTTGTTTTT 154971
          |||||||||||||||||||
Cp     35 ATTGTTTTATTTTGTTTTT 17


RESULT   35
LOCUS       AC003066   185402 bp    DNA             HTG       30-JAN-1998
DEFINITION  *** SEQUENCING IN PROGRESS *** Mus musculus Chromosome 16 BAC Clone
            tbx1 Syntenic To Homo sapiens 22q11.2 DGCR Region; HTGS phase 1, 8
            unordered pieces.
ACCESSION   AC003066
NID         g2815552
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 185402)
  AUTHORS   Zhang,G. and Roe,B.A.
  TITLE     Mouse Chromosome 16 BAC Clone tbx1
```



```
  JOURNAL   Unpublished (1997)
REFERENCE   2  (bases 1 to 185402)
  AUTHORS   Emanuel,B. and Budarf,M.
  TITLE     Mouse Chromosome 16 BAC Clone tbx1
  JOURNAL   Unpublished (1997)
REFERENCE   3  (bases 1 to 185402)
  AUTHORS   Zhang,G., Lao,V., Zhan,M. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-NOV-1997) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On Jan 28, 1998 this sequence version replaced gi:2588595.
            ***                                                          ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                          ***
            * This sequence is unfinished. It consists of 8 contigs for
            * which the order is not known; their order in this record is
            * arbitrary. In some cases, the exact lengths of the gaps
            * between the contigs are also unknown; these gaps are  presented
            * as runs of N as a convenience only. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
            *       1     2770: contig of 2770 bp in length
            *    2771     2849: gap of unknown length
            *    2850     7558: contig of 4709 bp in length
            *    7559     7637: gap of unknown length
            *    7638    14539: contig of 6902 bp in length
            *   14540    14617: gap of unknown length
            *   14618    26462: contig of 11845 bp in length
            *   26463    26540: gap of unknown length
            *   26541    50021: contig of 23481 bp in length
            *   50022    50099: gap of unknown length
            *   50100    85188: contig of 35089 bp in length
            *   85189    85266: gap of unknown length
            *   85267   127801: contig of 42535 bp in length
            *  127802   127879: gap of unknown length
            *  127880   185402: contig of 57523 bp in length.
FEATURES             Location/Qualifiers
     source          1. .185402
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="tbx1"
                     /chromosome="16"
BASE COUNT    46102 a  45861 c  45604 g  47283 t    552 others
ORIGIN

  Query Match          31.1%;  Score 19;  DB 17;  Length 185402;
  Best Local Similarity 100.0%;  Pred. No. 4.66e+00;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  72725 TTGTTTTATTTTGTTTTTG 72743
```

```
          ||||||||||||||||||||
Cp     34 TTGTTTTATTTTGTTTTTG 16


RESULT   36
LOCUS       AC004885   192078 bp    DNA             HTG       12-JUN-1998
DEFINITION  *** SEQUENCING IN PROGRESS *** Homo sapiens clone DJ0781A18; HTGS
            phase 1, 3 unordered pieces.
ACCESSION   AC004885
NID         g3213112
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 192078)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 192078)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-1998) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     ***                                                        ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                        ***
            * This sequence is unfinished. It consists of 3 contigs for
            * which the order is not known; their order in this record is
            * arbitrary. In some cases, the exact lengths of the gaps
            * between the contigs are also unknown; these gaps are  presented
            * as runs of N as a convenience only. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
            *     1     1722: contig of 1722 bp in length
            *     1723  1740: gap of unknown length
            *     1741  111151: contig of 109411 bp in length
            *   111152  111169: gap of unknown length
            *   111170  192078: contig of 80909 bp in length.
FEATURES             Location/Qualifiers
     source          1. .192078
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="DJ0781A18"
BASE COUNT    54011 a  37929 c  40588 g  59514 t     36 others
ORIGIN

  Query Match           31.1%;  Score 19;  DB 17;  Length 192078;
  Best Local Similarity 100.0%;  Pred. No. 4.66e+00;
  Matches    19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```



```
Db  60655 TTTTATTTTGTTTTTGAAA 60673
          ||||||||||||||||||||
Cp     31 TTTTATTTTGTTTTTGAAA 13


RESULT   37
LOCUS       CEY1A5     196643 bp    DNA             HTG       09-NOV-1997
DEFINITION  Caenorhabditis elegans DNA *** SEQUENCING IN PROGRESS *** from
            clone Y1A5; HTGS phase 1.
ACCESSION   AL008872
NID         g2612774
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryotae; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 196643)
  AUTHORS   Sulston,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-NOV-1997) Nematode Sequencing Project, Sanger Centre,
            Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics,
            Washington University, St. Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and the release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc. Order of segments is not known; 800 n's separate
            segments.
            ***                                                        ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                        ***
            * This sequence is unfinished. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
FEATURES             Location/Qualifiers
     source          1. .196643
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /chromosome="III"
                     /clone="Y1A5"
BASE COUNT    63438 a  34352 c  35398 g  61405 t   2050 others
ORIGIN

  Query Match           31.1%;  Score 19;  DB 17;  Length 196643;
  Best Local Similarity 100.0%;  Pred. No. 4.66e+00;
  Matches    19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  10603 AAAATAAAACAATAAAGAC 10621
```



```
          ||||||||||||||||||||
Qy     23 AAAATAAAACAATAAAGAC 41


RESULT   38
LOCUS       G08858       309 bp    DNA             STS       05-FEB-1997
DEFINITION  human STS CHLC.GATA22D05.P8973 clone GATA22D05, sequence tagged
            site.
ACCESSION   G08858
NID         g939408
KEYWORDS    STS; STS sequence; primer; sequence tagged site.
SOURCE      human vector=pJCP1 host=E.coli dut+ung+ (DH10B) Marker Selected
            genomic DNA prepared from XY individual of French nationality.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 309)
  AUTHORS   Murray,J., Sheffield,V, Weber,J.L., Duyk,G. and Buetow,K.H.
  TITLE     Cooperative Human Linkage Center
  JOURNAL   Unpublished (1995)
COMMENT     Synonyms: GATA22D05, CHLC.GATA22D05.#T8972
            GDB: G00-364-213
            Contact: Dr. Jeffrey C. Murray
            UofI
            The University of Iowa
            Department of Pediatrics, Iowa City, IA 52242, USA
            Tel: (319) 356-3508
            Fax: (319) 356-3347
            Email: jeff-murray@uiowa.edu

            Primer A: CATCATCTGCACTCCACAAA
            Primer B: TGTCCATCGACAGATGAATG
            STS size: 286
            PCR Profile:
                        denature:    30 seconds at 94 degrees C
                        annealing:   75 seconds at 55 degrees C
                        extension:   15 seconds at 72 degrees C
                        PCR cycles:  27
                        extension:    6 minutes at 72 degrees C
            Protocol:
                        Template:        30ng genomic DNA
                        Primer:          each 1.5 pmole
                        dNTPs:           each 200 uM
                        Taq Polymerase:  0.3 units
                        Total Vol:       10 ul

            Buffer:
                        MgCl2:  1.5mM
                        KCl:     50mM
                        Tris:    10mM
                        pH:     8.3.
```



```
FEATURES             Location/Qualifiers
     source          1. .309
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     STS             12. .297
     primer_bind     12. .31
     primer_bind     complement(278. .297)
BASE COUNT      122 a      35 c      55 g      93 t       4 others
ORIGIN

  Query Match           29.5%;  Score 18;  DB 30;  Length 309;
  Best Local Similarity 100.0%;  Pred. No. 1.72e+01;
  Matches    18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     57 AAAAACAAAATAAAACAA 74
          ||||||||||||||||||
Qy     17 AAAAACAAAATAAAACAA 34


RESULT   39
LOCUS       G35487       351 bp    DNA             STS       07-OCT-1997
DEFINITION    STS h14a397 5, sequence tagged site.
ACCESSION   G35487
NID         g2483247
KEYWORDS    STS; STS sequence; primer; sequence tagged site.
SOURCE      human vector=ml3mp18 Genomic repeat element-mediated PCR products
            cloned into M13.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 351)
  AUTHORS   Dear,P.H., Bankier,A.T. and Piper,M.B.
  TITLE     A high resolution metric HAPPY map of human chromosome 14
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Paul H. Dear
            Laboratory of Molecular Biology
            Medical Research Council
            Addenbrookes Hospital, Hills Road, Cambridge, CB2 2QH, England
            Fax: 44 (0) 1223 412178
            Email: phd@mrc-lmb.cam.ac.uk

            Primer A: TCACATCAAGTCCAGACACC
            Primer B: GAAGCAGGCTTCTCCATGG
            STS size: 132
            Protocol:
                 Template:     10 - 50 ng
                 Primer:         1 uM each
                 dNTPs           200 uM each
                 Taq Pol:        0.025 units/ul
                 Total Vol:    10 ul
```

```
           Buffer:
                MgCl2:          1 mM
                KCl:            50mM
                Tris HCl        10 mM
                pH:              8.3

           Alu-PCR product from mouse/human hybrid cell line 10479 (Coriell
           Cell Repositories).
FEATURES             Location/Qualifiers
     source          1. .351
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     STS             89. .220
     primer_bind     89. .108
     primer_bind     complement(202. .220)
BASE COUNT      117 a      92 c      78 g      60 t       4 others
ORIGIN

  Query Match           29.5%;  Score 18;  DB 30;  Length 351;
  Best Local Similarity 100.0%;  Pred. No. 1.72e+01;
  Matches    18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     45 AAAAACAAAATAAAACAA 62
          ||||||||||||||||||
Qy     17 AAAAACAAAATAAAACAA 34


RESULT   40
LOCUS       HSU3B2        367 bp    DNA              PRI       03-JAN-1998
DEFINITION  Homo sapiens U3 small nuclear RNA (U3b1) gene, 5' flanking region.
ACCESSION   AF020533
NID         g2736108
KEYWORDS    .
SEGMENT     2 of 3
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 367)
  AUTHORS   Gao,L., Frey,M.R. and Matera,A.G.
  TITLE     Human genes encoding U3 snRNA associate with coiled bodies in
            interphase cells and are clustered on chromosome 17p11.2 in a
            complex inverted repeat structure
  JOURNAL   Nucleic Acids Res. 25 (23), 4740-4747 (1997)
  MEDLINE   98033383
REFERENCE   2  (bases 1 to 367)
  AUTHORS   Matera,A.G. and Gao,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-AUG-1997) Genetics, Case Western Reserve University,
            10900 Euclid Ave, Cleveland, OH 44106-4955, USA
```



```
FEATURES             Location/Qualifiers
     source          1. .367
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17p11.2"
                     /clone="P1 2981"
                     /note="RNU3 locus; this is a multigene cluster on 17p11.2"
     misc_feature    <1. .>367
                     /gene="U3b1"
                     /note="5' flanking region; this sequence is approximately
                     2 kb upstream of the transcription start site"
BASE COUNT      109 a      84 c      86 g      88 t
ORIGIN

  Query Match           29.5%;  Score 18;  DB 26;  Length 367;
  Best Local Similarity 100.0%;  Pred. No. 1.72e+01;
  Matches    18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    269 AAAAACAAAATAAAACAA 286
          ||||||||||||||||||
Qy     17 AAAAACAAAATAAAACAA 34


RESULT   41
LOCUS       G01658        394 bp    DNA              STS       06-DEC-1994
DEFINITION  chicken STS ADL238.
ACCESSION   G01658
NID         g595182
KEYWORDS    STS sequence; primer; sequence tagged site.
SOURCE      chicken vector=pBluescript II KS+ host=E. coli.
  ORGANISM  Gallus domesticus
            Eukaryotae; Mitochondria; Metazoa; Chordata; Vertebrata; Tetrapoda;
            Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 394)
  AUTHORS   Cheng,H.H.
  JOURNAL   Unpublished (1994)
COMMENT     Synonyms: B272
            Contact: Hans H. Cheng
            Avian Disease and Oncology Laboratory
            USDA-ARS
            3606 E. Mount Hope Rd, East Lansing, MI 48823, USA
            Tel: 5173376758
            Fax: 5173376776
            Email: hcheng@pilot.msu.edu

            Primer A: AAACCCAAACAAAAGCAGAC
            Primer B: GCTCCTCATAAGCAAAATGC
            STS size: 160
            PCR Profile:
                 Denaturation:    94 degrees C for 1 minute
```



```
                 Annealing:       49 degrees C for 1 minute
                 Polymerization:  72 degrees C for 1 minute
                 PCR Cycles:      30
                 Thermal Cycler:  MJ Research
            Protocol:
                 Template:        50-100   ng
                 Primer:            0.1 uM
                 dNTPs:             200   uM
                 Taq polymerase:      0.4 units/ul
                 Total Vol:          25   ul

            Buffer:
                 MgCl2:          1.5 mM
                 KCl:            50   mM
                 Tris-HCl:       10   mM
                 pH:              9.0
                 Triton X-100:    0.1 %.
FEATURES             Location/Qualifiers
     source          1. .394
                     /organism="Gallus domesticus"
     STS             88. .247
     primer_bind     88. .107
     primer_bind     complement(228. .247)
BASE COUNT      138 a      76 c      65 g      115 t
ORIGIN

  Query Match           29.5%;  Score 18;  DB 30;  Length 394;
  Best Local Similarity 100.0%;  Pred. No. 1.72e+01;
  Matches    18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     64 AAAAACAAAATAAAACAA 81
          ||||||||||||||||||
Qy     17 AAAAACAAAATAAAACAA 34


RESULT   42
LOCUS       HSGLNRS13     435 bp    DNA              PRI       11-MAR-1993
DEFINITION  H.sapiens gene for glutaminyl-tRNA synthetase, exon 13.
ACCESSION   X72408
NID         g287779
KEYWORDS    glutaminyl-tRNA synthetase.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 435)
  AUTHORS   Kaiser,E., Eberhard,D. and Knippers,R.
  TITLE     Exons encoding the highly conserved part of human glutaminyl-tRNA
            synthetase
  JOURNAL   J. Mol. Evol. 34 (1), 45-53 (1992)
  MEDLINE   92211721
```



```
REFERENCE   2  (bases 1 to 435)
  AUTHORS   Kunze,N., Bittler,E., Fett,R., Schray,B., Hameister,H.,
            Wiedorn,K.H. and Knippers,R.
  TITLE     The human QARS locus: assignment of the human gene for
            glutaminyl-tRNA synthetase to chromosome 1q32-42
  JOURNAL   Hum. Genet. 85 (5), 527-530 (1990)
  MEDLINE   91033787
REFERENCE   3  (bases 1 to 435)
  AUTHORS   Kaiser,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-FEB-1993) E. Kaiser, AG Prof. Dr. R. Knippers,
            Institut f. Mol. Genetik, Univ. Konstanz, Fakultaet fuer Biologie,
            Universitaetsstr. 10 D-7750 Konstanz 1, FRG
FEATURES             Location/Qualifiers
     source          1. .435
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="HeLa"
                     /chromosome="1"
                     /map="1q32-42"
     intron          <1. .197
                     /number=12
     exon            198. .394
                     /number=13
     intron          395. .>435
                     /number=13
BASE COUNT      168 a      67 c      62 g      138 t
ORIGIN

  Query Match           29.5%;  Score 18;  DB 25;  Length 435;
  Best Local Similarity 100.0%;  Pred. No. 1.72e+01;
  Matches    18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     83 ACAAAATAAAACAATAAA 100
          ||||||||||||||||||
Qy     21 ACAAAATAAAACAATAAA 38


RESULT   43
LOCUS       RATANTREL     441 bp    mRNA             ROD       09-MAY-1995
DEFINITION  Rat carcinoembryonic antigen-related protein mRNA.
ACCESSION   L00686
NID         g202926
KEYWORDS    antigen-related protein; carcinoembryonic antigen-related protein.
SOURCE      Rattus norvegicus fetus placenta cDNA to mRNA.
  ORGANISM  Rattus norvegicus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;
            Murinae; Rattus.
REFERENCE   1  (bases 1 to 441)
  AUTHORS   Rebstock,S., Lucas,K., Thompson,J.A. and Zimmermann,W.
```

```
  TITLE     cDNA and gene analyses imply a novel structure for a rat
            carcinoembryonic antigen-related protein
  JOURNAL   J. Biol. Chem. 265 (14), 7872-7879 (1990)
  MEDLINE   90243655
REFERENCE   2  (bases 1 to 441)
  AUTHORS   Rebstock,S., Lucas,K., Weiss,M., Thompson,J. and Zimmermann,W.
  TITLE     Spatiotemporal expression of pregnancy-specific glycoprotein gene
            rnCGM1 in rat placenta
  JOURNAL   Dev. Dyn. 198 (3), 171-181 (1993)
  MEDLINE   94184012
FEATURES             Location/Qualifiers
     source          1. .441
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /cell_type="trophoblast"
                     /dev_stage="fetus"
                     /tissue_type="placenta"
     3'UTR           1. .441
                     /partial
                     /standard_name="rnCGM1"
                     /note="putative"
     polyA_signal    423. .428
                     /note="putative"
BASE COUNT      116 a    104 c     91 g    130 t
ORIGIN

  Query Match           29.5%;  Score 18;  DB 28;  Length 441;
  Best Local Similarity 100.0%;  Pred. No. 1.72e+01;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    415 AAAAACAAAATAAAACAA 432
          ||||||||||||||||||
Qy     17 AAAAACAAAATAAAACAA 34


RESULT   44
LOCUS       G41237        491 bp    DNA             STS       21-AUG-1998
DEFINITION  B689016/SP6 Human Chromosome 12 Homo sapiens STS genomic clone
            RPCI11-689016 SP6, sequence tagged site.
ACCESSION   G41237
NID         g3449151
KEYWORDS    STS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 491)
  AUTHORS   Montgomery,K.T., Renault,B., Yoon,S.J., Marondel,I., Merscher,S.,
            Kneitz,S., Yang,R., Lau,S., Lee,E., Chang,J., Miller,A., Issar,A.,
            Shim,C., Krauter,K.S. and Kucherlapati,R.
  TITLE     A Sequence Ready Map of Human Chromosome 12
```



```
  JOURNAL   Unpublished (1998)
COMMENT
            Contact: Raju Kucherlapati
            Department of Molecular Genetics
            Albert Einstein College of Medicine (AECOM)
            1300 Morris Park Ave., Bronx, NY 10461, USA
            Tel: 718-430-2824
            Fax: 718-430-8778
            Email: Kucherla@aecom.yu.edu
            Primer A: CTGGTCCTTACTTTGTACCTTGC
            Primer B: TTTATGAACACCTTCTGACCTCC
            STS size: 158
            Protocol:
              template          30-100ng
              each primer        50ng
              dNTP               200mM
              TAQ Gold           0.6U
              MgCl2              2.5mM
              BSA                50mg/ml
              Final Volume       15ml
            Preheat:          10 min 94 deg C
            Denaturation:     20 sec 94 deg C
            Annealing:        20 sec 56 deg C
            Polymerization:   20 sec 72 deg C
            PCR Cycles:       36

            Buffer:
                      KCl:              50 mM
                      Tris-HCl:         10 mM
                      pH:               8.3

            BAC 689016 from RPCI-11 human BAC library
            (http://bacpac.med.buffalo.edu/). Maps between AFM318zh1 and
            AFM234wb6.
FEATURES             Location/Qualifiers
     source          1. .491
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RPCI11-689016"
                     /clone_lib="Human Chromosome 12"
     STS             324. .481
     primer_bind     324. .346
     primer_bind     complement(459. .481)
BASE COUNT      169 a     59 c     100 g    162 t      1 others
ORIGIN

  Query Match           29.5%;  Score 18;  DB 30;  Length 491;
  Best Local Similarity 100.0%;  Pred. No. 1.72e+01;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    142 TTGTTTTATTTTGTTTTT 159
```



```
          ||||||||||||||||||
Cp     34 TTGTTTTATTTTGTTTTT 17


RESULT   45
LOCUS       AB005216     1548 bp    mRNA            PRI       14-NOV-1997
DEFINITION  Homo sapiens mRNA for Nck, Ash and phospholipase C gamma-binding
            protein NAP4, partial cds.
ACCESSION   AB005216
NID         g2443366
KEYWORDS    Nck, Ash and phospholipase C gamma-binding protein NAP4.
SOURCE      Homo sapiens fetus brain cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Matuoka,K., Miki,H., Takahashi,K. and Takenawa,T.
  TITLE     A novel ligand for an SH3 domain of the adaptor protein nck bears
            an SH2 domain and nuclear signaling motifs
  JOURNAL   Biochem. Biophys. Res. Commun. 239 (2), 488-492 (1997)
  MEDLINE   98008866
REFERENCE   2  (bases 1 to 1548)
  AUTHORS   Matuoka,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUN-1997) to the DDBJ/EMBL/GenBank databases. Koozi
            Matuoka, Tokyo Metropolitan Institute of Gerontology, Biosignal
            Research; Sakae-cho 35-2, Itabashi-ku, Tokyo 173, Japan
            (E-mail:kmatuoka@mailhost.net, Tel:+81-3-3964-3241,
            Fax:+81-3-3579-4776)
FEATURES             Location/Qualifiers
     source          1. .1548
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /dev_stage="fetus"
                     /tissue_type="brain"
     CDS             <1. .1459
                     /codon_start=2
                     /product="Nck, Ash and phospholipase C gamma-binding
                     protein NAP4"
                     /db_xref="PID:g2443367"
                     /translation="ERGLETNSCSEEELSSPGRGGGGGGRLLLQPPGPELPPVPPPLQ
                     DLVPLGRLSRGEQQQQQQQQPPPPPPPPGPLRPLAGPSRRGSFKIRLSRLFRTKSCNG
                     GSGGGDGTGKRPSGELAASAASLTDMGGSAGRELDAGRKPKLTRTQSAFSPVSFSPLF
                     TGETVSLVDVDISQRGLTSPHPPTPPPPPRRSLSLLDDISGTLPTSVLVAPMGSSLQS
                     FPLPPPPPPHAPDAFPRIAPIRAAESLHSQPPQHLQCPLYRPDSSSFAASLRELEKCG
                     WYWGPMNWEDAEMKLKGKPDGSFLVRDSSDPRIILSLSFRSQGITHHTRMEHYRGTFS
                     LWCHPKFEDRCQSVVEFIKRAIMHSKNGKFLYFLRSRVPGLPPPPVQLLVPVSRPSNV
                     KSLQHLCRFRIRQLVRIDHIPDLPLPKFLISYIRKFTYYDPQEEVYLSLKEAQLISKQ
                     KQEVEPSTWRGAPRWSPPRAFGCQAPALKNQIKLP"
BASE COUNT      336 a    475 c    410 g    327 t
```



```
  Query Match           29.5%;  Score 18;  DB 26;  Length 1548;
  Best Local Similarity 100.0%;  Pred. No. 1.72e+01;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    640 GTCTTCCTTGCAGTCTTT 657
          ||||||||||||||||||
Cp     53 GTCTTCCTTGCAGTCTTT 36


RESULT   46
LOCUS       HSU42389     1628 bp    mRNA            PRI       11-SEP-1996
DEFINITION  human neuropeptide y/peptide YY receptor type 2 mRNA, complete cds.
ACCESSION   U42389
NID         g1314329
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1628)
  AUTHORS   Yan,H., Yang,J., Marasco,J., Yamaguchi,K., Brenner,S., Collins,F.
            and Karbon,W.
  TITLE     Cloning and functional expression of cDNAs encoding human and rat
            pancreatic polypeptide receptors
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 93 (10), 4661-4665 (1996)
  MEDLINE   96209788
REFERENCE   2  (bases 1 to 1628)
  AUTHORS   Yan,H., Yang,J., Marasco,J., Yamaguchi,K., Brenner,S., Collins,F.
            and Karbon,W.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-DEC-1995) Hai Yan, Neuroscience, Amgen Inc., 1840
            DeHavilland Drive, Thousand Oaks, CA 91362, USA
FEATURES             Location/Qualifiers
     source          1. .1628
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             432. .1577
                     /codon_start=1
                     /product="neuropeptide y/peptide YY receptor type 2"
                     /db_xref="PID:g1314330"
                     /translation="MGPIGAEADENQTVEEMKVEQYGPQTTPRGELVPDPEPELIDST
                     KLIEVQVVLILAYCSIILLGVIGNSLVIHVVIKFKSMRTVTNFFIANLAVADLLVNTL
                     CLPFTLTYTLMGEWKMGPVLCHLVPYAQGLAVQVSTITLTVIALDRHRCIVYHLESKI
                     SKRISFLIIGLGWRISALLASPLAIFREYSLIEIIPDFEIVPCTEKWPAEEKSIYGTV
                     YSLSSLLILYVLPLGIISFSYTRIWSKLKNHVSPGAANDHYHQRRQKTTKMLVCVVVV
                     FAVSWLPLHAFQLAVDIDSQVLDLKEYKLIFTVFHIIAMCSTFANPLLYGWMNSNYRK
                     AFLSAFRCEQRLDAIHSEVSVTFKAKKNLEVRKNSGPNDSFTEATNV"
BASE COUNT      403 a    369 c    369 g    487 t
ORIGIN
```

```
  Query Match           29.5%;  Score 18;  DB 25;  Length 1628;
  Best Local Similarity 100.0%;  Pred. No. 1.72e+01;
  Matches    18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    358 TTGTTTTATTTTGTTTTT 375
          ||||||||||||||||||
Cp     34 TTGTTTTATTTTGTTTTT 17


RESULT   47
LOCUS       DROORFB      2012 bp    DNA             INV       24-FEB-1994
DEFINITION  D.virilis homeobox protein (antennapedia) gene, 3' end.
ACCESSION   M95826
NID         g158022
KEYWORDS    antennapedia protein; homeobox protein.
SOURCE      Drosophila virilis DNA.
  ORGANISM  Drosophila virilis
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Arthropoda;
            Tracheata; Insecta; Pterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 2012)
  AUTHORS   Hooper,J.E., Perez-Alonso,M., Bermingham,J.R., Prout,M.,
            Rocklein,B.A., Wagenbach,M., Edstrom,J.-E., de Frutos,R. and
            Scott,M.P.
  TITLE     Comparative studies of Drosophila antennapedia genes
  JOURNAL   Unpublished (1992)
FEATURES             Location/Qualifiers
     source          1. .2012
                     /organism="Drosophila virilis"
                     /db_xref="taxon:7244"
                     /germline
     gene            1. .2012
                     /note="Antennapedia"
                     /gene="Dvir\Antp"
                     /allele=""
                     /db_xref="FlyBase:FBgn0013066"
     intron          <1. .243
                     /gene="Dvir\Antp"
                     /db_xref="FlyBase:FBgn0013066"
     exon            244. .2012
                     /gene="Dvir\Antp"
                     /db_xref="FlyBase:FBgn0013066"
     CDS             <244. .494
                     /gene="Dvir\Antp"
                     /note="putative"
                     /codon_start=3
                     /function="DNA binding"
                     /product="homeobox protein"
                     /db_xref="PID:g158023"
                     /translation="RKRGRQTYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQ
                     IKIWFQNRRMKWKKENKTKGEPGSGGEGDEITPPNSPQ"
```



```
     3'UTR           492. .2012
                     /gene="Dvir\Antp"
                     /db_xref="FlyBase:FBgn0013066"
BASE COUNT      783 a    336 c    341 g    552 t
ORIGIN

  Query Match           29.5%;  Score 18;  DB 18;  Length 2012;
  Best Local Similarity 100.0%;  Pred. No. 1.72e+01;
  Matches    18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1826 AAAAACAAAATAAAACAA 1843
          ||||||||||||||||||
Qy     17 AAAAACAAAATAAAACAA 34


RESULT   48
LOCUS       AF056933     2180 bp    DNA             BCT       07-OCT-1998
DEFINITION  Bacillus thuringiensis strain C18 crystal protein toxin gene,
            complete cds.
ACCESSION   AF056933
NID         g3702806
KEYWORDS    .
SOURCE      Bacillus thuringiensis.
  ORGANISM  Bacillus thuringiensis
            Eubacteria; Firmicutes; Bacillus/Clostridium group; Bacillaceae;
            Bacillus.
REFERENCE   1  (bases 1 to 2180)
  AUTHORS   Osman,Y.A., Madkour,M.A. and Bulla,L.A. Jr.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAR-1998) Molecular Biology, University of Wyoming,
            P.O.Box 3944, Laramie, WY 82071, USA
FEATURES             Location/Qualifiers
     source          1. .2180
                     /organism="Bacillus thuringiensis"
                     /strain="C18"
                     /db_xref="taxon:1428"
                     /plasmid="large plasmid"
                     /note="country: Egypt"
     CDS             1. .2160
                     /codon_start=1
                     /transl_table=11
                     /product="crystal protein toxin"
                     /db_xref="PID:g3702807"
                     /translation="MKLKNPDKHQTLSSNAKVDKIATDSLKNETDIELKNNNNEDYLR
                     MSEHESIDPFVSASTIQTGIQIAGKILGTLGVPFPGQIASLYSFILGELWPKGKSQWE
                     IFMEHVEAIINRKISTYARNKALTDLKGLGDALAVYHESLESWVGNRNNTRARSVVKN
                     QYIALELMFVQKLPSFAVSGEEVPLLPIYAQAANLHLLLLRDASIFEKNGGLSASEIS
                     TFYNRQVERTRDYSYHCVKWNNTGLNNLRATNGQSWVRYNQFRKDIELMVLDLVRVFP
                     SYDTLVYPIKTTSQLTREVYTDAIGTVDPNQALRSTTWYNNNAPSFSAIEAAVIRSPH
                     LLDFLEKVTIYSLLSRWSNTQYNNMWGGHRLESRPIGGALNTSTQGSTNTSINPVTLQ
                     FTSRDFYRTESWAGLNLFLTQPVNGVPRVDFHWKFPTLPIASDNFYYLGYAGVGTQLQ
```



```
                     DSENELPPETTGQPNYESYSHRLSHIGLISGSHVKALVYSWTHRSADRTNTIEPNSIT
                     QIPLVKAFNLSSGAAVVRGPGFTGGHILRRTKSGTFGHIRVNINPPFAQRYRVRMSYA
                     STTDLQFHTSINGKAINQGNFSATMNRGEDLDYKTFRTVGFTTPFSFSDVQSTFTIGA
                     WNFSSGNEVYIGRIEFVPVEVTYEAEYDFEKAQEKVTALFTSTNPRGLKTDVKDYHID
                     QVSNLVESLSDELYLDEKRELFEIVKYAKQIHIERNM"
BASE COUNT      736 a    378 c    436 g    630 t
ORIGIN

  Query Match           29.5%;  Score 18;  DB 16;  Length 2180;
  Best Local Similarity 100.0%;  Pred. No. 1.72e+01;
  Matches    18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    614 TTTTTGAAAAGAATGGGG 631
          ||||||||||||||||||
Cp     21 TTTTTGAAAAGAATGGGG 4


RESULT   49
LOCUS       HSU17077     2280 bp    mRNA            PRI       30-SEP-1995
DEFINITION  Human BENE mRNA, partial cds.
ACCESSION   U17077
NID         g1000711
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2280)
  AUTHORS   Lautner-Rieske,A., Thiebe,R. and Zachau,H.G.
  TITLE     Searching for non-V kappa transcripts from the human immunoglobulin
            kappa locus
  JOURNAL   Gene 159 (2), 199-202 (1995)
  MEDLINE   95347597
REFERENCE   2  (bases 1 to 2280)
  AUTHORS   Lautner-Rieske,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-NOV-1994) Andrea Lautner-Rieske, Institut fuer
            Physiologische Chemie, Schillerstrasse 44, Muenchen, D-80336,
            Germany
FEATURES             Location/Qualifiers
     source          1. .2280
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="pc7-1"
                     /cell_type="PC-3"
     gene            1. .447
                     /gene="BENE"
     CDS             <1. .447
                     /gene="BENE"
                     /note="similar to human T-lymphocyte maturation factor,
                     Swiss-Prot Accession Number P21145"
```



```
                     /codon_start=1
                     /product="unknown"
                     /db_xref="PID:g1000712"
                     /translation="PPATSYAPSDVPSGVALFLTIPFAFFLPELIFGFLVWTMVAATH
                     IVYPLLQGWVMYVSLTSFLISLMFLLSYLPGFYKRPESWRVLDSLYHGTTGILYMSAA
                     VLQVHATIVSEKLLDPRIYYINSAASPFAPIATLLYILHAFSIYYH"
     repeat_region   849. .1159
                     /rpt_family="Alu"
BASE COUNT      521 a    616 c    531 g    612 t
ORIGIN

  Query Match           29.5%;  Score 18;  DB 25;  Length 2280;
  Best Local Similarity 100.0%;  Pred. No. 1.72e+01;
  Matches    18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    857 TTGTTTTATTTTGTTTTT 874
          ||||||||||||||||||
Cp     34 TTGTTTTATTTTGTTTTT 17


RESULT   50
LOCUS       D88027       3192 bp    DNA             INV       13-JUN-1997
DEFINITION  Dictyostelium discoideum gene for GAPA, complete cds.
ACCESSION   D88027
NID         g2196561
KEYWORDS    GAPA.
SOURCE      Dictyostelium discoideum (strain:AX2) vegetative DNA.
  ORGANISM  Dictyostelium discoideum
            Eukaryotae; mitochondrial eukaryotes; Dictyosteliida;
            Dictyostelium.
REFERENCE   1  (bases 1 to 3192)
  AUTHORS   Adachi,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-SEP-1996) to the DDBJ/EMBL/GenBank databases.
            Hiroyuki Adachi, University of Tokyo, Department of Life Sciences,
            Graduate School of Arts & Sciences; 3-8-1, Komaba, Meguro-ku, Tokyo
            153, Japan (E-mail:f37771@m-unix.cc.u-tokyo.ac.jp,
            Tel:81-3-5454-6769, Fax:81-3-5454-6769)
REFERENCE   2  (sites)
  AUTHORS   Adachi,H., Hasebe,T., Yoshinaga,K., Ohta,T. and Sutoh,K.
  TITLE     Isolation of Dictyostelium discoideum cytokinesis mutants by
            restriction enzyme-mediated integration of the blasticidin S
            resistance marker
  JOURNAL   Biochem. Biophys. Res. Commun. 205 (3), 1808-1814 (1994)
  MEDLINE   95110328
  REMARK    Erratum:[Biochem Biophys Res Commun 1995 Mar 28;208(3):1181]
REFERENCE   3  (sites)
  AUTHORS   Adachi,H., Takahashi,Y., Hasebe,T., Shirouzu,M., Yokoyama,S. and
            Sutoh,K.
  TITLE     Dictyostelium IQGAP-related protein specifically involved in the
            completion of cytokinesis
```

```
    JOURNAL   J. Cell Biol. 137 (4), 891-898 (1997)
    MEDLINE   97296496
  FEATURES             Location/Qualifiers
       source          1. .3192
                       /organism="Dictyostelium discoideum"
                       /strain="AX2"
                       /db_xref="taxon:44689"
                       /dev_stage="vegetative"
       CDS             join(237. .797,899. .1165,1253. .1837,1954. .3123)
                       /codon_start=1
                       /product="GAPA"
                       /db_xref="PID:d1021268"
                       /db_xref="PID:g2196562"
                       /translation="MEGLEIEDEDVILLDEDDDSSSSSTVNNSSSNIKNNGNTNNNIG
                       NDDSNKVTLMTSLREKGWGSGMLVDKEKNQYGTIKSYKDDKSSPWFEERQVIATLYKA
                       RVLLHEMIYTKMNQERLLSGNLCVGEIQSLLNTQKEDVETDWIAEIQELKRNMVAEIR
                       RNHLLERDVNKLDKRIALLIKHRSNIKDLLLEQNKGKKDKKKGDDKAEYITLDQKQL
                       ESYQNLFYLLQTEPHYLAKIVTLIQADQMEDFLDTVFLTLPGDDFSPREEFLILSLFR
                       LAIGQEMSRIKSAGDLLAVESGVVPKMITTTRREQGHEFLEQIIAPILENNVVNAPDL
                       NLELNAVQVYQNMISEQEIQTGAKSTLNRGLAEDQIIQLKEVQSILEPRVEKCIQICE
                       RFFTGIIQSLNRLPYGIRWICKQIQSIAQKNFDSKPDEIAKVIGYFVYYRFINLAIVT
                       PDAFEILDKELSITSRKNLVNIAKVLQNLFTLKTFQNQGSERWMQPLNKWILSKTSIV
                       RQYLEDLIQVTDPSEYLRVDKYNELTLKLNPVVVISLGEISQTHRLLIANLAALKMKE
                       KEDPLELILKALPAPLEVVDENDREIQLTLINRFKENIEKEISISASLLAETKELVIS
                       VLRSIPIQQKQQQQQHDDEKRDDLISILQNAIKHGKETNNPQLSSNAEKIINNLKKME
                       AEGSIQSENNQYEGFIKVIALEVINRQEIREQQRKERMRLTIALRDLRKHQSYLNDQI
                       QHYTSYLKDVLLHYGPKDKKKSTKPMKISFKELTKKGVIVESDIPKLSHGSTSFYISS
                       DAPGIFDIEARIGVASVGTLSLSLDDLLDKSSAGIPYLKLENIVLDVRMTLHLLNRHF
                       LKNI"
  BASE COUNT     1289 a    369 c    430 g   1104 t
  ORIGIN

    Query Match           29.5%;  Score 18;  DB 18;  Length 3192;
    Best Local Similarity 100.0%;  Pred. No. 1.72e+01;
    Matches    18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

  Db    175 AAAAACAAAATAAAACAA 192
            ||||||||||||||||||
  Qy     17 AAAAACAAAATAAAACAA 34


  Search completed: Tue Dec 22 21:14:29 1998
  Job time : 418 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Tue Dec 22 22:25:04 1998; MasPar time 446.56 Seconds
1259.202 Million cell updates/sec

Tabular output not generated.

Title: >US-08-530-112A-4937
Description: (1-314) from US08530112A.seq
Perfect Score: 298
N.A. Sequence: 1 GATCCTCTCTCCCAGGGAAT.....CCTNCNTGTGGCCCCTNTTN 314
Comp: CTAGGAGAGAGGGTCCCTTA.....GGANGNACACCGGGGANAAN

Scoring table: TABLE jmetric
Gap 60

Nmatch STD : Dbase 0; Query 0

Searched: 2275026 seqs, 895388244 bases x 2

Post-processing: Minimum Match 0%
Listing first 1000 summaries

Database: embl-est55
1:em_est1 2:em_gss1 3:em_gss2 4:em_gss3
Database: genbank-est107
5:gb_est1 6:gb_est10 7:gb_est11 8:gb_est12 9:gb_est13
10:gb_est14 11:gb_est15 12:gb_est16 13:gb_est17
14:gb_est18 15:gb_est19 16:gb_est2 17:gb_est20
18:gb_est21 19:gb_est3 20:gb_est4 21:gb_est5 22:gb_est6
23:gb_est7 24:gb_est8 25:gb_est9 26:gb_gss1 27:gb_gss2
28:gb_gss3 29:gb_gss4

Statistics: Mean 8.347; Variance 1.175; scale 7.103

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
	1	298	100.0	313	6	C00217	HUMGS0005879, Human Ge	0.00e+00
	2	171	57.4	457	6	AA432289	zw69a11.r1 Soares test	0.00e+00
	3	156	52.3	542	22	N76293	yv38f01.r1 Soares feta	0.00e+00
	4	143	48.0	289	21	W24101	zb46f06.r1 Soares feta	0.00e+00
	5	80	26.8	277	22	AA090434	y0386.seq.F Fetal hear	3.90e-159
c	6	43	14.4	509	8	AA503247	ne45h02.s1 NCI_CGAP_Co	1.15e-60
c	7	43	14.4	511	14	AA934377	op27g12.s1 Soares_NFL_	1.15e-60
c	8	21	7.0	619	26	FR0006119	F.rubripes GSS sequenc	8.09e-11
	9	20	6.7	325	14	H31765	EST106132 Rat PC-12 ce	5.91e-09
	10	19	6.4	622	28	AQ157283	nbxb0009023f CUGI Rice	3.74e-07
	11	18	6.0	117	19	HSU50541	Human MATC-1 mRNA sequ	2.03e-05
	12	18	6.0	226	7	AA509406	vg18f08.r1 Soares mous	2.03e-05
	13	18	6.0	265	5	T09208	EST07101 Homo sapiens	2.03e-05
c	14	18	6.0	265	13	AA644289	af63f04.s1 Soares NhHM	2.03e-05
	15	18	6.0	303	11	AA716959	vu69f05.r1 Stratagene	2.03e-05
c	16	18	6.0	303	13	AA599839	ag33e05.s1 Jia bone ma	2.03e-05
	17	18	6.0	309	19	HSC27P061	H. sapiens partial cDN	2.03e-05
c	18	18	6.0	312	10	W11728	mb17e04.r1 Soares mous	2.03e-05
	19	18	6.0	340	22	AA105531	mo57g06.r1 Life Tech m	2.03e-05
	20	18	6.0	349	24	AA286324	va28c05.r1 GuayWoodfor	2.03e-05
	21	18	6.0	363	22	AA102860	mo10e12.r1 Life Tech m	2.03e-05
	22	18	6.0	365	15	AA921733	om43e01.s1 Soares_NFL_	2.03e-05
	23	18	6.0	366	23	AA152541	mr85c07.r1 Stratagene	2.03e-05
	24	18	6.0	368	12	AA760296	vv71b06.r1 Stratagene	2.03e-05
	25	18	6.0	373	10	AA656955	vr25d02.r1 Barstead mo	2.03e-05
	26	18	6.0	375	15	AA998836	UI-R-C0-ii-g-01-0-UI.s	2.03e-05
	27	18	6.0	390	16	M62106	EST00165 Homo sapiens	2.03e-05
	28	18	6.0	394	13	AA838853	vw78a10.r1 Stratagene	2.03e-05
c	29	18	6.0	395	13	AA599463	ag26d01.s1 Jia bone ma	2.03e-05
c	30	18	6.0	400	18	AI139353	qc25e04.x1 Soares_preg	2.03e-05
	31	18	6.0	403	12	AA755779	vv35c01.r1 Stratagene	2.03e-05
	32	18	6.0	411	5	R11980	yf51b04.r1 Homo sapien	2.03e-05
c	33	18	6.0	416	17	AI123210	oo17c01.x1 Soares_NSF_	2.03e-05
	34	18	6.0	421	7	MUS68E12	Mouse embryonal carcin	2.03e-05
	35	18	6.0	426	5	T80918	yd25g05.r1 Homo sapien	2.03e-05
	36	18	6.0	432	23	AA065589	ml71d07.r1 Stratagene	2.03e-05
c	37	18	6.0	434	20	W51994	zc37g04.s1 Soares sene	2.03e-05
c	38	18	6.0	435	17	AI092100	qa57h06.s1 Soares_NhHM	2.03e-05
	39	18	6.0	439	20	W70917	me23h04.r1 Soares mous	2.03e-05
	40	18	6.0	445	6	AA115038	zk89f08.r1 Soares preg	2.03e-05
	41	18	6.0	448	27	AQ089629	HS_3008_A1_D01_T7 CIT	2.03e-05
c	42	18	6.0	458	6	AA114966	zk89f08.s1 Soares preg	2.03e-05

c	43	18	6.0	487	11	AA669796	ag36b12.s1 Jia bone ma	2.03e-05
	44	18	6.0	499	13	AA845481	ai89d04.s1 Soares para	2.03e-05
	45	18	6.0	501	14	AA848340	EST191100 Normalized r	2.03e-05
c	46	18	6.0	506	28	AQ180720	HS_3220_A2_B04_T7 CIT	2.03e-05
	47	18	6.0	507	17	AI116452	ue98d02.y1 Sugano mous	2.03e-05
	48	18	6.0	530	21	W52834	zc55e06.r1 Soares sem	2.03e-05
c	49	18	6.0	540	21	W58425	zd25e08.r1 Soares feta	2.03e-05
c	50	18	6.0	565	9	AA523420	ng30b04.s1 NCI_CGAP_Co	2.03e-05
	51	18	6.0	570	13	AA871871	vq42a12.r1 Barstead bo	2.03e-05
	52	18	6.0	594	21	N95418	zb66e11.s1 Soares feta	2.03e-05
	53	18	6.0	775	18	AU036076	Mus musculus cDNA, clo	2.03e-05
	54	17	5.7	148	7	HUML13456	Human fetal-lung cDNA	9.25e-04
	55	17	5.7	199	20	HHEA18W	H. sapiens partial cDN	9.25e-04
	56	17	5.7	202	13	AA486364	ab13d06.r1 Stratagene	9.25e-04
c	57	17	5.7	208	13	AA488072	ab13d06.s1 Stratagene	9.25e-04
c	58	17	5.7	215	20	W62145	md86g07.r1 Soares mous	9.25e-04
c	59	17	5.7	280	8	AA553568	nk81g06.s1 NCI_CGAP_Sc	9.25e-04
c	60	17	5.7	282	14	H31250	EST105113 Rat PC-12 ce	9.25e-04
c	61	17	5.7	312	16	R44560	yg29f01.s1 Homo sapien	9.25e-04
c	62	17	5.7	315	5	M75736	CEL01A10S1 Caenorhabdi	9.25e-04
	63	17	5.7	329	12	AA390336	LD03279.5prime LD Dros	9.25e-04
	64	17	5.7	351	16	H02317	yj40e04.s1 Homo sapien	9.25e-04
	65	17	5.7	358	5	R13937	yf62h05.r1 Homo sapien	9.25e-04
c	66	17	5.7	360	9	C61714	C.elegans cDNA clone y	9.25e-04
c	67	17	5.7	365	5	T01421	wEST02142 Caenorhabdit	9.25e-04
	68	17	5.7	370	5	R54688	yj74g11.r1 Homo sapien	9.25e-04
	69	17	5.7	427	6	AA388819	vb25d06.r1 Soares mous	9.25e-04
c	70	17	5.7	428	23	AA139864	mq95b04.r1 Stratagene	9.25e-04
	71	17	5.7	435	20	W00954	za56f06.r1 Soares feta	9.25e-04
c	72	17	5.7	435	6	AA383308	EST96667 Testis I Homo	9.25e-04
c	73	17	5.7	437	23	AA138584	mr08h09.r1 Soares mous	9.25e-04
c	74	17	5.7	454	15	AA998553	UI-R-C0-ic-h-12-0-UI.s	9.25e-04
c	75	17	5.7	473	20	W59030	md62g10.r1 Soares mous	9.25e-04
c	76	17	5.7	478	18	AI178657	EST222338 Normalized r	9.25e-04
	77	17	5.7	500	26	B19456	pEMS730-H9 Flow Sorted	9.25e-04
c	78	17	5.7	502	10	AA624319	vm99g11.r1 Knowles Sol	9.25e-04
	79	17	5.7	512	13	AA845605	ak04h07.s1 Soares para	9.25e-04
	80	17	5.7	520	12	AA264743	LD08416.5prime LD Dros	9.25e-04
c	81	17	5.7	528	29	AQ142574	HS_2222_B1_G05_MR CIT	9.25e-04
c	82	17	5.7	530	23	AA174893	ms77h06.r1 Soares mous	9.25e-04
	83	17	5.7	532	22	AA161679	MBAFCG0F07T3 Brugia ma	9.25e-04
c	84	17	5.7	549	12	AA391814	LD10717.5prime LD Dros	9.25e-04
c	85	17	5.7	563	13	AA858821	UI-R-A0-bn-h-09-0-UI.s	9.25e-04
	86	17	5.7	621	12	AA820473	LD24220.5prime LD Dros	9.25e-04
	87	17	5.7	623	27	AQ026080	l(2)k14523 Drosophila	9.25e-04
c	88	17	5.7	660	18	AI176848	EST220451 Normalized r	9.25e-04
	89	17	5.7	873	29	AQ253295	HS_2046_B1_E03_MR CIT	9.25e-04
c	90	17	5.7	903	10	HSAF001543	Human clone kappa_200	9.25e-04
c	91	17	5.7	905	15	AI009283	EST203734 Normalized r	9.25e-04
c	92	16	5.4	76	20	W54653	md02a09.r1 Soares mous	3.47e-02
	93	16	5.4	116	27	AQ056907	RPCI11-51H3.TJ RPCI11	3.47e-02

	94	16	5.4	175	7	AA443376	zw85d07.r1 Soares tota	3.47e-02
	95	16	5.4	194	16	T24532	EST107 Homo sapiens cD	3.47e-02
	96	16	5.4	200	6	AA436802	zv71e12.r1 Soares tota	3.47e-02
	97	16	5.4	208	5	T37542	EST102694 Saccharomyce	3.47e-02
	98	16	5.4	211	19	N56323	JJ65311F Homo sapiens c	3.47e-02
	99	16	5.4	219	6	AA054509	zk83c05.s1 Soares preg	3.47e-02
c	100	16	5.4	223	13	AA665350	nt89e06.s1 NCI_CGAP_Pr	3.47e-02
	101	16	5.4	237	10	AA621703	af48h05.s1 Soares tota	3.47e-02
	102	16	5.4	250	7	AA470251	vd88e06.r1 Soares mous	3.47e-02
c	103	16	5.4	250	24	AA320556	EST22926 Adipose tissu	3.47e-02
c	104	16	5.4	250	18	AI155972	ud84g01.r1 Soares mous	3.47e-02
c	105	16	5.4	255	13	AA822432	vw37f11.r1 Soares mous	3.47e-02
c	106	16	5.4	258	22	T93397	lmEST0160 LmLV39cDNA L	3.47e-02
c	107	16	5.4	274	14	AA985493	oq83b09.s1 NCI_CGAP_Kl	3.47e-02
c	108	16	5.4	275	15	AA972563	op96h03.s1 NCI_CGAP_Lu	3.47e-02
	109	16	5.4	276	13	AA520583	ae60g02.s1 Stratagene	3.47e-02
	110	16	5.4	280	19	N46068	yy35g08.r1 Homo sapien	3.47e-02
	111	16	5.4	281	23	AA250549	mz60h04.r1 Soares mous	3.47e-02
c	112	16	5.4	281	28	AQ105084	HS_2170_B1_D11_MR CIT	3.47e-02
	113	16	5.4	282	5	R10900	yf38e07.s1 Homo sapien	3.47e-02
c	114	16	5.4	291	23	AA139866	mq95c03.r1 Stratagene	3.47e-02
c	115	16	5.4	296	17	AI018623	ou47g03.x1 NCI_CGAP_Br	3.47e-02
	116	16	5.4	300	6	C07777	C.elegans cDNA clone y	3.47e-02
c	117	16	5.4	300	13	AA856104	vw83a03.r1 Stratagene	3.47e-02
	118	16	5.4	302	15	AA909049	om12b02.s1 Soares_NFL_	3.47e-02
c	119	16	5.4	302	23	AA247673	cah0646.seq.F Human fe	3.47e-02
	120	16	5.4	307	7	AA518690	vi28g05.r1 Barstead mo	3.47e-02
	121	16	5.4	311	23	AA111164	mp58h10.r1 Soares 2NbM	3.47e-02
c	122	16	5.4	312	24	HSPD02440	H.sapiens EST sequence	3.47e-02
	123	16	5.4	314	24	AA303280	EST16047 Aorta endothe	3.47e-02
	124	16	5.4	316	7	AA447053	zw86b07.r1 Soares tota	3.47e-02
	125	16	5.4	318	20	W00934	za53e09.r1 Soares feta	3.47e-02
	126	16	5.4	322	7	CELK071E9R	C.elegans cDNA clone y	3.47e-02
c	127	16	5.4	323	11	AA697436	HL02466.5prime HL Dros	3.47e-02
	128	16	5.4	324	7	RICS0725A	Rice cDNA, partial seq	3.47e-02
	129	16	5.4	328	5	R27522	yh47f07.r1 Homo sapien	3.47e-02
c	130	16	5.4	328	11	AA081453	cn06h04.r1 Stratagene	3.47e-02
c	131	16	5.4	328	28	AQ109656	CIT-HSP-2374N19.TR CIT	3.47e-02
	132	16	5.4	330	8	AA465357	aa23c06.s1 NCI_CGAP_GC	3.47e-02
c	133	16	5.4	332	27	AQ083516	RPCI11-56K4.TJ RPCI11	3.47e-02
	134	16	5.4	336	13	AA869719	vq15a05.r1 Barstead st	3.47e-02
	135	16	5.4	337	17	AI084969	ow86f09.s1 Soares_feta	3.47e-02
	136	16	5.4	338	19	HSC2IG011	H. sapiens partial cDN	3.47e-02
	137	16	5.4	341	28	AQ095504	HS_3012_B1_D03_MR CIT	3.47e-02
c	138	16	5.4	343	25	AA375324	EST87595 HSC172 cells	3.47e-02
c	139	16	5.4	347	21	AA084672	zn20f05.s1 Stratagene	3.47e-02
	140	16	5.4	350	7	AA461293	zx63d10.r1 Soares tota	3.47e-02
	141	16	5.4	353	8	AA491498	ab20h08.r1 Stratagene	3.47e-02
	142	16	5.4	354	23	AA198103	mv13c07.r1 GuayWoodfor	3.47e-02
c	143	16	5.4	355	26	FR0009507	F.rubripes GSS sequenc	3.47e-02
c	144	16	5.4	358	5	T72232	yc68c05.r1 Homo sapien	3.47e-02

	145	16	5.4	360	9	C61018	C.elegans cDNA clone y 3.47e-02
c	146	16	5.4	360	9	C42498	C.elegans cDNA clone y 3.47e-02
c	147	16	5.4	362	19	H87032	yw15g12.r1 Homo sapien 3.47e-02
	148	16	5.4	363	27	AQ006852	CIT-HSP-2293J11.TF CIT 3.47e-02
	149	16	5.4	364	22	AA161699	MBAFCE8F03T3 Brugia ma 3.47e-02
c	150	16	5.4	365	11	AA709314	z160b11.s1 Soares preg 3.47e-02
c	151	16	5.4	369	27	AQ078803	CIT-HSP-2354P9.TR CIT- 3.47e-02
c	152	16	5.4	369	22	AA021400	ze66d02.r1 Soares reti 3.47e-02
	153	16	5.4	370	14	AA685833	EST108290 Rat PC-12 ce 3.47e-02
	154	16	5.4	371	25	AA349738	EST56604 Infant brain 3.47e-02
c	155	16	5.4	376	25	AA351885	EST59766 Infant brain 3.47e-02
	156	16	5.4	378	28	AQ099380	HS_3042_B2_H10_MF CIT 3.47e-02
	157	16	5.4	379	27	AQ004660	CIT-HSP-2133I18.TF CIT 3.47e-02
c	158	16	5.4	382	28	AQ131202	HS_3029_B1_H06_MF CIT 3.47e-02
	159	16	5.4	384	24	AA309184	EST179991 Frontal cort 3.47e-02
	160	16	5.4	392	20	W54503	md01c01.r1 Soares mous 3.47e-02
	161	16	5.4	394	18	AI144431	qb77h11.x1 Soares_feta 3.47e-02
	162	16	5.4	396	13	AA858044	ob13h09.s1 NCI_CGAP_Ki 3.47e-02
	163	16	5.4	397	16	N70998	yi50e08.r1 Homo sapien 3.47e-02
c	164	16	5.4	397	27	AQ028323	CIT-HSP-2325J20.TR CIT 3.47e-02
	165	16	5.4	401	10	AA425507	zw46g09.r1 Soares tota 3.47e-02
	166	16	5.4	403	11	AA668310	ab79d03.s1 Stratagene 3.47e-02
c	167	16	5.4	403	26	FR0009955	P.rubripes GSS sequenc 3.47e-02
c	168	16	5.4	404	28	AQ219666	HS_3243_B2_C09_MR CIT 3.47e-02
	169	16	5.4	408	16	R87730	yq33b02.s1 Homo sapien 3.47e-02
	170	16	5.4	409	7	AA450005	zx33e08.s1 Soares tota 3.47e-02
	171	16	5.4	411	10	AA644505	af73b08.r1 Soares NhHM 3.47e-02
c	172	16	5.4	418	26	AQ019718	CIT-HSP-2309G15.TR CIT 3.47e-02
c	173	16	5.4	420	27	AQ033231	HS_2226_A2_D08_MF CIT 3.47e-02
c	174	16	5.4	420	18	AI148056	qg61f09.r1 Soares_test 3.47e-02
	175	16	5.4	421	14	AA836772	of37d01.s1 NCI_CGAP_Ki 3.47e-02
	176	16	5.4	421	10	AA622912	np33b03.s1 NCI_CGAP_Lu 3.47e-02
	177	16	5.4	421	9	AA600569	vm75d09.r1 Knowles Sol 3.47e-02
	178	16	5.4	423	15	AA962528	or80g03.s1 NCI_CGAP_Lu 3.47e-02
c	179	16	5.4	423	17	AI027087	ow45b07.s1 Soares_para 3.47e-02
	180	16	5.4	424	18	AI183608	qe27f11.x1 Soares_feta 3.47e-02
	181	16	5.4	427	17	AI090547	qa64f04.x1 Soares_feta 3.47e-02
	182	16	5.4	427	6	AA127232	z117f05.s1 Soares preg 3.47e-02
	183	16	5.4	428	16	H52118	yq32h02.s1 Homo sapien 3.47e-02
	184	16	5.4	428	19	N26170	yx94g08.s1 Homo sapien 3.47e-02
	185	16	5.4	428	10	W12410	ma68d09.r1 Soares mous 3.47e-02
	186	16	5.4	429	29	AQ244730	HS_2056_B1_E07_T7 CIT 3.47e-02
	187	16	5.4	431	6	AA150611	z143h02.r1 Soares preg 3.47e-02
	188	16	5.4	431	18	AI076969	oy25c01.s1 Soares_sene 3.47e-02
	189	16	5.4	434	28	AQ122764	HS_3088_A2_C09_MF CIT 3.47e-02
c	190	16	5.4	434	5	R11953	yf54e02.r1 Homo sapien 3.47e-02
	191	16	5.4	437	13	AA863612	vx05f10.r1 Soares 2NbM 3.47e-02
c	192	16	5.4	438	21	W89642	mf75c05.r1 Soares mous 3.47e-02
	193	16	5.4	439	26	B63090	CIT978SK-103I13-2.TV C 3.47e-02
c	194	16	5.4	441	28	AQ205062	HS_3226_B2_E10_MR CIT 3.47e-02
c	195	16	5.4	445	12	AA757207	ah56e05.s1 Soares test 3.47e-02

c	196	16	5.4	445	16	R55717	yg92h02.s1 Homo sapien 3.47e-02
c	197	16	5.4	446	16	H24446	ym53g08.r1 Homo sapien 3.47e-02
	198	16	5.4	448	16	H29569	ym58f01.s1 Homo sapien 3.47e-02
	199	16	5.4	449	18	AI146699	qb32c11.x1 Soares_preg 3.47e-02
	200	16	5.4	455	21	W32699	zc06b11.s1 Soares para 3.47e-02
	201	16	5.4	455	9	AA554891	n115f10.s1 NCI_CGAP_Br 3.47e-02
	202	16	5.4	456	20	W16767	zb16a06.r1 Soares feta 3.47e-02
	203	16	5.4	458	28	AQ191028	HS_2212_B1_C03_MR CIT 3.47e-02
c	204	16	5.4	459	22	AA132893	zo22a02.r1 Stratagene 3.47e-02
c	205	16	5.4	461	15	AA936437	on65d03.s1 Soares_NFL_ 3.47e-02
c	206	16	5.4	461	9	AA534344	nf75g11.s1 NCI_CGAP_Co 3.47e-02
c	207	16	5.4	462	8	AA506221	nh44g02.s1 NCI_CGAP_Pr 3.47e-02
	208	16	5.4	463	5	R08933	yf25g07.s1 Homo sapien 3.47e-02
c	209	16	5.4	463	17	AI111910	UI-R-A1-ef-f-08-0-UI.s 3.47e-02
c	210	16	5.4	464	10	AA611923	vo86a02.r1 Barstead mo 3.47e-02
	211	16	5.4	465	13	AA626788	ad09g05.s1 Soares NbHF 3.47e-02
	212	16	5.4	466	17	AI056123	ow33d03.s1 Soares_para 3.47e-02
	213	16	5.4	466	19	N52801	ys18g05.s1 Homo sapien 3.47e-02
	214	16	5.4	467	16	H44705	yp24c06.r1 Homo sapien 3.47e-02
	215	16	5.4	467	6	AA035510	zk26a06.s1 Soares preg 3.47e-02
	216	16	5.4	468	14	AA910864	ok85c07.s1 NCI_CGAP_Ki 3.47e-02
	217	16	5.4	470	8	AA495897	zw04h05.r1 Soares NhHM 3.47e-02
	218	16	5.4	471	19	N59817	yx76e09.s1 Homo sapien 3.47e-02
	219	16	5.4	471	18	AI149126	qc71e02.x1 Soares_plac 3.47e-02
	220	16	5.4	472	17	AA961750	or60d06.s1 NCI_CGAP_GC 3.47e-02
c	221	16	5.4	472	5	R54653	yj74c08.r1 Homo sapien 3.47e-02
	222	16	5.4	473	13	AA830066	oc44a05.s1 NCI_CGAP_GC 3.47e-02
c	223	16	5.4	474	11	AA698947	HL06013.5prime HL Dros 3.47e-02
	224	16	5.4	475	11	AA660865	aa80e12.s1 NCI_CGAP_GC 3.47e-02
	225	16	5.4	475	16	H18008	ym42e12.r1 Homo sapien 3.47e-02
	226	16	5.4	475	6	AA156836	z118e01.r1 Soares preg 3.47e-02
	227	16	5.4	475	27	B74192	CIT-HSP-2033A1.TP CIT- 3.47e-02
	228	16	5.4	477	18	AI095827	qb26a06.x1 Soares_preg 3.47e-02
	229	16	5.4	477	5	T85568	yd82d11.s1 Homo sapien 3.47e-02
	230	16	5.4	480	6	AA431436	zw79g04.s1 Soares test 3.47e-02
c	231	16	5.4	482	15	AI011021	EST205472 Normalized r 3.47e-02
	232	16	5.4	483	6	AA039948	zk45f06.s1 Soares preg 3.47e-02
	233	16	5.4	484	13	AA732456	nz96b11.s1 NCI_CGAP_GC 3.47e-02
	234	16	5.4	485	29	AQ138175	HS_3074_B1_C01_MF CIT 3.47e-02
	235	16	5.4	488	27	AQ084686	HS_2252_B1_D02_MR CIT 3.47e-02
c	236	16	5.4	489	11	AA682166	vu13f03.r1 Barstead mo 3.47e-02
	237	16	5.4	493	29	AQ228344	HS_2014_B2_A09_MR CIT 3.47e-02
c	238	16	5.4	493	22	AA016208	ze33f05.s1 Soares reti 3.47e-02
c	239	16	5.4	495	12	AA729623	nx60e09.s1 NCI_CGAP_Al 3.47e-02
c	240	16	5.4	496	28	AQ222337	HS_2007_A2_F02_MR CIT 3.47e-02
c	241	16	5.4	496	21	W45611	zc26d07.r1 Soares sene 3.47e-02
	242	16	5.4	498	16	R98168	yq66a10.r1 Homo sapien 3.47e-02
c	243	16	5.4	499	24	AA275564	vc33a12.r1 Barstead MP 3.47e-02
	244	16	5.4	501	21	AA048859	mj50c04.s1 Soares mous 3.47e-02
	245	16	5.4	501	27	B92562	CIT-HSP-2164M9.TR CIT- 3.47e-02
c	246	16	5.4	504	27	B94957	CIT-HSP-2170H21.TR CIT 3.47e-02

c	247	16	5.4	504	8	AA465448	aa23e04.r1 NCI_CGAP_GC 3.47e-02
c	248	16	5.4	505	12	T45885	9148 Lambda-PRL2 Arabi 3.47e-02
	249	16	5.4	506	12	AA744785	ny26e11.s1 NCI_CGAP_GC 3.47e-02
	250	16	5.4	507	18	AI089437	qa06a09.x1 NCI_CGAP_Br 3.47e-02
	251	16	5.4	509	20	W00905	za52e09.r1 Soares feta 3.47e-02
c	252	16	5.4	516	12	AA745673	ob05f08.s1 NCI_CGAP_Ki 3.47e-02
c	253	16	5.4	521	18	AI145084	UI-R-BT0-pt-h-03-0-UI. 3.47e-02
	254	16	5.4	523	21	W67790	zd38b08.s1 Soares feta 3.47e-02
c	255	16	5.4	524	13	AA196926	zq09c08.r1 Stratagene 3.47e-02
	256	16	5.4	530	28	AQ171846	HS_3180_B1_D02_T7 CIT 3.47e-02
c	257	16	5.4	530	6	AA431838	zw79g04.r1 Soares test 3.47e-02
	258	16	5.4	536	23	AA118867	mp59a01.r1 Soares 2NbM 3.47e-02
	259	16	5.4	537	11	R90323	16678 Lambda-PRL2 Arab 3.47e-02
c	260	16	5.4	538	11	AA669908	ag42g04.s1 Jia bone ma 3.47e-02
	261	16	5.4	542	15	C81013	Mus musculus 3.5-dpc b 3.47e-02
c	262	16	5.4	548	8	AA481612	aa35g07.r1 NCI_CGAP_GC 3.47e-02
c	263	16	5.4	558	23	AA228811	mx92e04.r1 Soares mous 3.47e-02
c	264	16	5.4	562	9	AA577682	nn22f12.s1 NCI_CGAP_Co 3.47e-02
	265	16	5.4	567	26	B52751	CIT-HSP-2006M1.TF CIT- 3.47e-02
c	266	16	5.4	575	23	AA124471	mp96g02.r1 Soares 2NbM 3.47e-02
c	267	16	5.4	576	26	B08477	T2F4-T7 TAMU Arabidops 3.47e-02
c	268	16	5.4	577	26	FR0025973	F.rubripes GSS sequenc 3.47e-02
	269	16	5.4	581	19	N66434	yx40f06.s1 Homo sapien 3.47e-02
c	270	16	5.4	581	21	W67848	zd38b08.r1 Soares feta 3.47e-02
c	271	16	5.4	585	6	AA410913	zv39d10.r1 Soares ovar 3.47e-02
c	272	16	5.4	586	6	AA001234	zh82g08.r1 Soares feta 3.47e-02
	273	16	5.4	587	26	FR0027771	F.rubripes GSS sequenc 3.47e-02
	274	16	5.4	587	17	AU022138	Mus musculus unfertili 3.47e-02
c	275	16	5.4	590	26	B15398	344P23.TP CIT978SKA1 H 3.47e-02
	276	16	5.4	591	21	W52990	zc02e09.r1 Soares para 3.47e-02
	277	16	5.4	597	15	AI052686	oz26b01.x1 Soares_tota 3.47e-02
	278	16	5.4	602	15	C80075	Mus musculus 3.5-dpc b 3.47e-02
c	279	16	5.4	602	9	AA573957	nk09g05.s1 NCI_CGAP_Co 3.47e-02
	280	16	5.4	619	26	FR0002702	F.rubripes GSS sequenc 3.47e-02
c	281	16	5.4	620	10	AA619737	vl60c08.s1 Knowles Sol 3.47e-02
c	282	16	5.4	629	14	AA921288	vz37d03.r1 Soares 2NbM 3.47e-02
	283	16	5.4	632	27	AQ003047	CIT-HSP-2036A2.TF CIT 3.47e-02
c	284	16	5.4	638	26	CI18F1	Ciona intestinalis gen 3.47e-02
c	285	16	5.4	661	11	AA692056	vt06h03.r1 Barstead mo 3.47e-02
	286	16	5.4	663	29	AQ250680	T9P13-T7 TAMU Arabidop 3.47e-02
	287	16	5.4	665	28	AQ195204	RPCI11-61D15.TK RPCI11 3.47e-02
	288	16	5.4	693	11	AA697657	HL02943.3prime HL Dros 3.47e-02
c	289	16	5.4	696	10	AA635222	SWOvL3CA1055SK Onchoce 3.47e-02
	290	16	5.4	714	26	AG004147	Homo sapiens genomic D 3.47e-02
	291	16	5.4	714	27	AG010273	Homo sapiens genomic D 3.47e-02
	292	16	5.4	714	26	AG004034	Homo sapiens genomic D 3.47e-02
c	293	16	5.4	780	20	W22511	69D9 Human retina cDNA 3.47e-02
c	294	16	5.4	808	20	W22150	62F3 Human retina cDNA 3.47e-02
c	295	16	5.4	909	26	B09733	F23H11-Sp6 IGF Arabido 3.47e-02
	296	15	5.0	85	5	T60949	yc46a04.r1 Homo sapien 1.05e+00
	297	15	5.0	88	20	ENU44376	Emericella nidulans cl 1.05e+00

	298	15	5.0	122	7	AA511786	vg16f02.r1 Soares mous 1.05e+00
c	299	15	5.0	129	27	AQ009603	CIT-HSP-2285J14.TFB CI 1.05e+00
c	300	15	5.0	138	16	H46582	yp19a03.s1 Homo sapien 1.05e+00
	301	15	5.0	140	7	AA519046	vi26e07.r1 Barstead mo 1.05e+00
	302	15	5.0	144	23	AA064498	ml49h02.r1 Stratagene 1.05e+00
c	303	15	5.0	148	20	HSAAABIUS	H. sapiens putatively 1.05e+00
c	304	15	5.0	151	20	HSAAABIET	H. sapiens putatively 1.05e+00
	305	15	5.0	171	5	T20045	B325F Homo sapiens cDN 1.05e+00
c	306	15	5.0	173	12	AA758724	ah75f06.s1 Soares test 1.05e+00
	307	15	5.0	177	21	AA052484	mc21h02.r1 Soares mous 1.05e+00
c	308	15	5.0	177	24	AA318832	EST21189 Adrenal gland 1.05e+00
	309	15	5.0	185	22	AA089647	chn0227.seq.F Fetal he 1.05e+00
	310	15	5.0	190	22	N59702	yv55h04.r1 Soares feta 1.05e+00
c	311	15	5.0	191	16	T35071	EST79785 Homo sapiens 1.05e+00
c	312	15	5.0	196	19	H57414	yr11g01.r1 Homo sapien 1.05e+00
c	313	15	5.0	197	20	HSCZRH011	H. sapiens partial cDN 1.05e+00
c	314	15	5.0	204	20	HSAAABJWW	H. sapiens putatively 1.05e+00
c	315	15	5.0	207	8	AA235391	zt31g05.s1 Soares ovar 1.05e+00
c	316	15	5.0	208	21	W67968	zd39f07.s1 Soares feta 1.05e+00
c	317	15	5.0	210	6	AA127846	zl12b04.s1 Soares preg 1.05e+00
c	318	15	5.0	210	9	AA522588	ni70e05.s1 NCI_CGAP_Pr 1.05e+00
c	319	15	5.0	212	8	AA235871	zs41e07.s1 Soares NhHM 1.05e+00
	320	15	5.0	213	20	W41867	mc64c03.r1 Soares mous 1.05e+00
c	321	15	5.0	213	18	AI174441	an46c10.s1 Gessler Wil 1.05e+00
c	322	15	5.0	214	21	AA034489	zk22e09.s1 Soares preg 1.05e+00
c	323	15	5.0	215	16	H44408	yo74d02.s1 Homo sapien 1.05e+00
c	324	15	5.0	218	21	AA046274	zf12c03.s1 Soares feta 1.05e+00
c	325	15	5.0	218	16	H02023	yj38e08.s1 Homo sapien 1.05e+00
	326	15	5.0	220	14	AA895316	vy33a11.r1 Stratagene 1.05e+00
c	327	15	5.0	221	20	ATTS5510	A. thaliana transcribe 1.05e+00
	328	15	5.0	222	7	CELK101D6R	C.elegans cDNA clone y 1.05e+00
	329	15	5.0	223	14	AA903162	ok53d08.s1 NCI_CGAP_Le 1.05e+00
	330	15	5.0	224	21	W67832	zd39f07.r1 Soares feta 1.05e+00
c	331	15	5.0	224	10	AA429226	zw19e05.r1 Soares ovar 1.05e+00
	332	15	5.0	225	8	AA280148	zt09h03.r1 NCI_CGAP_GC 1.05e+00
	333	15	5.0	229	16	R89397	yq01b03.r1 Homo sapien 1.05e+00
c	334	15	5.0	230	22	N98208	0126c3 csapPFDd2.1, De 1.05e+00
c	335	15	5.0	230	20	HSC34G082	H. sapiens partial cDN 1.05e+00
	336	15	5.0	231	25	AA360585	EST69929 T-cell lympho 1.05e+00
	337	15	5.0	232	25	AA370427	EST82281 Prostate glan 1.05e+00
c	338	15	5.0	235	16	H16301	ym23b03.s1 Homo sapien 1.05e+00
	339	15	5.0	242	20	W41824	mc65e06.r1 Soares mous 1.05e+00
	340	15	5.0	243	21	AA031036	mi47h07.r1 Soares mous 1.05e+00
c	341	15	5.0	243	7	HUM214G07B	Human aorta cDNA 5'-en 1.05e+00
	342	15	5.0	244	27	AQ070486	HS_3035_B1_G05_MR CIT 1.05e+00
	343	15	5.0	248	20	W78621	me88f06.r1 Soares mous 1.05e+00
	344	15	5.0	249	10	AA663528	ab80d10.s1 Stratagene 1.05e+00
c	345	15	5.0	251	5	T95095	ye44a08.s1 Homo sapien 1.05e+00
	346	15	5.0	252	7	HUMGSO0664	Human HL60 3'directed 1.05e+00
	347	15	5.0	255	11	AA680848	LmFrAm0409 Leishmania 1.05e+00
	348	15	5.0	261	19	H98412	SMNHADA016064T3 Schist 1.05e+00

c	349	15	5.0	261	20	N73409	EST59e105 BATM2 Homo s 1.05e+00
	350	15	5.0	262	5	R34946	yh86c05.r1 Homo sapien 1.05e+00
c	351	15	5.0	262	6	AA401151	zu52d03.s1 Soares ovar 1.05e+00
	352	15	5.0	263	8	AA478329	zu45a12.r1 Soares ovar 1.05e+00
	353	15	5.0	263	14	AA686536	EST110599 Rat PC-12 ce 1.05e+00
c	354	15	5.0	264	22	N54796	yv31b08.s1 Soares feta 1.05e+00
	355	15	5.0	265	8	AA526209	ni59c03.s1 NCI_CGAP_Ov 1.05e+00
	356	15	5.0	267	7	HUM319C10B	Human aorta cDNA 5' en 1.05e+00
c	357	15	5.0	268	5	T10631	hbc653 Homo sapiens cD 1.05e+00
	358	15	5.0	271	25	AA378222	EST90962 Synovial sarc 1.05e+00
c	359	15	5.0	275	5	R25104	yh36a11.s1 Homo sapien 1.05e+00
	360	15	5.0	276	5	R01076	ye84c07.s1 Homo sapien 1.05e+00
c	361	15	5.0	276	26	B23015	F24K5TR IGF Arabidopsi 1.05e+00
c	362	15	5.0	279	14	AA982978	ua24g04.r1 Soares mous 1.05e+00
c	363	15	5.0	282	25	AA345138	EST51202 Gall bladder 1.05e+00
	364	15	5.0	285	19	HSC20D041	H. sapiens partial cDN 1.05e+00
	365	15	5.0	286	25	AA348447	EST55088 Hippocampus I 1.05e+00
	366	15	5.0	286	5	R27082	yh52b06.r1 Homo sapien 1.05e+00
	367	15	5.0	288	27	B99013	CIT-HSP-2287E15.TF CIT 1.05e+00
c	368	15	5.0	288	7	AA454691	zx76c10.s1 Soares ovar 1.05e+00
	369	15	5.0	290	8	AA262925	zr71a01.s1 Soares NhHM 1.05e+00
c	370	15	5.0	290	7	MUSA1B07	Mouse embryonal carcin 1.05e+00
c	371	15	5.0	292	13	AA599410	ag23b05.s1 Jia bone ma 1.05e+00
	372	15	5.0	294	10	SSC17G05	Sus scrofa mRNA; expre 1.05e+00
c	373	15	5.0	294	6	AA099821	zk87c06.s1 Soares preg 1.05e+00
c	374	15	5.0	295	17	AI072123	UI-R-C2-nf-h-03-0-UI.s 1.05e+00
	375	15	5.0	296	22	N59716	yv56b04.r1 Soares feta 1.05e+00
c	376	15	5.0	296	16	T34025	EST61420 Homo sapiens 1.05e+00
	377	15	5.0	298	25	AA373407	EST85439 HSC172 cells 1.05e+00
	378	15	5.0	298	5	T12239	A448F Homo sapiens cDN 1.05e+00
	379	15	5.0	300	10	AA623323	vi19g02.r1 Barstead mo 1.05e+00
c	380	15	5.0	300	9	C53972	C.elegans cDNA clone y 1.05e+00
c	381	15	5.0	300	9	C31509	C.elegans cDNA clone y 1.05e+00
	382	15	5.0	302	15	AI044975	UI-R-C1-ke-a-11-0-UI.s 1.05e+00
	383	15	5.0	304	19	H63374	yz56f09.s1 Homo sapien 1.05e+00
	384	15	5.0	307	5	T87087	yd88b12.s1 Homo sapien 1.05e+00
	385	15	5.0	308	21	AA049823	mj38h04.r1 Soares mous 1.05e+00
c	386	15	5.0	309	9	C75163	Homo sapiens mRNA expr 1.05e+00
	387	15	5.0	311	9	AA528580	nf02a03.s1 NCI_CGAP_Ki 1.05e+00
c	388	15	5.0	314	13	AAB36694	Tc-EST-145 (5' read) T 1.05e+00
c	389	15	5.0	314	25	AA376729	EST89210 Small intesti 1.05e+00
c	390	15	5.0	314	6	AA401737	zt66d06.s1 Soares test 1.05e+00
c	391	15	5.0	315	13	AA609507	af14d07.s1 Soares test 1.05e+00
c	392	15	5.0	316	14	H34497	EST111488 Rat PC-12 ce 1.05e+00
	393	15	5.0	319	8	AA235984	zs41g05.s1 Soares NhHM 1.05e+00
c	394	15	5.0	319	29	AQ236546	HS_2033_B2_E05_7 CIT 1.05e+00
c	395	15	5.0	319	18	AI086677	oz57d04.x1 Soares_sene 1.05e+00
c	396	15	5.0	321	8	AA482858	nf49c07.s1 NCI_CGAP_Al 1.05e+00
c	397	15	5.0	322	8	AA494517	ne36h09.s1 NCI_CGAP_Co 1.05e+00
	398	15	5.0	323	23	AA209824	mu37e04.r1 Soares 2NbM 1.05e+00
	399	15	5.0	324	15	AI029854	UI-R-C0-jq-c-05-0-UI.s 1.05e+00

	400	15	5.0	324	8	AA393388	zt75d10.r1 Soares test 1.05e+00
	401	15	5.0	325	16	R63888	yi21g09.r1 Homo sapien 1.05e+00
c	402	15	5.0	325	25	AA363248	EST73077 Ovary I Homo 1.05e+00
	403	15	5.0	327	5	R19362	yf46b12.s1 Homo sapien 1.05e+00
	404	15	5.0	327	7	AA461026	zx64d11.r1 Soares tota 1.05e+00
c	405	15	5.0	327	8	AA287036	zs57d11.s1 NCI_CGAP_GC 1.05e+00
c	406	15	5.0	327	5	R12749	yf58f07.r1 Homo sapien 1.05e+00
	407	15	5.0	331	13	AA826972	nr87b05.s1 NCI_CGAP_Pr 1.05e+00
	408	15	5.0	331	22	AA188757	zp77a07.s1 Stratagene 1.05e+00
c	409	15	5.0	332	22	AA169643	zo94b03.s1 Stratagene 1.05e+00
c	410	15	5.0	332	26	FR0021317	F.rubripes GSS sequenc 1.05e+00
	411	15	5.0	333	15	AA970789	on57h08.s1 Soares_NFL_ 1.05e+00
c	412	15	5.0	333	26	AQ019158	CIT-HSP-2303C15.TF CIT 1.05e+00
	413	15	5.0	334	20	W44128	mc74c01.r1 Soares mous 1.05e+00
c	414	15	5.0	334	5	R23445	yh39a01.r1 Homo sapien 1.05e+00
	415	15	5.0	335	18	AI143528	qa51h12.s1 Soares_NhHM 1.05e+00
c	416	15	5.0	335	20	W39007	zb29f03.r1 Soares para 1.05e+00
	417	15	5.0	337	11	AA705586	ab42b02.s1 Stratagene 1.05e+00
	418	15	5.0	337	8	AA195496	zr36c05.r1 Soares NhHM 1.05e+00
	419	15	5.0	337	5	R27739	yh64f12.r1 Homo sapien 1.05e+00
c	420	15	5.0	337	21	W83568	mf29d11.r1 Soares mous 1.05e+00
	421	15	5.0	338	24	AA296644	EST11178 Uterus Homo s 1.05e+00
	422	15	5.0	339	20	W75289	me53g03.r1 Soares mous 1.05e+00
c	423	15	5.0	339	5	M79753	wEST00290 Caenorhabdit 1.05e+00
	424	15	5.0	340	27	AQ069322	HS_2265_B2_H03_MR CIT 1.05e+00
	425	15	5.0	341	26	B41716	HS-1054-A2-B06-MR.abi 1.05e+00
	426	15	5.0	342	16	M79081	EST01229 Homo sapiens 1.05e+00
c	427	15	5.0	342	18	AI093981	qa33e04.s1 Soares_NhHM 1.05e+00
c	428	15	5.0	342	20	W78194	zd68c04.r1 Soares feta 1.05e+00
c	429	15	5.0	346	27	AQ088459	HS_2208_B1_D06_MF CIT 1.05e+00
	430	15	5.0	347	28	AQ103681	HS_3093_A2_H07_MF CIT 1.05e+00
	431	15	5.0	347	16	R66500	yi32e06.r1 Homo sapien 1.05e+00
	432	15	5.0	348	19	H89578	yw28d09.r1 Homo sapien 1.05e+00
	433	15	5.0	348	20	ATTS1897	A. thaliana transcribe 1.05e+00
c	434	15	5.0	349	5	R41455	yf86a05.s1 Homo sapien 1.05e+00
c	435	15	5.0	350	23	AA219210	sq15g06.r1 Stratagene 1.05e+00
c	436	15	5.0	350	17	AI050976	ow34e11.s1 Soares_para 1.05e+00
	437	15	5.0	351	5	T77929	yd16f06.r1 Homo sapien 1.05e+00
c	438	15	5.0	351	25	AA334410	EST38644 Embryo, 9 wee 1.05e+00
c	439	15	5.0	352	13	AA804359	nz37h12.s1 NCI_CGAP_GC 1.05e+00
	440	15	5.0	353	14	AA825307	oc67e08.s1 NCI_CGAP_GC 1.05e+00
c	441	15	5.0	353	23	AA176063	mt01h02.r1 Soares mous 1.05e+00
	442	15	5.0	356	5	R17807	yg09b07.r1 Homo sapien 1.05e+00
	443	15	5.0	360	9	C65072	C.elegans cDNA clone y 1.05e+00
	444	15	5.0	360	20	HSCEYH081	H. sapiens partial cDN 1.05e+00
c	445	15	5.0	360	6	CELK032GXF	C.elegans cDNA clone y 1.05e+00
c	446	15	5.0	360	8	AA292740	zt55g08.s1 Soares ovar 1.05e+00
c	447	15	5.0	360	23	AA213164	mw85e07.r1 Soares mous 1.05e+00
c	448	15	5.0	361	27	AQ001641	CIT-HSP-2282D7.TR CIT- 1.05e+00
	449	15	5.0	362	5	T49553	ya77a02.s1 Homo sapien 1.05e+00
c	450	15	5.0	363	5	R11292	yf41a09.r1 Homo sapien 1.05e+00

c	451	15	5.0	363	10	AA426124	zv52e12.s1 Soares test 1.05e+00
	452	15	5.0	364	16	R68089	yi05c07.s1 Homo sapien 1.05e+00
c	453	15	5.0	364	21	W52356	zc48b07.s1 Soares sene 1.05e+00
c	454	15	5.0	364	6	C06366	similar to none, mRNA 1.05e+00
c	455	15	5.0	364	7	AA446317	zw67b06.s1 Soares test 1.05e+00
c	456	15	5.0	365	19	H66375	yu16e08.r1 Homo sapien 1.05e+00
c	457	15	5.0	365	5	R13832	yf65a02.r1 Homo sapien 1.05e+00
	458	15	5.0	366	19	N31870	yx71e03.r1 Homo sapien 1.05e+00
	459	15	5.0	366	5	T95491	ye43e10.r1 Homo sapien 1.05e+00
c	460	15	5.0	366	12	AA817708	UI-R-A0-aa-e-03-0-UI.s 1.05e+00
c	461	15	5.0	366	5	T69607	yc44f12.s1 Homo sapien 1.05e+00
c	462	15	5.0	366	27	AQ057191	CIT-HSP-2335P13.TF CIT 1.05e+00
c	463	15	5.0	369	20	W02895	za04g04.r1 Soares mela 1.05e+00
c	464	15	5.0	370	16	M78872	EST01020 Homo sapiens 1.05e+00
	465	15	5.0	371	5	R02826	ye80g11.s1 Homo sapien 1.05e+00
	466	15	5.0	371	22	AA020543	mh62c06.r1 Soares mous 1.05e+00
c	467	15	5.0	373	24	AA311919	EST182625 Jurkat T-cel 1.05e+00
	468	15	5.0	374	5	T95195	ye44a08.r1 Homo sapien 1.05e+00
c	469	15	5.0	375	9	C66378	C.elegans cDNA clone y 1.05e+00
c	470	15	5.0	375	18	AI094645	oy61d08.s1 NCI_CGAP_Br 1.05e+00
	471	15	5.0	376	16	H47500	yp74c10.s1 Homo sapien 1.05e+00
c	472	15	5.0	376	6	AA150195	zl04b01.s1 Soares preg 1.05e+00
	473	15	5.0	377	9	C65044	C.elegans cDNA clone y 1.05e+00
c	474	15	5.0	377	15	AI029229	UI-R-C0-ip-h-05-0-UI.s 1.05e+00
	475	15	5.0	378	28	AQ106207	HS_3051_B1_A06_MF CIT 1.05e+00
	476	15	5.0	378	8	AA476409	zx02a01.s1 Soares tota 1.05e+00
c	477	15	5.0	378	8	AA508551	nh67b07.s1 NCI_CGAP_Pr 1.05e+00
	478	15	5.0	379	5	R18097	yg16f10.r1 Homo sapien 1.05e+00
	479	15	5.0	379	13	AA830659	oc52h03.s1 NCI_CGAP_GC 1.05e+00
c	480	15	5.0	379	14	AA934537	ny35a09.s1 NCI_CGAP_Pr 1.05e+00
	481	15	5.0	381	25	AA332368	EST36300 Embryo, 8 wee 1.05e+00
c	482	15	5.0	381	12	AA772406	ai44e06.s1 Soares para 1.05e+00
	483	15	5.0	382	5	T49126	yb09c02.r1 Homo sapien 1.05e+00
	484	15	5.0	382	27	B82284	RPCI11-14G16.TP RPCI11 1.05e+00
c	485	15	5.0	382	11	AA708440	zl57c01.s1 Soares preg 1.05e+00
c	486	15	5.0	382	17	AI122551	qa95h05.x1 Soares_feta 1.05e+00
c	487	15	5.0	382	14	AA899314	UI-R-E0-crh-05-0-UI.s 1.05e+00
	488	15	5.0	383	12	AA739572	337 PtIFG2 Pinus taeda 1.05e+00
	489	15	5.0	383	8	AA135259	zo24h06.r1 Stratagene 1.05e+00
	490	15	5.0	384	5	T49570	ya77c02.s1 Homo sapien 1.05e+00
	491	15	5.0	384	5	T37198	EST102278 Saccharomyce 1.05e+00
c	492	15	5.0	384	14	C91195	Dictyostelium discoide 1.05e+00
c	493	15	5.0	385	29	AF046473	Mus musculus clone OST 1.05e+00
	494	15	5.0	387	20	W59568	md72g08.r1 Soares mous 1.05e+00
c	495	15	5.0	387	17	AI025635	ow04d06.s1 Soares_para 1.05e+00
c	496	15	5.0	390	26	B30612	HS-1002-B2-H07-MR.abi 1.05e+00
c	497	15	5.0	390	16	H14180	ym62h08.r1 Homo sapien 1.05e+00
c	498	15	5.0	390	21	W36949	mb82c12.r1 Soares mous 1.05e+00
	499	15	5.0	391	27	AQ006277	CIT-HSP-2291G9.TR CIT- 1.05e+00
c	500	15	5.0	391	19	H61137	yu37b04.s1 Homo sapien 1.05e+00
c	501	15	5.0	391	17	AI026015	ow03d12.s1 Soares_para 1.05e+00

c	502	15	5.0	391	28	AQ219105	HS_2165_B2_E12_MF CIT 1.05e+00
	503	15	5.0	392	15	AA998749	UI-R-C0-ii-f-11-0-UI.s 1.05e+00
	504	15	5.0	392	27	AQ060227	CIT-HSP-2356D22.TR CIT 1.05e+00
c	505	15	5.0	392	16	R94450	yq40b09.s1 Homo sapien 1.05e+00
c	506	15	5.0	392	11	AA772167	zg88e10.s1 Soares feta 1.05e+00
c	507	15	5.0	392	8	AA468594	ne07g12.s1 NCI_CGAP_Co 1.05e+00
c	508	15	5.0	393	5	R09310	yf22h04.r1 Homo sapien 1.05e+00
	509	15	5.0	394	16	H15970	ym20d08.r1 Homo sapien 1.05e+00
	510	15	5.0	394	12	AA761496	nz27a10.s1 NCI_CGAP_GC 1.05e+00
c	511	15	5.0	394	21	AA048756	mj43a09.r1 Soares mous 1.05e+00
c	512	15	5.0	394	13	AA855576	vw66g04.r1 Stratagene 1.05e+00
	513	15	5.0	395	14	AA938390	oi99a03.s1 NCI_CGAP_Me 1.05e+00
	514	15	5.0	395	5	R31136	yh61c07.r1 Homo sapien 1.05e+00
c	515	15	5.0	396	28	AQ191660	HS_3225_B1_H12_MR CIT 1.05e+00
c	516	15	5.0	396	14	AA903344	ok45h12.s1 NCI_CGAP_Le 1.05e+00
c	517	15	5.0	396	26	B42773	HS-1056-A2-H08-MF.abi 1.05e+00
	518	15	5.0	397	21	AA030644	mi13c10.r1 Soares mous 1.05e+00
c	519	15	5.0	397	15	AA934043	on78a05.s1 Soares_NFL_ 1.05e+00
c	520	15	5.0	397	22	AA103876	mo43g04.r1 Life Tech m 1.05e+00
	521	15	5.0	398	20	W53363	md42d06.r1 Soares mous 1.05e+00
	522	15	5.0	398	5	T65328	yc74e05.r1 Homo sapien 1.05e+00
c	523	15	5.0	398	26	B44743	HS-1060-A1-A05-MF.abi 1.05e+00
	524	15	5.0	399	26	AQ020763	CIT-HSP-2304H17.TR CIT 1.05e+00
	525	15	5.0	400	28	AQ093937	HS_3025_A1_H06_7 CIT 1.05e+00
	526	15	5.0	400	6	C18864	Human placenta cDNA 5' 1.05e+00
c	527	15	5.0	400	10	AA631987	np74a01.s1 NCI_CGAP_Br 1.05e+00
	528	15	5.0	401	26	AQ010243	HS_2173_A1_MR_F07 CIT 1.05e+00
	529	15	5.0	401	11	AA692601	vt60c12.r1 Barstead mo 1.05e+00
	530	15	5.0	401	12	AA763779	vo99f12.r1 Soares mous 1.05e+00
c	531	15	5.0	402	14	AA825743	od29h11.s1 NCI_CGAP_GC 1.05e+00
c	532	15	5.0	402	19	H99148	yx14g05.s1 Homo sapien 1.05e+00
	533	15	5.0	403	12	AA773339	ab65f06.s1 Stratagene 1.05e+00
	534	15	5.0	404	14	AA981518	ua21b07.r1 Soares mous 1.05e+00
	535	15	5.0	404	28	AQ108843	HS_3185_B1_B03_MR CIT 1.05e+00
	536	15	5.0	404	28	AQ122567	HS_3083_B1_F03_MR CIT 1.05e+00
c	537	15	5.0	404	11	AA723379	zg73a11.s1 Soares feta 1.05e+00
	538	15	5.0	405	19	N74491	za54f04.s1 Homo sapien 1.05e+00
c	539	15	5.0	405	24	AA065261	c04502s Testis 5 Homo 1.05e+00
c	540	15	5.0	405	12	AA798438	vy06h03.r1 Stratagene 1.05e+00
	541	15	5.0	406	12	AA768494	oa64h10.s1 NCI_CGAP_GC 1.05e+00
	542	15	5.0	407	14	AA850445	EST193212 Normalized r 1.05e+00
c	543	15	5.0	407	26	B45864	HS-1062-B1-B04-MR.abi 1.05e+00
c	544	15	5.0	407	16	R79439	yi87e12.r1 Homo sapien 1.05e+00
	545	15	5.0	408	5	R33563	yh79g06.r1 Homo sapien 1.05e+00
	546	15	5.0	408	27	AQ064448	HS_2198_B2_H08_MR CIT 1.05e+00
c	547	15	5.0	410	10	AA635926	nr71a06.s1 NCI_CGAP_Pr 1.05e+00
c	548	15	5.0	410	21	W15362	zc17d12.s1 Soares para 1.05e+00
	549	15	5.0	411	8	AA251281	zs04a05.s1 NCI_CGAP_GC 1.05e+00
c	550	15	5.0	411	6	HSPD06081	H.sapiens EST sequence 1.05e+00
	551	15	5.0	412	19	N56872	yy55d12.s1 Homo sapien 1.05e+00
	552	15	5.0	412	15	AA998658	UI-R-C0-ig-d-02-0-UI.s 1.05e+00

c	553	15	5.0	412	14	AA980907	ua47b05.r1 Soares mous	1.05e+00
c	554	15	5.0	413	10	AA614917	vo31g06.r1 Barstead mo	1.05e+00
c	555	15	5.0	415	28	AQ220948	HS_3244_A2_C05_MR CIT	1.05e+00
c	556	15	5.0	415	12	AA805841	nu71g12.s1 NCI_CGAP_Al	1.05e+00
c	557	15	5.0	415	20	W20358	zb44g07.r1 Soares feta	1.05e+00
	558	15	5.0	416	12	AA764542	vp05d10.r1 Soares mous	1.05e+00
	559	15	5.0	416	10	AA477389	zu37g06.r1 Soares ovar	1.05e+00
c	560	15	5.0	416	24	AA286034	vc34f02.r1 Barstead MP	1.05e+00
c	561	15	5.0	416	27	AQ072287	HS_3011_A2_B09_T7 CIT	1.05e+00
	562	15	5.0	417	17	AI095760	qb30b08.x1 Soares_preg	1.05e+00
	563	15	5.0	417	18	AI151438	qc88g07.x1 Soares_preg	1.05e+00
	564	15	5.0	417	26	B30698	HS-1003-A1-D08-MF.abi	1.05e+00
c	565	15	5.0	417	28	AQ223950	HS_2259_B2_E12_T7 CIT	1.05e+00
c	566	15	5.0	417	14	AA887827	nr09d04.s1 NCI_CGAP_Co	1.05e+00
	567	15	5.0	418	15	AI045887	UI-R-C0-jq-c-05-0-UI.s	1.05e+00
c	568	15	5.0	418	13	AA885031	am41e05.s1 Soares NFL	1.05e+00
c	569	15	5.0	418	11	AA071104	sm66e10.r1 Stratagene	1.05e+00
c	570	15	5.0	418	13	AA829261	qg07a01.s1 NCI_CGAP_Lu	1.05e+00
	571	15	5.0	420	13	AA856290	vw98h12.r1 Soares 2NbM	1.05e+00
c	572	15	5.0	420	6	AA398380	zt62c07.s1 Soares test	1.05e+00
	573	15	5.0	421	6	AA058710	zk70b06.s1 Soares preg	1.05e+00
c	574	15	5.0	421	15	AA976212	oo30a11.s1 NCI_CGAP_Lu	1.05e+00
	575	15	5.0	422	10	AA629388	zu78c10.s1 Soares test	1.05e+00
c	576	15	5.0	423	22	AA181551	zp51a11.s1 Stratagene	1.05e+00
c	577	15	5.0	424	5	R21070	yg52d08.r1 Homo sapien	1.05e+00
c	578	15	5.0	424	13	AA883814	al59e08.s1 Soares NFL	1.05e+00
	579	15	5.0	425	18	AI171278	EST217233 Normalized r	1.05e+00
c	580	15	5.0	425	5	R43381	yg08a07.s1 Homo sapien	1.05e+00
c	581	15	5.0	425	28	AQ154462	HS_3006_B2_H04_MR CIT	1.05e+00
	582	15	5.0	426	22	W94541	ze04a08.s1 Soares feta	1.05e+00
	583	15	5.0	426	16	R78826	yi90c02.r1 Homo sapien	1.05e+00
c	584	15	5.0	426	14	AA964332	UI-R-C0-gu-h-07-0-UI.s	1.05e+00
	585	15	5.0	427	29	AQ243969	HS_2058_B1_E04_MR CIT	1.05e+00
	586	15	5.0	427	12	AA791136	vv91c12.r1 Stratagene	1.05e+00
	587	15	5.0	427	24	N45828	T1586 MVAT4 bloodstrea	1.05e+00
	588	15	5.0	427	14	AA900041	UI-R-E0-dh-g-04-0-UI.s	1.05e+00
c	589	15	5.0	428	21	AA052211	mb89h04.r1 Soares mous	1.05e+00
c	590	15	5.0	428	6	AA398204	zt59b04.s1 Soares test	1.05e+00
	591	15	5.0	430	5	R25240	yh41d12.r1 Homo sapien	1.05e+00
c	592	15	5.0	430	9	AA516117	ni07f05.s1 NCI_CGAP_Br	1.05e+00
c	593	15	5.0	430	12	AA713669	nv69c04.s1 NCI_CGAP_GC	1.05e+00
	594	15	5.0	431	28	AQ221914	HS_2189_A1_C04_MR CIT	1.05e+00
	595	15	5.0	431	6	AA039622	zk40f05.s1 Soares preg	1.05e+00
	596	15	5.0	431	18	AI156406	ue10f09.r1 Soares mous	1.05e+00
c	597	15	5.0	431	26	AQ020005	CIT-HSP-2312P5.TF CIT-	1.05e+00
	598	15	5.0	432	16	R66608	yi32e10.r1 Homo sapien	1.05e+00
	599	15	5.0	432	21	W89875	mf77d05.r1 Soares mous	1.05e+00
c	600	15	5.0	432	12	AA769826	ah81b12.s1 Soares test	1.05e+00
c	601	15	5.0	432	8	AA281406	zs95b10.r1 NCI_CGAP_GC	1.05e+00
c	602	15	5.0	432	5	T89930	ye12c11.r1 Homo sapien	1.05e+00
c	603	15	5.0	432	27	B82835	RPCI11-17P12.TV RPCI11	1.05e+00

	604	15	5.0	433	19	H62099	yu40e04.r1 Homo sapien	1.05e+00
	605	15	5.0	433	14	AA857284	oh96d07.s1 NCI_CGAP_HN	1.05e+00
	606	15	5.0	433	17	AI076139	oy92f09.x1 Soares_feta	1.05e+00
c	607	15	5.0	433	12	T75784	10562 Lambda-PRL2 Arab	1.05e+00
c	608	15	5.0	433	8	AA465386	aa23f07.s1 NCI_CGAP_GC	1.05e+00
	609	15	5.0	435	27	AQ036727	CIT-HSP-2334H7.TF CIT-	1.05e+00
	610	15	5.0	435	6	AA148570	sl04b01.r1 Soares preg	1.05e+00
	611	15	5.0	436	5	T71817	yc62f03.r1 Homo sapien	1.05e+00
	612	15	5.0	436	12	AA757367	ah98a09.s1 Soares NFL	1.05e+00
c	613	15	5.0	436	20	N95737	yy60f06.r1 Homo sapien	1.05e+00
c	614	15	5.0	436	20	W60078	zc93b08.s1 Pancreatic	1.05e+00
	615	15	5.0	438	11	N38180	19407 Lambda-PRL2 Arab	1.05e+00
c	616	15	5.0	439	29	AQ233309	HS_2259_B1_G12_T7 CIT	1.05e+00
	617	15	5.0	440	13	AA810658	ob58e02.s1 NCI_CGAP_GC	1.05e+00
c	618	15	5.0	440	15	AI016111	ot87d07.s1 Soares_tota	1.05e+00
c	619	15	5.0	440	19	N68402	za13b07.s1 Homo sapien	1.05e+00
	620	15	5.0	441	18	AI141628	oy91f11.x1 Soares_feta	1.05e+00
c	621	15	5.0	441	18	AI050740	ox94c02.s1 Soares_sene	1.05e+00
c	622	15	5.0	441	13	AA633660	ac15h11.s1 Stratagene	1.05e+00
c	623	15	5.0	441	20	W97406	mf94f09.r1 Soares mous	1.05e+00
	624	15	5.0	442	15	AI061354	an36h10.x1 Gessler Wil	1.05e+00
	625	15	5.0	442	28	AQ204248	HS_3116_B1_F04_MR CIT	1.05e+00
	626	15	5.0	442	15	AA893159	EST196962 Normalized r	1.05e+00
c	627	15	5.0	442	27	AQ013866	RPCI11-23J9.TKBR RPCI1	1.05e+00
c	628	15	5.0	442	18	AI177995	EST221651 Normalized r	1.05e+00
c	629	15	5.0	442	15	AU007397	Schizosaccharomyces po	1.05e+00
c	630	15	5.0	442	26	B48958	RPCI11-4C22.TP RPCI11	1.05e+00
	631	15	5.0	443	14	AA849841	EST192608 Normalized r	1.05e+00
	632	15	5.0	444	28	AQ091191	HS_3009_B1_G01_T7 CIT	1.05e+00
c	633	15	5.0	444	11	AA079318	zm98c02.s1 Stratagene	1.05e+00
c	634	15	5.0	444	17	AI093040	qa77h09.x1 Soares_feta	1.05e+00
c	635	15	5.0	445	8	AA459425	aa31a05.r1 NCI_CGAP_GC	1.05e+00
c	636	15	5.0	445	17	AA829263	og07a03.s1 NCI_CGAP_Lu	1.05e+00
	637	15	5.0	446	6	AA410404	zv12f08.s1 Soares NhHM	1.05e+00
	638	15	5.0	446	20	W47956	mc87c05.r1 Soares mous	1.05e+00
c	639	15	5.0	446	12	AA731613	nw60c10.s1 NCI_CGAP_GC	1.05e+00
c	640	15	5.0	446	10	AA594595	nl95d07.s1 NCI_CGAP_Co	1.05e+00
c	641	15	5.0	446	21	AA050822	mj22d06.r1 Soares mous	1.05e+00
	642	15	5.0	447	29	AQ247068	HS_2027_B2_G11_T7 CIT	1.05e+00
	643	15	5.0	447	17	AI119753	uc20a07.r1 Soares mous	1.05e+00
c	644	15	5.0	447	18	AI141843	qa66c06.x1 Soares_feta	1.05e+00
	645	15	5.0	448	18	AI175935	EST219508 Normalized r	1.05e+00
	646	15	5.0	448	23	AA146195	mq65f09.r1 Soares 2NbM	1.05e+00
	647	15	5.0	448	16	R67877	yi28h09.r1 Homo sapien	1.05e+00
	648	15	5.0	449	29	AQ141973	HS_3157_A2_B07_MR CIT	1.05e+00
	649	15	5.0	449	12	AA796164	vp18c09.r1 Soares mous	1.05e+00
c	650	15	5.0	449	29	AQ254192	HS_2243_A1_F05_MR CIT	1.05e+00
c	651	15	5.0	449	9	AA548769	nj13f10.s1 NCI_CGAP_Pr	1.05e+00
c	652	15	5.0	450	18	AI183968	qd69e11.x1 Soares_test	1.05e+00
c	653	15	5.0	450	19	H87547	yw17b01.r1 Homo sapien	1.05e+00
c	654	15	5.0	450	28	AQ198360	RPCI11-47M9.TJ RPCI11	1.05e+00

c	655	15	5.0	450	13	AA867588	vy02d06.r1 Stratagene	1.05e+00
	656	15	5.0	451	22	AA114129	zn65f04.r1 Stratagene	1.05e+00
	657	15	5.0	451	13	AA174006	zp01a04.r1 Stratagene	1.05e+00
c	658	15	5.0	451	6	AA134822	zl20e09.s1 Soares preg	1.05e+00
c	659	15	5.0	451	8	AA262761	zs24d02.s1 NCI_CGAP_GC	1.05e+00
c	660	15	5.0	451	5	R07911	yf16c01.r1 Homo sapien	1.05e+00
c	661	15	5.0	451	9	AA565032	nj11g09.s1 NCI_CGAP_Pr	1.05e+00
	662	15	5.0	452	22	AA156426	zl51b07.s1 Soares preg	1.05e+00
c	663	15	5.0	452	18	AI086704	oz61b10.x1 Soares_sene	1.05e+00
	664	15	5.0	453	29	AQ232063	HS_2021_A1_F07_T7 CIT	1.05e+00
	665	15	5.0	453	5	R44030	yg21e10.s1 Homo sapien	1.05e+00
	666	15	5.0	453	28	AQ167494	HS_2193_B1_D11_MR CIT	1.05e+00
	667	15	5.0	453	18	AI168083	oz89d01.x1 Soares_sene	1.05e+00
	668	15	5.0	453	29	AQ136211	HS_3063_A1_D10_MF CIT	1.05e+00
c	669	15	5.0	453	17	AI001884	ot39c09.s1 Soares_test	1.05e+00
c	670	15	5.0	455	28	AQ164260	HS_2207_B1_G10_MR CIT	1.05e+00
	671	15	5.0	456	21	W25135	zb69d07.r1 Soares feta	1.05e+00
	672	15	5.0	456	23	AA125330	mp80e08.r1 Soares 2NbM	1.05e+00
	673	15	5.0	456	15	AI012468	EST206919 Normalized r	1.05e+00
c	674	15	5.0	456	27	AQ004509	CIT-HSP-2292K13.TF CIT	1.05e+00
c	675	15	5.0	456	22	AA120854	zn83c10.r1 Stratagene	1.05e+00
c	676	15	5.0	457	14	AF064923	Mus musculus clone CR1	1.05e+00
c	677	15	5.0	458	11	AA071558	zm74f06.s1 Stratagene	1.05e+00
c	678	15	5.0	458	13	AA169185	zp20e07.s1 Stratagene	1.05e+00
c	679	15	5.0	459	17	AI001760	ot38h02.s1 Soares_test	1.05e+00
c	680	15	5.0	459	14	C90965	Dictyostelium discoide	1.05e+00
c	681	15	5.0	459	5	R12270	yf40e09.s1 Homo sapien	1.05e+00
	682	15	5.0	460	23	AA253670	mx84e01.r1 Soares mous	1.05e+00
	683	15	5.0	460	22	AA023737	mh78h11.r1 Soares mous	1.05e+00
	684	15	5.0	461	22	AA016847	mh43a04.r1 Soares mous	1.05e+00
c	685	15	5.0	461	21	N94522	zb78a01.s1 Soares sene	1.05e+00
c	686	15	5.0	461	18	AI141851	qa66d05.x1 Soares_feta	1.05e+00
	687	15	5.0	462	16	R94781	yq43c05.r1 Homo sapien	1.05e+00
c	688	15	5.0	462	13	AA633896	ac73a02.s1 Stratagene	1.05e+00
	689	15	5.0	463	16	H22784	ym54e03.r1 Homo sapien	1.05e+00
c	690	15	5.0	463	16	H16686	ym26d10.s1 Homo sapien	1.05e+00
c	691	15	5.0	463	22	AA185657	mt89a06.r1 Soares mous	1.05e+00
c	692	15	5.0	463	8	AA542877	ni80h01.s1 NCI_CGAP_Pr	1.05e+00
	693	15	5.0	464	19	N50723	yz21b09.s1 Homo sapien	1.05e+00
c	694	15	5.0	464	29	AQ245781	HS_2058_A2_B08_T7 CIT	1.05e+00
	695	15	5.0	465	19	N32717	yx75a09.r1 Homo sapien	1.05e+00
	696	15	5.0	466	10	AA417296	zu18c04.s1 Soares NhHM	1.05e+00
	697	15	5.0	466	7	AA462479	vg74e06.r1 Soares mous	1.05e+00
	698	15	5.0	466	10	AA421779	zu26a02.s1 Soares ovar	1.05e+00
	699	15	5.0	466	17	AI022698	ow67d03.x1 Soares_feta	1.05e+00
	700	15	5.0	466	10	AA477876	zu34d08.r1 Soares ovar	1.05e+00
c	701	15	5.0	466	6	AA127923	zl13h05.s1 Soares preg	1.05e+00
	702	15	5.0	467	15	AI008360	EST202811 Normalized r	1.05e+00
	703	15	5.0	467	21	W61095	zd30e05.r1 Soares feta	1.05e+00
c	704	15	5.0	467	18	AI139053	qc22g11.x1 Soares_preg	1.05e+00
c	705	15	5.0	467	15	AI015539	ov01f10.x1 NCI_CGAP_Ki	1.05e+00

	706	15	5.0	468	22	AA152377	zo07e05.s1 Stratagene	1.05e+00
	707	15	5.0	468	14	AA940036	vz96g05.r1 Soares 2NbM	1.05e+00
c	708	15	5.0	468	18	AI178071	EST221730 Normalized r	1.05e+00
c	709	15	5.0	469	18	AI159886	qb55g02.x1 NCI_CGAP_Br	1.05e+00
	710	15	5.0	470	20	AA002696	mg54e09.r1 Soares mous	1.05e+00
c	711	15	5.0	470	27	B60396	CIT-HSP-2007M17.TFB CI	1.05e+00
c	712	15	5.0	470	22	AA147554	zl51b07.r1 Soares preg	1.05e+00
c	713	15	5.0	470	9	AA534753	nf77e06.s1 NCI_CGAP_Co	1.05e+00
c	714	15	5.0	471	23	AA175517	ms92d06.r1 Soares mous	1.05e+00
c	715	15	5.0	471	9	AA533669	nj96f02.s1 NCI_CGAP_Pr	1.05e+00
	716	15	5.0	472	21	W83255	mf25b11.r1 Soares mous	1.05e+00
c	717	15	5.0	472	28	AQ220988	HS_3244_A2_A04_MR CIT	1.05e+00
c	718	15	5.0	472	18	AI143093	qb76a01.x1 Soares_feta	1.05e+00
	719	15	5.0	473	29	AQ229428	HS_2024_A1_F07_T7 CIT	1.05e+00
c	720	15	5.0	475	18	AI095823	qb26a02.x1 Soares_preg	1.05e+00
c	721	15	5.0	476	13	AA866276	UI-R-A0-bg-b-06-0-UI.s	1.05e+00
	722	15	5.0	477	22	AA037119	zc51a06.r1 Soares sene	1.05e+00
c	723	15	5.0	477	6	AA388575	vc95h07.r1 Barstead MP	1.05e+00
	724	15	5.0	478	28	AQ210526	HS_2212_A2_A04_MF CIT	1.05e+00
c	725	15	5.0	478	14	AA825664	od29b08.s1 NCI_CGAP_GC	1.05e+00
c	726	15	5.0	478	22	N75725	yv31c08.r1 Soares feta	1.05e+00
c	727	15	5.0	478	16	H08732	yl89d07.s1 Homo sapien	1.05e+00
	728	15	5.0	480	10	AA430752	zw52g08.r1 Soares tota	1.05e+00
	729	15	5.0	480	23	AA153587	mq57d05.r1 Soares 2NbM	1.05e+00
	730	15	5.0	481	12	AA540453	LD19947.5prime LD Dros	1.05e+00
c	731	15	5.0	481	13	C87917	Mus musculus fertilize	1.05e+00
c	732	15	5.0	482	19	N26421	yx23a04.r1 Homo sapien	1.05e+00
c	733	15	5.0	482	19	N35810	yy27f02.s1 Homo sapien	1.05e+00
	734	15	5.0	483	24	AA277380	va81h10.r1 Soares mous	1.05e+00
c	735	15	5.0	483	26	B96126	F24A2TPB IGF Arabidops	1.05e+00
c	736	15	5.0	483	23	AA097841	mm35c09.r1 Stratagene	1.05e+00
	737	15	5.0	484	28	AQ182403	HS_3058_B1_E04_MF CIT	1.05e+00
	738	15	5.0	485	20	AA003190	mg41h05.r1 Soares mous	1.05e+00
	739	15	5.0	485	21	W29751	mc07d06.r1 Soares mous	1.05e+00
c	740	15	5.0	485	10	AA419118	zv34d08.s1 Soares ovar	1.05e+00
c	741	15	5.0	485	21	AA008173	mg69g04.r1 Soares mous	1.05e+00
	742	15	5.0	486	29	AQ234713	HS_2050_B1_B12_MR CIT	1.05e+00
c	743	15	5.0	486	28	AQ177410	HS_2206_A2_C10_MR CIT	1.05e+00
c	744	15	5.0	487	28	AQ117447	HS_2191_A2_H12_MR CIT	1.05e+00
	745	15	5.0	488	10	AA622178	nq55g08.s1 NCI_CGAP_Co	1.05e+00
	746	15	5.0	488	26	B51476	CIT978SK-A-61H9.TP CIT	1.05e+00
c	747	15	5.0	488	23	AA261159	va50c05.r1 Soares mous	1.05e+00
c	748	15	5.0	488	13	AA823828	vr84c05.s1 Knowles Sol	1.05e+00
	749	15	5.0	489	23	AA209939	mu40e12.r1 Soares 2NbM	1.05e+00
	750	15	5.0	489	10	AA427107	ve82e10.r1 Soares mous	1.05e+00
	751	15	5.0	490	15	AA943257	EST198756 Normalized r	1.05e+00
	752	15	5.0	490	8	AA215537	zr97c05.r1 NCI_CGAP_GC	1.05e+00
c	753	15	5.0	490	12	AA724866	ah97d06.s1 Soares NFL	1.05e+00
c	754	15	5.0	490	12	AA781145	aj24a10.s1 Soares test	1.05e+00
	755	15	5.0	492	13	AA833413	ub58c03.r1 Soares mous	1.05e+00
	756	15	5.0	492	13	AA822748	vw35a11.r1 Soares mous	1.05e+00

c	757	15	5.0	492	20	W92501	ze44c04.r1 Soares reti	1.05e+00
	758	15	5.0	493	11	AA716166	zg63f07.s1 Soares feta	1.05e+00
	759	15	5.0	493	23	AA261556	mz62h04.r1 Soares mous	1.05e+00
	760	15	5.0	494	17	AI004303	ou56e02.x1 NCI_CGAP_Br	1.05e+00
c	761	15	5.0	494	16	H43874	yo69e09.s1 Homo sapien	1.05e+00
	762	15	5.0	495	23	AA222144	mw09f01.r1 Soares mous	1.05e+00
	763	15	5.0	495	28	AQ132430	HS_3021_A1_E11_MR CIT	1.05e+00
	764	15	5.0	495	16	H45688	yp23e02.s1 Homo sapien	1.05e+00
	765	15	5.0	495	22	AA155402	mt64h12.r1 Soares mous	1.05e+00
	766	15	5.0	495	24	AA274937	va96f04.r1 Soares mous	1.05e+00
	767	15	5.0	496	24	AA288341	mr52b03.r1 Life Tech m	1.05e+00
c	768	15	5.0	496	10	W12447	ma64g09.r1 Soares mous	1.05e+00
	769	15	5.0	497	28	AQ154693	HS_3021_A1_E11_T7 CIT	1.05e+00
c	770	15	5.0	497	9	AA534131	nj97f02.s1 NCI_CGAP_Pr	1.05e+00
	771	15	5.0	498	19	H31875	ym71f02.r1 Homo sapien	1.05e+00
	772	15	5.0	498	14	AA834012	of25b03.s1 NCI_CGAP_Ki	1.05e+00
c	773	15	5.0	498	11	AA727817	vp33c02.r1 Barstead mo	1.05e+00
c	774	15	5.0	499	27	FR0031496	Fugu rubripes GSS sequ	1.05e+00
c	775	15	5.0	499	14	AA895111	vy07b12.r1 Stratagene	1.05e+00
	776	15	5.0	500	20	W64224	me04h01.r1 Soares mous	1.05e+00
c	777	15	5.0	500	16	H27220	yl63a12.r1 Homo sapien	1.05e+00
c	778	15	5.0	501	15	AI049470	ub37d12.r1 Soares 2NbM	1.05e+00
c	779	15	5.0	502	11	AA673641	vo57f01.r1 Soares mous	1.05e+00
c	780	15	5.0	503	17	AI096847	qb40a12.x1 NCI_CGAP_Br	1.05e+00
	781	15	5.0	504	14	AA867035	ua39d10.r1 Soares mous	1.05e+00
	782	15	5.0	504	8	AA546238	vk60h05.r1 Stratagene	1.05e+00
	783	15	5.0	504	15	AI043877	UI-R-C0-jk-b-01-0-UI.s	1.05e+00
	784	15	5.0	505	23	AA118395	mp74c09.r1 Soares 2NbM	1.05e+00
	785	15	5.0	505	13	AA859260	UI-R-E0-cj-f-04-0-UI.s	1.05e+00
	786	15	5.0	506	23	AA239849	mx81c06.r1 Soares mous	1.05e+00
c	787	15	5.0	506	7	AA475293	vh20d09.r1 Soares mous	1.05e+00
c	788	15	5.0	507	26	B35548	HS-1029-A2-C01-MR.abi	1.05e+00
	789	15	5.0	508	28	AQ155210	HS_3051_A1_D06_MF CIT	1.05e+00
c	790	15	5.0	508	26	FR0014865	F.rubripes GSS sequenc	1.05e+00
	791	15	5.0	509	14	AA990324	ua56b12.r1 Soares 2NbM	1.05e+00
c	792	15	5.0	509	10	AA621353	af91b01.s1 Soares tota	1.05e+00
	793	15	5.0	510	17	AA922084	oi29g05.s1 NCI_CGAP_HN	1.05e+00
	794	15	5.0	510	8	AA553076	vl08d02.r1 Soares mous	1.05e+00
	795	15	5.0	510	23	AA216891	mu93g08.r1 Soares mous	1.05e+00
	796	15	5.0	511	26	B55922	CIT-HSP-2005I11.TR CIT	1.05e+00
	797	15	5.0	511	8	AA278953	zs81b01.s1 NCI_CGAP_GC	1.05e+00
	798	15	5.0	511	24	AA288578	vb21d08.r1 Soares mous	1.05e+00
	799	15	5.0	512	5	R20590	yg49d06.r1 Homo sapien	1.05e+00
c	800	15	5.0	512	21	AA030381	mi23a02.r1 Soares mous	1.05e+00
	801	15	5.0	513	8	AA455713	aa22e05.r1 NCI_CGAP_GC	1.05e+00
	802	15	5.0	513	22	AA171253	ms51c07.r1 Life Tech m	1.05e+00
c	803	15	5.0	513	10	AA658744	fa66f11.s1 zebrafish f	1.05e+00
	804	15	5.0	514	12	AA813952	oc06f01.s1 NCI_CGAP_GC	1.05e+00
	805	15	5.0	515	11	AA650942	31056 Lambda-PRL2 Arab	1.05e+00
c	806	15	5.0	516	17	AI023185	ow66f03.s1 Soares_sene	1.05e+00
c	807	15	5.0	516	17	AI032700	ox07f12.s1 Soares_feta	1.05e+00

	808	15	5.0	517	19	N31846	yx71a03.r1 Homo sapien	1.05e+00
c	809	15	5.0	517	26	FR0013128	F.rubripes GSS sequenc	1.05e+00
	810	15	5.0	519	19	H71679	ys01c05.r1 Homo sapien	1.05e+00
	811	15	5.0	520	8	AA546290	vk63b07.r1 Stratagene	1.05e+00
c	812	15	5.0	520	29	AQ140272	HS_3113_A2_F06_MR CIT	1.05e+00
	813	15	5.0	521	7	AA454120	zx45e05.r1 Soares test	1.05e+00
	814	15	5.0	521	13	AA805656	ns41a10.s1 NCI_CGAP_GC	1.05e+00
c	815	15	5.0	521	28	AQ167028	HS_3147_B1_G10_T7 CIT	1.05e+00
	816	15	5.0	522	12	AA739394	vv53a03.r1 Soares 2NbM	1.05e+00
	817	15	5.0	523	11	AA689037	vs02d12.r1 Barstead mo	1.05e+00
	818	15	5.0	523	13	AA839014	vw93a02.r1 Stratagene	1.05e+00
	819	15	5.0	524	5	R18299	yg01f05.r1 Homo sapien	1.05e+00
	820	15	5.0	524	23	AA120573	mp69g12.r1 Soares 2NbM	1.05e+00
c	821	15	5.0	525	24	AA286570	vb76a03.r1 Soares mous	1.05e+00
c	822	15	5.0	525	21	AA016535	mg90b05.r1 Soares mous	1.05e+00
	823	15	5.0	526	16	H40054	yl44e11.r1 Homo sapien	1.05e+00
	824	15	5.0	528	14	AA944973	EST200472 Normalized r	1.05e+00
	825	15	5.0	528	28	AQ213135	HS_3251_B1_E09_T7 CIT	1.05e+00
	826	15	5.0	531	27	B81135	CIT-HSP-2015F12.TFC CI	1.05e+00
	827	15	5.0	531	13	AA832781	vw46c09.r1 Soares mous	1.05e+00
c	828	15	5.0	533	18	AI183966	qd69e09.x1 Soares_test	1.05e+00
c	829	15	5.0	533	27	AQ043136	CIT-HSP-2327F16.TF CIT	1.05e+00
	830	15	5.0	534	28	AQ117797	HS_3000_A2_E11_MF CIT	1.05e+00
	831	15	5.0	535	26	B75029	RPCI11-12K7.TV RPCI11	1.05e+00
c	832	15	5.0	535	18	AI182003	ud75d01.x1 Sugano mous	1.05e+00
	833	15	5.0	536	15	AI037144	uc90h12.y1 Sugano mous	1.05e+00
	834	15	5.0	536	24	AA315000	EST186807 HCC cell lin	1.05e+00
	835	15	5.0	536	16	H16400	ym24h05.s1 Homo sapien	1.05e+00
c	836	15	5.0	536	28	AQ118804	HS_3016_B2_A11_T7 CIT	1.05e+00
	837	15	5.0	537	15	AI012469	EST206920 Normalized r	1.05e+00
	838	15	5.0	538	16	H12550	yj12b09.r1 Homo sapien	1.05e+00
c	839	15	5.0	540	29	AQ153255	HS_2230_B2_B06_MR CIT	1.05e+00
c	840	15	5.0	544	14	AA905636	oj95c06.s1 Soares_NFL_	1.05e+00
c	841	15	5.0	544	24	AA271926	vb75c04.r1 Soares mous	1.05e+00
	842	15	5.0	551	10	AA611531	vc85a05.r1 Barstead mo	1.05e+00
c	843	15	5.0	552	16	R91353	yp95g06.r1 Homo sapien	1.05e+00
	844	15	5.0	555	14	AA920074	vz25e05.r1 Soares 2NbM	1.05e+00
	845	15	5.0	556	17	AU024173	Mus musculus unfertili	1.05e+00
	846	15	5.0	560	29	AQ233608	HS_2045_B2_G08_MR CIT	1.05e+00
c	847	15	5.0	560	26	AQ023656	CpGR0094A Cryptosporid	1.05e+00
	848	15	5.0	561	17	AA787970	r3f10a1.r1 Aspergillus	1.05e+00
	849	15	5.0	563	17	AI119372	ue98h05.y1 Sugano mous	1.05e+00
	850	15	5.0	563	20	W25819	14d9 Human retina cDNA	1.05e+00
	851	15	5.0	564	23	AA145622	ms07d11.r1 Stratagene	1.05e+00
	852	15	5.0	567	16	R70286	yj81d03.r1 Homo sapien	1.05e+00
c	853	15	5.0	568	7	AA516555	vh96a01.r1 Barstead mo	1.05e+00
c	854	15	5.0	569	26	FR0025578	F.rubripes GSS sequenc	1.05e+00
	855	15	5.0	570	21	W55883	zc03a12.s1 Soares para	1.05e+00
	856	15	5.0	570	27	B89157	CIT-HSP-2173K13.TR CIT	1.05e+00
	857	15	5.0	570	6	AA127913	zl13h05.r1 Soares preg	1.05e+00
	858	15	5.0	574	23	AA145127	mr08d01.r1 Soares mous	1.05e+00

c	859	15	5.0	576	22	AA107554	mp09h10.r1 Life Tech m	1.05e+00
	860	15	5.0	579	18	AI151837	ui46f07.y1 Sugano mous	1.05e+00
	861	15	5.0	581	8	AA507008	ni04e09.s1 NCI_CGAP_Br	1.05e+00
	862	15	5.0	582	17	AI118381	ue40d08.x1 Sugano mous	1.05e+00
c	863	15	5.0	582	13	AA863947	vx87b09.r1 Soares 2NbM	1.05e+00
	864	15	5.0	583	6	AA398975	xt93a06.r1 Soares test	1.05e+00
c	865	15	5.0	584	8	AA527357	ng40b04.s1 NCI_CGAP_Co	1.05e+00
c	866	15	5.0	584	26	FR0013842	F.rubripes GSS sequenc	1.05e+00
	867	15	5.0	587	12	AA719264	ah46h09.s1 Soares test	1.05e+00
	868	15	5.0	587	27	AQ008399	CIT-HSP-2173K13.TR.1 C	1.05e+00
c	869	15	5.0	587	13	C85966	Mus musculus fertilize	1.05e+00
	870	15	5.0	588	23	AA120618	mp71a12.r1 Soares 2NbM	1.05e+00
	871	15	5.0	589	17	AA873349	oh64h01.s1 NCI_CGAP_Ki	1.05e+00
	872	15	5.0	590	27	B59270	CIT-HSP-2015L2.TRB CIT	1.05e+00
	873	15	5.0	591	6	AA127875	zl12b04.r1 Soares preg	1.05e+00
c	874	15	5.0	591	8	AA542185	vj80d10.r1 Soares mous	1.05e+00
c	875	15	5.0	591	6	AA121002	zl01a12.r1 Soares preg	1.05e+00
c	876	15	5.0	592	22	AA148013	zo51c09.r1 Stratagene	1.05e+00
	877	15	5.0	593	23	AA097397	mm36c07.r1 Stratagene	1.05e+00
	878	15	5.0	595	15	AA946763	oq50c11.s1 NCI_CGAP_Ki	1.05e+00
c	879	15	5.0	595	8	AA527588	ng42d08.s1 NCI_CGAP_Co	1.05e+00
	880	15	5.0	597	9	AA596418	vm55d11.r1 Stratagene	1.05e+00
	881	15	5.0	599	27	FR0032345	Fugu rubripes GSS sequ	1.05e+00
c	882	15	5.0	600	22	AA037310	zc52g07.r1 Soares sene	1.05e+00
c	883	15	5.0	601	21	N95233	zb53h10.s1 Soares feta	1.05e+00
c	884	15	5.0	602	12	AA763285	vw60h05.r1 Soares mous	1.05e+00
	885	15	5.0	603	26	AQ019278	CIT-HSP-2306H14.TF CIT	1.05e+00
c	886	15	5.0	605	26	B52384	CIT-HSP-2003G1.TR CIT-	1.05e+00
c	887	15	5.0	605	10	AA641637	nr80c06.s1 NCI_CGAP_Pr	1.05e+00
	888	15	5.0	609	23	AA162033	ms16f01.r1 Stratagene	1.05e+00
	889	15	5.0	612	26	FR0027605	F.rubripes GSS sequenc	1.05e+00
	890	15	5.0	614	27	FR0031533	Fugu rubripes GSS sequ	1.05e+00
	891	15	5.0	614	26	B01349	cSRL-130d11-u cSRL flo	1.05e+00
	892	15	5.0	615	26	B55701	CIT-HSP-2003E11.TR CIT	1.05e+00
	893	15	5.0	618	14	AA946018	EST201517 Normalized r	1.05e+00
c	894	15	5.0	623	26	B77677	T2E122TR TAMU Arabidop	1.05e+00
c	895	15	5.0	628	27	AQ039559	CIT-HSP-2325F3.TV CIT-	1.05e+00
	896	15	5.0	629	20	DCU47090	Daucus carota somatic	1.05e+00
	897	15	5.0	630	26	B27916	T2lC4TFB TAMU Arabidop	1.05e+00
	898	15	5.0	631	28	AQ160048	mgxb0002N24r CUGI Rice	1.05e+00
c	899	15	5.0	643	23	AA153888	mq56d05.r1 Soares 2NbM	1.05e+00
c	900	15	5.0	648	15	AI042799	uc84d03.x1 Sugano mous	1.05e+00
	901	15	5.0	651	29	AQ238234	RPCI11-74G14.TK RPCI11	1.05e+00
c	902	15	5.0	652	11	AA567993	HL02094.5prime HL Dros	1.05e+00
c	903	15	5.0	652	11	AA567983	HL02082.5prime HL Dros	1.05e+00
	904	15	5.0	656	27	AQ057528	CIT-HSP-2340L12.TR CIT	1.05e+00
c	905	15	5.0	658	12	AA739977	742 PtIFG2 Pinus taeda	1.05e+00
	906	15	5.0	669	26	B24100	F18K21TF IGF Arabidops	1.05e+00
	907	15	5.0	677	23	AA122635	mq09g07.r1 Barstead MP	1.05e+00
	908	15	5.0	677	28	AQ109387	CIT-HSP-2370N9.TR CIT-	1.05e+00
c	909	15	5.0	684	23	AA186185	mt63d04.r1 Soares 2NbM	1.05e+00

	910	15	5.0	686	12	AA803457	GM13219.5prime GM Dros	1.05e+00
c	911	15	5.0	690	15	AI043103	uc86f09.x1 Sugano mous	1.05e+00
	912	15	5.0	692	28	AG013171	Homo sapiens genomic D	1.05e+00
	913	15	5.0	692	26	AG005526	Homo sapiens genomic D	1.05e+00
c	914	15	5.0	697	23	AA154926	ms02h02.r1 Stratagene	1.05e+00
	915	15	5.0	698	11	AA690494	vu52g06.r1 Soares mous	1.05e+00
c	916	15	5.0	710	11	AA697787	HL03153.5prime HL Dros	1.05e+00
c	917	15	5.0	724	26	B56102	CIT-HSP-2005P21.TV CIT	1.05e+00
	918	15	5.0	732	17	AI128037	qc47a01.x1 Soares_preg	1.05e+00
c	919	15	5.0	733	12	AA764076	vw09f05.r1 Soares 2NbM	1.05e+00
	920	15	5.0	738	14	AA981749	ua36a01.r1 Soares mous	1.05e+00
	921	15	5.0	758	10	AA633250	nq57g10.s1 NCI_CGAP_Co	1.05e+00
	922	15	5.0	775	29	AQ253337	HS_2047_B1_A02_T7 CIT	1.05e+00
c	923	15	5.0	779	26	B20293	T2lP20-T7 TAMU Arabido	1.05e+00
	924	15	5.0	782	18	AI097555	qa04f12.x1 NCI_CGAP_Br	1.05e+00
	925	15	5.0	790	20	W28657	49g5 Human retina cDNA	1.05e+00
c	926	15	5.0	835	23	AA174888	ms77g05.r1 Soares mous	1.05e+00
	927	15	5.0	839	17	AI114961	ui41a08.y1 Sugano mous	1.05e+00
	928	15	5.0	1020	26	B08884	F1K4-T7 IGF Arabidopsi	1.05e+00
	929	15	5.0	1121	26	B11013	T2lC4-T7 TAMU Arabidop	1.05e+00
c	930	15	5.0	1248	26	B09703	F2lE18-T7 IGF Arabidop	1.05e+00
c	931	15	5.0	1345	24	HSD66681	Human clone EST157481	1.05e+00
c	932	14	4.7	204	5	T40715	ya09g01.s2 Homo sapien	2.48e+01
	933	14	4.7	210	14	AA878773	oe80g04.s1 NCI_CGAP_Lu	2.48e+01
	934	14	4.7	226	24	AA328466	EST32008 Embryo, 12 we	2.48e+01
	935	14	4.7	230	7	AA453300	zx12b03.s1 Soares tota	2.48e+01
c	936	14	4.7	254	25	AA378668	EST91395 Synovial sarc	2.48e+01
c	937	14	4.7	255	8	AA465398	aa23h04.s1 NCI_CGAP_GC	2.48e+01
	938	14	4.7	264	21	AA055320	zf21a01.s1 Soares feta	2.48e+01
	939	14	4.7	281	25	AA367164	EST78395 Pancreas tumo	2.48e+01
	940	14	4.7	283	28	AQ199226	RPCI11-58M12.TJ RPCI11	2.48e+01
	941	14	4.7	291	21	T16216	IB968 Infant brain, Be	2.48e+01
	942	14	4.7	293	5	T67068	ya52f01.s3 Homo sapien	2.48e+01
c	943	14	4.7	294	12	AA769982	ah81e05.s1 Soares test	2.48e+01
	944	14	4.7	295	15	AI052874	SWBmL35BH1OP01SK Brugi	2.48e+01
c	945	14	4.7	308	12	AA737461	oa49c05.s1 NCI_CGAP_GC	2.48e+01
c	946	14	4.7	311	19	N56704	yw76b07.r1 Homo sapien	2.48e+01
c	947	14	4.7	311	17	AU023583	Mus musculus unfertili	2.48e+01
	948	14	4.7	312	19	H69670	yr92c07.s1 Homo sapien	2.48e+01
	949	14	4.7	343	16	M86114	EST02640 Homo sapiens	2.48e+01
	950	14	4.7	361	28	AQ128096	HS_3094_A1_D02_MR CIT	2.48e+01
	951	14	4.7	363	20	R81088	EST22e063 WATM1 Homo s	2.48e+01
	952	14	4.7	366	16	H54134	yq90g06.r1 Homo sapien	2.48e+01
	953	14	4.7	368	15	AI043678	UI-R-C0-jl-c-02-0-UI.s	2.48e+01
	954	14	4.7	386	24	AA318334	EST20402 Retina II Hom	2.48e+01
c	955	14	4.7	395	20	N80413	yr92d05.r1 Homo sapien	2.48e+01
c	956	14	4.7	399	18	AI168045	os92h06.x1 Soares_para	2.48e+01
c	957	14	4.7	403	21	AA048277	mj28f06.r1 Soares mous	2.48e+01
	958	14	4.7	407	6	AA405882	zu57b04.r1 Soares ovar	2.48e+01
	959	14	4.7	411	16	H42578	yo09b07.r1 Homo sapien	2.48e+01
	960	14	4.7	412	6	AA410867	zv35g12.r1 Soares ovar	2.48e+01

	961	14	4.7	414	28	AQ189735	HS_3218_A2_F12_MR CIT 2.48e+01
	962	14	4.7	417	28	AQ208835	HS_3229_A2_H02_MR CIT 2.48e+01
	963	14	4.7	418	17	AU011124	Schizosaccharomyces po 2.48e+01
c	964	14	4.7	418	22	AA099773	zl81a04.r1 Stratagene 2.48e+01
c	965	14	4.7	419	21	W58420	zd20c03.s1 Soares feta 2.48e+01
c	966	14	4.7	423	20	W62383	md98f04.r1 Soares mous 2.48e+01
	967	14	4.7	425	23	AA232837	zr44g03.r1 Soares NhHM 2.48e+01
c	968	14	4.7	429	7	AA455292	zx80c09.s1 Soares ovar 2.48e+01
c	969	14	4.7	430	8	AA292916	zt66e10.r1 Soares test 2.48e+01
	970	14	4.7	435	5	R35845	yg67g10.r1 Homo sapien 2.48e+01
c	971	14	4.7	443	17	AI051855	ox44g04.s1 Soares_tota 2.48e+01
c	972	14	4.7	445	29	AQ147709	HS_3079_A1_D04_MR CIT 2.48e+01
c	973	14	4.7	450	18	AI146999	qb34h01.x1 Soares_preg 2.48e+01
c	974	14	4.7	454	5	R24963	yg41e03.r1 Homo sapien 2.48e+01
	975	14	4.7	458	7	AA459301	zx89c05.s1 Soares ovar 2.48e+01
c	976	14	4.7	458	15	AA825547	od55a09.s1 NCI_CGAP_GC 2.48e+01
c	977	14	4.7	472	23	AA062576	zf68c04.r1 Soares pine 2.48e+01
	978	14	4.7	477	23	AA164152	mq12a10.r1 Barstead MP 2.48e+01
	979	14	4.7	478	13	AA188796	zp77c11.s1 Stratagene 2.48e+01
c	980	14	4.7	478	15	AI019685	ua92e10.r1 Soares mous 2.48e+01
	981	14	4.7	496	22	AA033700	zf01c12.s1 Soares feta 2.48e+01
	982	14	4.7	496	17	AI121103	ud69g09.x1 Sugano mous 2.48e+01
	983	14	4.7	499	23	AA237244	mx17g06.r1 Soares mous 2.48e+01
	984	14	4.7	501	28	AQ191428	HS_3233_A1_B07_T7 CIT 2.48e+01
	985	14	4.7	510	18	AI161165	qb50e03.x1 NCI_CGAP_Br 2.48e+01
c	986	14	4.7	526	10	AA617805	nq04f10.s1 NCI_CGAP_Lu 2.48e+01
	987	14	4.7	529	6	AA193391	zr40b01.r1 Soares NhHM 2.48e+01
c	988	14	4.7	535	24	AA288865	mr43b02.r1 Life Tech m 2.48e+01
	989	14	4.7	548	9	AA574029	nk11a06.s1 NCI_CGAP_Co 2.48e+01
	990	14	4.7	553	28	AQ186316	HS_3084_A1_D10_MF CIT 2.48e+01
c	991	14	4.7	588	17	AU020976	Mus musculus 8-cell em 2.48e+01
c	992	14	4.7	590	19	N32746	yx75h11.r1 Homo sapien 2.48e+01
	993	14	4.7	594	12	AA779457	af22d11.s1 Soares tota 2.48e+01
	994	14	4.7	627	9	AA573839	nk08c04.s1 NCI_CGAP_Co 2.48e+01
c	995	14	4.7	638	14	AA892565	EST196368 Normalized r 2.48e+01
c	996	14	4.7	702	20	W28319	45c1 Human retina cDNA 2.48e+01
	997	14	4.7	717	18	AI172967	ud62h09.x1 Sugano mous 2.48e+01
c	998	14	4.7	726	29	AQ146163	HS_2231_A1_D10_MR CIT 2.48e+01
c	999	14	4.7	759	15	AI027777	ov99f04.x1 Soares_test 2.48e+01
c	1000	14	4.7	841	21	W36370	mb63e11.r1 Soares mous 2.48e+01

ALIGNMENTS

```
RESULT    1
LOCUS       C00217         313 bp    DNA             EST       11-JUL-1996
DEFINITION  HUMGS0005879, Human Gene Signature, 3'-directed cDNA sequence, mRNA
            sequence.
ACCESSION   C00217
NID         g1432447
KEYWORDS    EST; Gene Signature; GS; EST(expressed sequence tag); BodyMap; gene
```



```
            expression.
SOURCE      One or more human adult tissue.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 313)
  AUTHORS   Okubo,K.
  TITLE     BodyMap; human gene expression database
  JOURNAL   Unpublished (1995)
REFERENCE   2  (bases 1 to 313)
  AUTHORS   Okubo,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-DEC-1995) to the DDBJ/EMBL/GenBank databases. Kousaku
            Okubo, Osaka University, Institute for Molecular and Cellular Bio;
            1-3,Yamada-oka, Suita, Osaka Pref. 565, Japan
            (E-mail:kousaku@imcb.osaka-u.ac.jp, Tel:06-877-5111(ex.3315),
            Fax:06-877-1922)
COMMENT     On Jul 18, 1996 this sequence version replaced gi:1416731.
            We are not submitting the same cDNA sequence redundantly to DDBJ
            since 1993. For the abundance information of clones with this
            sequence in this library and as well as in other 3'-directed
            libraries, see ' http://www.imcb.osaka-u.ac.jp/bodymap'.  The
            sequences of the clones represented by this GS sequences is also
            found there.
FEATURES             Location/Qualifiers
     source          1. .313
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT       75 a      71 c      75 g      77 t      15 others
ORIGIN

  Query Match          100.0%;  Score 298;  DB 6;  Length 313;
  Best Local Similarity 100.0%;  Pred. No. 0.00e+00;
  Matches   313;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db       1 GATCCTCTCTCCCAGGGAATCCGACACAGGAGGAACCCCTTCTNTGGTTGANCTGGGCCA 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy       1 GATCCTCTCTCCCAGGGAATCCGACACAGGAGGAACCCCTTCTNTGGTTGANCTGGGCCA 60

Db      61 GGCCTAAGAGTAGCAGGAACTCTAAGACCACAGAGTTTTTTATAAATGTATAAATGTATC 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      61 GGCCTAAGAGTAGCAGGAACTCTAAGACCACAGAGTTTTTTATAAATGTATAAATGTATC 120

Db     121 AAGCCAAATNTGCAGATGCTAACTGGACATTCTGGGGAACTGGGCACCAGGNGTGCCTTC 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     121 AAGCCAAATNTGCAGATGCTAACTGGACATTCTGGGGAACTGGGCACCAGGNGTGCCTTC 180

Db     181 ATACACTGTACCCCAGCTCTNTTCTAAAAGAGAAGTNGGTGGGCACACTTGAACTGTTTG 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     181 ATACACTGTACCCCAGCTCTNTTCTAAAAGAGAAGTNGGTGGGCACACTTGAACTGTTTG 240
```



```
Db     241 GTGGNCCNAACCACAGGANGCTGCAATTNTNTGGNTTAGGGTGATACTTTTGGCCCTNCN 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     241 GTGGNCCNAACCACAGGANGCTGCAATTNTNTGGNTTAGGGTGATACTTTTGGCCCTNCN 300

Db     301 TGTGGCCCCTNTT 313
           |||||||||||||
Qy     301 TGTGGCCCCTNTT 313


RESULT    2
LOCUS       AA432289       457 bp    mRNA            EST       22-MAY-1997
DEFINITION  zw69a11.r1 Soares testis NHT Homo sapiens cDNA clone 781436 5',
            mRNA sequence.
ACCESSION   AA432289
NID         g2114677
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 457)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J.,
            Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-Merck EST Project 1997
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 401.
FEATURES             Location/Qualifiers
     source          1. .457
                     /organism="Homo sapiens"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from mRNA obtained from Clontech
                     Laboratories, Inc., and primed with  a Not I - oligo(dT)
                     primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGGCCCAATTTTTTTTTTTTTTTT 3']
                     Double-stranded cDNA was ligated to Eco RI adaptors
```



```
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization to Cot5, and was
                     constructed  by Bento Soares and M. Fatima Bonaldo."
                     /db_xref="taxon:9606"
                     /clone="781436"
                     /clone_lib="Soares testis NHT"
                     /sex="male"
                     /lab_host="DH10B"
     mRNA            <1. .>457
BASE COUNT      120 a    108 c    106 g    123 t
ORIGIN

  Query Match          57.4%;  Score 171;  DB 6;  Length 457;
  Best Local Similarity 97.7%;  Pred. No. 0.00e+00;
  Matches   171;  Conservative    0;  Mismatches   4;  Indels   0;  Gaps   0;

Db     283 GATCCTCTCTCCCAGGGAATCCGACACAGGAGGAACCCCTTCTCTGGTTGAGCTGGGCCA 342
           |||||||||||||||||||||||||||||||||||||||||||| ||||||| ||||||||
Qy       1 GATCCTCTCTCCCAGGGAATCCGACACAGGAGGAACCCCTTCTNTGGTTGANCTGGGCCA 60

Db     343 GGCCTAAGAGTAGCAGGAACTCTAAGACCACAGAGTTTTTTATAAATGTATAAATGTATC 402
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      61 GGCCTAAGAGTAGCAGGAACTCTAAGACCACAGAGTTTTTTATAAATGTATAAATGTATC 120

Db     403 AAGCCAAATGTGCAGATGCTAACTGGACATTCTGGGGAACTGGGCACCAGGAGTG 457
           |||||||| ||||||||||||||||||||||||||||||||||||||||||| |||
Qy     121 AAGCCAAATNTGCAGATGCTAACTGGACATTCTGGGGAACTGGGCACCAGGNGTG 175


RESULT    3
LOCUS       N76293         542 bp    mRNA            EST       28-JAN-1997
DEFINITION  yv38f01.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone
            245017 5', mRNA sequence.
ACCESSION   N76293
NID         g1238871
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 542)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
            Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
```

```
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 957   Std Error: 0.00
            Seq primer: reverse ET
            High quality sequence stop: 382.
FEATURES              Location/Qualifiers
     source           1. .542
                      /organism="Homo sapiens"
                      /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                      with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                      1st strand cDNA was primed with a Pac I - oligo(dT) primer
                      [5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTT 3'],
                      double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Pac I and cloned into the Pac I
                      and Eco RI sites of the modified pT7T3 vector.  Library
                      went through one round of normalization. Library
                      constructed by Bento Soares and M.Fatima Bonaldo."
                      /db_xref="taxon:9606"
                      /clone="245017"
                      /clone_lib="Soares fetal liver spleen 1NFLS"
                      /sex="male"
                      /dev_stage="20 week-post conception fetus"
                      /lab_host="DH10B (ampicillin resistant)"
     mRNA             <1. .>542
BASE COUNT      137 a    136 c    133 g    134 t      2 others
ORIGIN

  Query Match           52.3%;  Score 156;  DB 22;  Length 542;
  Best Local Similarity  94.3%;  Pred. No. 0.00e+00;
  Matches   216;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Db    199 CTAAGAGTAGCAGGAACTCTAAGACCACAGAGTTTTTTATAAATGTATAAATGTATCAAG 258
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     64 CTAAGAGTAGCAGGAACTCTAAGACCACAGAGTTTTTTATAAATGTATAAATGTATCAAG 123

Db    259 CCAAATGTGCAGATGCTAACTGGACATTCTGGGGAACTGGGCACCAGGAGTGCTNTCATA 318
          |||||| |||||||||||||||||||||||||||||||||||||||||||| ||||  |||||
Qy    124 CCAAATNTGCAGATGCTAACTGGACATTCTGGGGAACTGGGCACCAGGNGTGCCTTCATA 183

Db    319 CACTGTACCCCAGCTCTCTTCTAAAAGAGAAGTGGGTGGGCACACTCGAACTGTTTGGTG 378
          |||||||||||||||| |||||||||||||| |||||||||| ||||||||||||
Qy    184 CACTGTACCCCAGCTCTNTTCTAAAAGAGAAGTNGGTGGGCACACTTGAACTGTTTGGTG 243

Db    379 GCCCCAACCACAGGAAGCTGCAATTCTCTGGCTTAGGGTGATACTTTTG 427
          | || |||||||||| ||||||||| | ||| ||||||||||||||||||
Qy    244 GNCCNAACCACAGGANGCTGCAATTNTNTGGNTTAGGGTGATACTTTTG 292
```



```
RESULT    4
LOCUS       W24101         289 bp    mRNA            EST       20-AUG-1996
DEFINITION  zb48f06.r1 Soares fetal lung NbHL19W Homo sapiens cDNA clone 306851
            5', mRNA sequence.
ACCESSION   W24101
NID         g1300918
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 289)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
            Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 821   Std Error: 0.00
            Seq primer: mob.REGA+ET
            High quality sequence stop: 136.
FEATURES              Location/Qualifiers
     source           1. .289
                      /organism="Homo sapiens"
                      /note="Organ: lung; Vector: pT7T3D (Pharmacia) with a
                      modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                      strand cDNA was primed with a Not I - oligo(dT) primer
                      [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCAATTTTTTTTTTTTTTTT-3'],
                      double-stranded cDNA was size selected, ligated to Eco RI
                      adapters (Pharmacia), digested with Not I and cloned into
                      the Not I and Eco RI sites of a modified pT7T3 vector
                      (Pharmacia). Library went through one round of
                      normalization to a Cot = 5. Library constructed by Bento
                      Soares and M.Fatima Bonaldo. This library was constructed
                      from the same fetus as the fetal heart library, Soares
                      fetal heart NbHH19W."
                      /db_xref="taxon:9606"
                      /clone="306851"
                      /clone_lib="Soares fetal lung NbHL19W"
```



```
                      /dev_stage="19 weeks"
                      /lab_host="DH10B (ampicillin resistant)"
     mRNA             <1. .>289
BASE COUNT       75 a     68 c     74 g     72 t
ORIGIN

  Query Match           48.0%;  Score 143;  DB 21;  Length 289;
  Best Local Similarity  97.9%;  Pred. No. 0.00e+00;
  Matches   143;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Db     39 CTAAGAGTAGCAGGAACTCTAAGACCACAGAGTTTTTTATAAATGTATAAATGTATCAAG 98
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     64 CTAAGAGTAGCAGGAACTCTAAGACCACAGAGTTTTTTATAAATGTATAAATGTATCAAG 123

Db     99 CCAAATGTGCAGATGCTAACTGGACATTCTGGGGAACTGGGCACCAGGAGTGCCTTCATA 158
          |||||| |||||||||||||||||||||||||||||||||||||||||| ||||||||||
Qy    124 CCAAATNTGCAGATGCTAACTGGACATTCTGGGGAACTGGGCACCAGGNGTGCCTTCATA 183

Db    159 CACTGTACCCCAGCTCTCTTCTAAAA 184
          |||||||||||||||| ||||||||
Qy    184 CACTGTACCCCAGCTCTNTTCTAAAA 209


RESULT    5
LOCUS       AA090434       277 bp    mRNA            EST       24-OCT-1996
DEFINITION  y0386.seq.F Fetal heart, Lambda ZAP Express Homo sapiens cDNA 5',
            mRNA sequence.
ACCESSION   AA090434
NID         g1635018
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 277)
  AUTHORS   Liew,C.C.
  TITLE     cDNAs from fetal heart (1996)
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Liew CC
            Molecular Cardiology
            University of Toronto
            Banting Institute, 100 College St., Toronto, Ontario, M5G1L5
            Tel: 4169788758
            Fax: 4169785650
            Email: liewcc@utcc.utoronto.ca
            PCR Primers
            FORWARD: 5' GCCAAGCTCGAAATTAACCCTCACTAAAGGG 3'
            BACKWARD: 5' CCAGTGAATTGTAATACGACTCACTATAGGGCG 3'
            Seq primer: 5' GAAATTAACCCTCACTAAAGGG 3'.
FEATURES              Location/Qualifiers
```



```
     source           1. .277
                      /organism="Homo sapiens"
                      /note="Vector: Lambda ZAP Express; Site_1: EcoRI; Site_2:
                      XhoI; mRNA was purified from human fetal hearts (8-10
                      weeks).  cDNA was synthesized using a XhoI-Oligo dT
                      adaptor-primer.  EcoRI adaptors were ligated, followed by
                      digestion with XhoI, for directional cloning into
                      predigested lambda ZAP Express."
                      /db_xref="taxon:9606"
                      /clone_lib="Fetal heart, Lambda ZAP Express"
                      /lab_host="E. coli XL1-Blue"
     mRNA             <1. .>277
BASE COUNT       72 a     68 c     71 g     66 t
ORIGIN

  Query Match           26.8%;  Score 80;  DB 22;  Length 277;
  Best Local Similarity  97.6%;  Pred. No. 3.90e-159;
  Matches    80;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Db    126 ATCCGACACAGGAGGAACCCCTTCTCTGGTTGAGCTGGGCCAGGCCTAAGAGTAGCAGGA 185
          |||||||||||||||||||||||||| ||||||| ||||||||||||||||||||||||||
Qy     19 ATCCGACACAGGAGGAACCCCTTCTNTGGTTGANCTGGGCCAGGCCTAAGAGTAGCAGGA 78

Db    186 ACTCTAAGACCACAGAGTTTTT 207
          |||||||||||||||||||||
Qy     79 ACTCTAAGACCACAGAGTTTTT 100


RESULT    6
LOCUS       AA503247       509 bp    mRNA            EST       19-AUG-1997
DEFINITION  ne45h02.s1 NCI_CGAP_Co3 Homo sapiens cDNA clone IMAGE:900339, mRNA
            sequence.
ACCESSION   AA503247
NID         g2238214
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 509)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck,
            M.D., Ph.D.
```

```
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arraying: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           www-bio.llnl.gov/bbrp/image/image.html

           Insert Length: 891   Std Error: 0.00
           Seq primer: -40m13 fwd. ET from Amersham
           High quality sequence stop: 474.
FEATURES             Location/Qualifiers
     source          1. .509
                     /organism="Homo sapiens"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from 12 pooled bulk tumor samples and primed
                     with a Not I - oligo(dT) primer. Double-stranded cDNA was
                     ligated to Eco RI adaptors (Pharmacia), digested with Not
                     I and cloned into the Not I and Eco RI sites of the
                     modified pT7T3 vector. Library went through one round of
                     normalization."
                     /db_xref="taxon:9606"
                     /clone="IMAGE:900339"
                     /clone_lib="NCI_CGAP_Co3"
                     /sex="pooled"
                     /tissue_type="colon"
                     /lab_host="DH10B"
     mRNA            <1. .>509
BASE COUNT      144 a    130 c    114 g    121 t
ORIGIN

  Query Match           14.4%;  Score 43;  DB 8;  Length 509;
  Best Local Similarity  87.8%;  Pred. No. 1.15e-60;
  Matches    43;  Conservative    0;  Mismatches   6;  Indels   0;  Gaps   0;

Db    461 CAAAAGTATCACCCTAAGCCAGAGAATTGCAGCTTCCTGTGGTTGGGGC 509
          ||||||||||||||||| ||| | |||||||||| ||||||||||| || |
Cp    292 CAAAAGTATCACCCTAANCCANANAATTGCAGCNTCCTGTGGTTNGGNC 244


RESULT    7
LOCUS       AA934377      511 bp    mRNA            EST       28-APR-1998
DEFINITION  op27g12.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
            IMAGE:1578118 3', mRNA sequence.
ACCESSION   AA934377
NID         g3091534
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
```



```
REFERENCE   1  (bases 1 to 511)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
           Contact: Robert Strausberg, Ph.D.
           Tel: (301) 496-1550
           Email: Robert_Strausberg@nih.gov
           This clone is available royalty-free through LLNL ; contact the
           IMAGE Consortium (info@image.llnl.gov) for further information.
           Seq primer: -40m13 fwd. ET from Amersham
           High quality sequence stop: 473.
FEATURES             Location/Qualifiers
     source          1. .511
                     /organism="Homo sapiens"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from three normalized
                     libraries (fetal lung NbHL19W, testis NHT, and B-cell
                     NCI_CGAP_GCB1) were mixed, and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR-amplified cDNAs from pools of 5,000 clones made
                     from the same 3 libraries. The pools consisted of
                     I.M.A.G.E. clones 297480-302087, 682632-687239,
                     726408-728711, and 729096-731399. Subtraction by Bento
                     Soares and M. Fatima Bonaldo."
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1578118"
                     /clone_lib="Soares_NFL_T_GBC_S1"
                     /lab_host="DH10B"
BASE COUNT      144 a    131 c    114 g    122 t
ORIGIN

  Query Match           14.4%;  Score 43;  DB 14;  Length 511;
  Best Local Similarity  87.8%;  Pred. No. 1.15e-60;
  Matches    43;  Conservative    0;  Mismatches   6;  Indels   0;  Gaps   0;

Db    463 CAAAAGTATCACCCTAAGCCAGAGAATTGCAGCTTCCTGTGGTTGGGGC 511
          ||||||||||||||||| ||| | |||||||||| ||||||||||| || |
Cp    292 CAAAAGTATCACCCTAANCCANANAATTGCAGCNTCCTGTGGTTNGGNC 244


RESULT    8
LOCUS       FR0006119     619 bp    DNA             GSS       02-MAR-1997
DEFINITION  F.rubripes GSS sequence, clone 106C21bA7, genomic survey sequence.
ACCESSION   Z89929
NID         g1867143
KEYWORDS    GSS; genome survey sequence.
SOURCE      Fugu rubripes.
```



```
  ORGANISM  Fugu rubripes
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Actinopterygii; Neopterygii; Teleoatei; Euteleostei;
            Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontoidei;
            Tetraodontidae; Fugu.
REFERENCE   1  (bases 1 to 619)
  AUTHORS   Elgar,G., Clark,M., Smith,S., Meek,S., Warner,S., Umrania,Y.,
            Williams,G. and Brenner,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-FEB-1997) MRC Human Genome Mapping Project Resource
            Centre Hinxton, Cambridge, CB10 1SB. Email: biohelp@hgmp.mrc.ac.uk
COMMENT     Vector: pBluescript II KS
            V_type: phagemid
            PRIMER: KS
            DESCR:
            One pass dye-terminator sequencing of cosmid cloned genomic
            sequence.
FEATURES             Location/Qualifiers
     source          1. .619
                     /organism="Fugu rubripes"
                     /db_xref="taxon:31033"
                     /clone_lib="cosmid 106C21"
                     /clone="106C21bA7"
BASE COUNT      131 a    132 c    148 g    146 t     62 others
ORIGIN

  Query Match            7.0%;  Score 21;  DB 26;  Length 619;
  Best Local Similarity  87.5%;  Pred. No. 8.09e-11;
  Matches    21;  Conservative    0;  Mismatches   3;  Indels   0;  Gaps   0;

Db     11 GTCGGANNCCCNGGGAGAGAGGAT 34
          ||||||  ||| ||||||||||||
Cp     25 GTCGGATTCCCTGGGAGAGAGGAT 2


RESULT    9
LOCUS       H31765        325 bp    mRNA            EST       02-APR-1998
DEFINITION  EST106132 Rat PC-12 cells, untreated Rattus sp. cDNA clone RPCAX55
            5' end, mRNA sequence.
ACCESSION   H31765
NID         g977182
KEYWORDS    EST.
SOURCE      Rattus sp.
  ORGANISM  Rattus sp.
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 325)
  AUTHORS   Lee,N.H., Weinstock,K.G., Kirkness,E.F., Earle-Hughes,J.A.,
            Fuldner,R.A., Marmaras,S., Glodek,A., Gocayne,J.D., Adams,M.D.,
            Kerlavage,A.R., Fraser,C.M. and Venter,J.C.
  TITLE     Comparative expressed-sequence-tag analysis of differential gene
```



```
            expression profiles in PC-12 cells before and after nerve growth
            factor treatment
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 92, 8303-8307 (1995)
  MEDLINE   95396786
COMMENT     Other_ESTs: EST106131
            Contact: Lee, NH
            ATCC
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
            For clone availability please contact  the TIGR Database
            (tdbinfo@tdb.tigr.org)
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .325
                     /organism="Rattus sp."
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; poly(A)+ RNA was purified from untreated PC12 cells
                     cultured for 9 days.  cDNA was constructed using an
                     oligo-dT primer and directionally cloned using the Lambda
                     ZAP II Vector Kit by Stratagene"
                     /db_xref="ATCC (inhost):2001033"
                     /db_xref="taxon:10118"
                     /clone="RPCAX55"
                     /clone_lib="Rat PC-12 cells, untreated"
BASE COUNT       81 a     73 c     95 g     74 t      2 others
ORIGIN

  Query Match            6.7%;  Score 20;  DB 14;  Length 325;
  Best Local Similarity  95.2%;  Pred. No. 5.91e-09;
  Matches    20;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db     84 AGCCAAATGTGCAGATGCTAA 104
          |||||||| ||||||||||||
Qy    122 AGCCAAATNTGCAGATGCTAA 142


RESULT   10
LOCUS       AQ157283      622 bp    DNA             GSS       10-SEP-1998
DEFINITION  nbxb0009023f CUGI Rice BAC Library Oryza sativa genomic clone
            nbxb0009023f, genomic survey sequence.
ACCESSION   AQ157283
NID         g3554308
KEYWORDS    GSS.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Charophyta/Embryophyta group;
            Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
```

```
REFERENCE   1  (bases 1 to 622)
  AUTHORS   Yu,Y., Budiman,M.A., Boyd,C.A., Gaudette,B., Gayle,A.,
            Kingsbury,R., Phillips,K., Sasinowski,M., Dean,R.A. and Wing,R.A.
  TITLE     A BAC End Sequencing Framework to Sequence the Rice Genome
  JOURNAL   Unpublished (1998)
COMMENT
            Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: TAATACGACTCACTATAGGG
            Class: BAC ends
            High quality sequence stop: 444.
FEATURES             Location/Qualifiers
     source          1. .622
                     /organism="Oryza sativa"
                     /strain="Japonica"
                     /cultivar="Nipponbare"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII; Rice is one of two most popular grains in the
                     world. Half of the world  population especially those
                     inhabiting highly populated areas of the  humid tropics
                     and subtropics, rely on rice as their primary source of
                     carbohydrate. Monocotyledonous rice is a diploid plant
                     (2n=24) with a  haploid genome equivalent of 431 Mbp
                     (Arumuganathan and Earle, 1991).  The relatively small
                     genome of rice, three times larger than that of
                     Arabidopsis, makes it suitable for genomic studies. In
                     order to  facilitate positional cloning, physical mapping
                     and genome sequencing  of rice, we have constructed a BAC
                     library from Oryza sativa, Nipponbare  variety. The
                     library contains 36,864 clones with an average insert size
                      of 128.5 Kb providing 10.9 haploid genome equivalents.
                     The deep  coverage allows the isolation a particular
                     sequence with a probability  of 99.9 %. Two high density
                     filters, each containing 18,432 clones  (doubly spotted),
                     represent the whole library for colony screening."
                     /db_xref="taxon:4530"
                     /clone="nbxb0009023f"
                     /clone_lib="CUGI Rice BAC Library"
                     /tissue_type="Leaf"
                     /lab_host="E. coli DH10B"
BASE COUNT      217 a     75 c     96 g    233 t      1 others
ORIGIN

  Query Match             6.4%;  Score 19;  DB 28;  Length 622;
  Best Local Similarity 100.0%;  Pred. No. 3.74e-07;
  Matches    19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```



```
Db    533 TATAAATGTATAAATGTAT 551
          ||||||||||||||||||
Qy    101 TATAAATGTATAAATGTAT 119


RESULT   11
LOCUS       HSU50541      117 bp    mRNA            EST       14-MAR-1996
DEFINITION  Human MATC-1 mRNA sequence.
ACCESSION   U50541
NID         g1224132
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 117)
  AUTHORS   Cucco,C., Fanciulli,M., Bruno,T. and Ricca,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-1996) Carla Cucco, CSP - MCF, Istituto Regina
            Elena, 156, Via delle Messi d'Oro, Roma RM 00158, Italia
FEATURES             Location/Qualifiers
     source          1. .117
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="MATC-1"
BASE COUNT       24 a     36 c     33 g     24 t
ORIGIN

  Query Match             6.0%;  Score 18;  DB 19;  Length 117;
  Best Local Similarity  90.0%;  Pred. No. 2.03e-05;
  Matches    18;  Conservative    0;  Mismatches   2;  Indels   0;  Gaps   0;

Db     13 CAAACCACAGGACGCTGCAA 32
          | |||||||||| |||||||
Qy    247 CNAACCACAGGANGCTGCAA 266


RESULT   12
LOCUS       AA509406      226 bp    mRNA            EST       08-JUL-1997
DEFINITION  vg18f08.r1 Soares mouse NbMH Mus musculus cDNA clone 861735 5'
            similar to TR:G965470 G965470 FK-506 BINDING PROTEIN HOMOLOGUE. :,
            mRNA sequence.
ACCESSION   AA509406
NID         g2247260
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae;
            Murinae; Mus.
```



```
REFERENCE   1  (bases 1 to 226)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:505823
            Possible reversed clone: similarity on wrong strand
            Seq primer: -28ml3 rev2 ET from Amersham
            High quality sequence stop: 116.
FEATURES             Location/Qualifiers
     source          1. .226
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGAAAGTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not
                     I and Eco RI sites of  the modified pT7T3 vector. RNA
                     provided by Dr. Minoru Ko, Wayne State Univ. Library
                     constructed and normalized by Bento Soares and M.Fatima
                     Bonaldo."
                     /db_xref="taxon:10090"
                     /clone="861735"
                     /clone_lib="Soares mouse NbMH"
                     /sex="male"
                     /tissue_type="heart"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
     mRNA            <1. .>226
BASE COUNT       50 a     49 c     87 g     40 t
ORIGIN

  Query Match             6.0%;  Score 18;  DB 7;  Length 226;
  Best Local Similarity 100.0%;  Pred. No. 2.03e-05;
  Matches    18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```



```
Db    191 CTGGACATTCTGGGGAAC 208
          ||||||||||||||||||
Qy    143 CTGGACATTCTGGGGAAC 160


RESULT   13
LOCUS       T09208        265 bp    mRNA            EST       03-AUG-1993
DEFINITION  EST07101 Homo sapiens cDNA clone HIBBQ63 5' end.
ACCESSION   T09208
NID         g390236
KEYWORDS    EST.
SOURCE      Human clone=HIBBQ63 library=Infant brain, Bento Soares primer=M13
            Reverse.
  ORGANISM  Homo sapiens
            Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 265)
  AUTHORS   Adams,M.D., Soares,M.B., Kerlavage,A.R., Fields,C. and Venter,J.C.
  TITLE     Rapid cDNA sequencing (expressed sequence tags) from a
            directionally cloned human infant brain cDNA library
  JOURNAL   Nature Genet. 4, 373-380 (1993)
COMMENT
            Contact: Adams, MD
            The Institute for Genomic Research
            932 Clopper Road, Gaithersburg, MD 20878
            Tel: 3018699056
            Fax: 3018699423
            Email: mdadams@tigr.org.
FEATURES             Location/Qualifiers
     source          1. .265
                     /organism="Homo sapiens"
                     /clone="HIBBQ63"
BASE COUNT       49 a     80 c     96 g     34 t      6 others
ORIGIN

  Query Match             6.0%;  Score 18;  DB 5;  Length 265;
  Best Local Similarity 100.0%;  Pred. No. 2.03e-05;
  Matches    18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    117 CTGGACATTCTGGGGAAC 134
          ||||||||||||||||||
Qy    143 CTGGACATTCTGGGGAAC 160


RESULT   14
LOCUS       AA644289      265 bp    mRNA            EST       02-MAR-1998
DEFINITION  af63f04.s1 Soares NhHMPu S1 Homo sapiens cDNA clone 1046719 3',
            mRNA sequence.
ACCESSION   AA644289
NID         g2569507
KEYWORDS    EST.
```

```
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 265)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
            Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
            Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 176  Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 250.
FEATURES             Location/Qualifiers
     source          1. .265
                     /organism="Homo sapiens"
                     /note="Organ: mixed (see below); Vector: pT7T3D-Pac
                     (Pharmacia) with a modified polylinker; Site_1: Not I;
                     Site_2: Eco RI; Equal amounts of plasmid DNA from three
                     normalized libraries (melanocyte 2NbHM, pregnant uterus
                     NbHPU, and fetal heart NbHHL19W) were mixed, and ss circles
                     were made in vitro. Following HAP purification, this DNA
                     was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from pools of
                     5,000 clones made from the same 3 libraries. The pools
                     consisted of I.M.A.G.E. clones 260232-265223,
                     340488-345479, and 484488-489479."
                     /db_xref="taxon:9606"
                     /clone="1046719"
                     /clone_lib="Soares NhHMPu S1"
                     /tissue_type="Pooled human melanocyte, fetal heart, and
                     pregnant uterus"
                     /lab_host="DH10B"
BASE COUNT       55 a     76 c     78 g     56 t
ORIGIN

  Query Match            6.0%;  Score 18;  DB 13;  Length 265;
  Best Local Similarity  90.0%;  Pred. No. 2.03e-05;
  Matches   18;  Conservative    0;  Mismatches   2;  Indels   0;  Gaps   0;

Db      92 TTGCAGCGTCCTGTGGTTTG 111
```



```
            ||||||| |||||||||| |
Cp     266 TTGCAGCNTCCTGTGGTTNG 247


RESULT   15
LOCUS       AA716959      303 bp    mRNA            EST       29-DEC-1997
DEFINITION  vu69f05.r1 Stratagene mouse skin (#937313) Mus musculus cDNA clone
            1196673 5' similar to TR:Q14318 Q14318 FK-506 BINDING PROTEIN
            HOMOLOGUE. ;, mRNA sequence.
ACCESSION   AA716959
NID         g2729233
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 303)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:643769
            Seq primer: -28m13 revl ET from Amersham
            High quality sequence stop: 248.
FEATURES             Location/Qualifiers
     source          1. .303
                     /organism="Mus musculus"
                     /strain="C57BL/6"
                     /note="Organ: skin; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. Whole skin from 11 week old C57BL/6 female mice.
                     Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'
                     adaptor sequence: 5' GAATTCGGCACGAG 3' -3' adaptor
                     sequence: 5' CTCGAGTTTTTTTTTTTTTTTTT 3'"
                     /db_xref="taxon:10090"
                     /clone="1196673"
                     /clone_lib="Stratagene mouse skin (#937313)"
                     /sex="females"
```



```
                     /tissue_type="whole skin"
                     /dev_stage="11 weeks old"
                     /lab_host="SOLR (kanamycin resistant)"
BASE COUNT       63 a     91 c    101 g     48 t
ORIGIN

  Query Match            6.0%;  Score 18;  DB 11;  Length 303;
  Best Local Similarity 100.0%;  Pred. No. 2.03e-05;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     107 CTGGACATTCTGGGGAAC 124
           ||||||||||||||||||
Qy     143 CTGGACATTCTGGGGAAC 160


RESULT   16
LOCUS       AA599839      303 bp    mRNA            EST       02-MAR-1998
DEFINITION  ag33e05.s1 Jia bone marrow stroma Homo sapiens cDNA clone 1091360
            3', mRNA sequence.
ACCESSION   AA599839
NID         g2433464
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 303)
  AUTHORS   Jia,L., Wilkin,D., Bittner,M., Robey,P., Young,M., Yamada,Y.,
            Krizman,D., Liotta,L., Bonner,R., Schuler,G., Boguski,M.,
            Powell,J., Lennon,G., Roodman,D., Hotchkiss,R., Meltzer,P.,
            Trent,J., Hillier,L., Allen,M., Bowles,L., Geisel,S., Kucaba,T.,
            Marra,M., Martin,J., Steptoe,M., Tan,F., Theising,B., Bowers,Y.,
            Wylie,T., Waterston,R., Wilson,R. and Francomano,C.
  TITLE     WashU-MGB/NHGRI EST Project
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Wilson RK / Jia L
            WashU-MGB/NHGRI EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1551  Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 257.
FEATURES             Location/Qualifiers
     source          1. .303
                     /organism="Homo sapiens"
```



```
                     /note="Vector: pBluescript; Site_1: EcoRI; Site_2: XhoI;
                     mRNA made from human bone marrow stroma, cDNA made by
                     oligo-dT priming. Directionally cloned. Size-selected for
                     average insert size >0.5 kb. Library supplied by Dr. Libin
                     Jia (NHGRI). "
                     /db_xref="taxon:9606"
                     /clone="1091360"
                     /clone_lib="Jia bone marrow stroma"
                     /sex="mixed"
                     /tissue_type="bone marrow stroma"
                     /dev_stage="mixed"
                     /lab_host="XL1-Blue MRF'/SOLR"
BASE COUNT       61 a     88 c     93 g     61 t
ORIGIN

  Query Match            6.0%;  Score 18;  DB 13;  Length 303;
  Best Local Similarity  90.0%;  Pred. No. 2.03e-05;
  Matches   18;  Conservative    0;  Mismatches   2;  Indels   0;  Gaps   0;

Db      60 TTGCAGCGTCCTGTGGTTTG 79
           ||||||| |||||||||| |
Cp     266 TTGCAGCNTCCTGTGGTTNG 247


RESULT   17
LOCUS       HSC27F061     309 bp    RNA             EST       21-SEP-1995
DEFINITION  H. sapiens partial cDNA sequence; clone c-27f06, mRNA sequence.
ACCESSION   Z44717
NID         g573876
KEYWORDS    EST; partial cDNA sequence; transcribed sequence fragment.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 309)
  AUTHORS   Genexpress.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-OCT-1994) Genethon, B.P. 60, 91002 Evry Cedex France
            and Genetique Moleculaire et Biologie du developpement, CNRS UPR420
            B.P. 8, 94801 Villejuif Cedex France.E-mail: genexpress@genethon.fr
REFERENCE   2  (bases 1 to 309)
  AUTHORS   Genexpress.
  TITLE     The Genexpress cDNA program
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 309)
  AUTHORS   Auffray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C.,
            Devignes,M.D., Duprat,S., Houlgatte,R., Jumeau,M.N., Lamy,B.,
            Lorenzo,F., Mitchell,H., Mariage-Samson,R., Pietu,G., Pouliot,Y.,
            Sebastiani-Kabaktchis,C. and Tessier,A.
  TITLE     IMAGE: molecular integration of the analysis of the human genome
            and its expression
```

```
  JOURNAL   C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
  MEDLINE   95277534
COMMENT     Clone library from B.Soares, Psychiatry Dept. Columbia University
            USA;

            Cloning_method: total mRNA was oligo-(dT) primed and directionally
            cloned 5' -> 3' into the HindIII -> NotI sites of the lafmid BA
            vector;
            Sequencing_method: single read, full automatic;
            Primer: M13_reverse
            cDNA sequence colinear to mRNA
            Stretch_removed: nothing
            Normalization_method: Bento Soares, P.N.A.S in press;
            Genexpress_library_idt: C;
            Genexpress_sequence_idt: y1c-27f06;

            No significant homology found with :
            genbank release 81 swissprot release 28.
FEATURES            Location/Qualifiers
     source         1. .309
                    /organism="Homo sapiens"
                    /isolate="muscular atrophy patient"
                    /db_xref="taxon:9606"
                    /dev_stage="3 months old"
                    /tissue_type="total brain"
                    /clone_lib="normalized infant brain cDNA"
                    /sex="Female"
BASE COUNT       56 a     87 c    112 g     47 t      7 others
ORIGIN

  Query Match           6.0%;  Score 18;  DB 19;  Length 309;
  Best Local Similarity 100.0%;  Pred. No. 2.03e-05;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    130 CTGGACATTCTGGGGAAC 147
          ||||||||||||||||||
Qy    143 CTGGACATTCTGGGGAAC 160


RESULT   18
LOCUS       W11728         312 bp    mRNA            EST       02-OCT-1997
DEFINITION  mb17e04.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone 329694
            5', mRNA sequence.
ACCESSION   W11728
NID         g1286031
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 312)
```



```
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI: 211094
            Seq primer: mob.REGA+ET
            High quality sequence stop: 293.
FEATURES            Location/Qualifiers
     source         1. .312
                    /organism="Mus musculus"
                    /note="Vector: pT7T3D (Pharmacia) with a modified
                    polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                    was primed with a Not I - oligo(dT) primer [5'
                    TGTTACCAATCTGAAGTGGGAGCGGCCGCATTTTTTTTTTTTTTTTTT 3'],
                    double-stranded cDNA was size selected, ligated to Eco RI
                    adapters (Pharmacia), digested with Not I and cloned into
                    the Not I and Eco RI sites of a modified pT7T3 vector
                    (Pharmacia). Library went through one round of
                    normalization to a Cot = 5. Library constructed by Bento
                    Soares and M.Fatima Bonaldo. RNA was kindly provided by
                    Dr. Minoru Ko (Wayne State University)."
                    /db_xref="taxon:10090"
                    /clone="329694"
                    /clone_lib="Soares mouse p3NMF19.5"
                    /dev_stage="19.5 dpc total fetus"
                    /lab_host="DH10B (ampicillin resistant)"
BASE COUNT       49 a    101 c     93 g     69 t
ORIGIN

  Query Match           6.0%;  Score 18;  DB 10;  Length 312;
  Best Local Similarity 100.0%;  Pred. No. 2.03e-05;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    174 GTTCCCCAGAATGTCCAG 191
          ||||||||||||||||||
Cp    160 GTTCCCCAGAATGTCCAG 143
```



```
RESULT   19
LOCUS       AA105531       340 bp    mRNA            EST       30-OCT-1996
DEFINITION  mo57g06.r1 Life Tech mouse embryo 8 5dpc 10664019 Mus musculus cDNA
            clone 557722 5' similar to TR:G965470 G965470 FX-506 BINDING
            PROTEIN HOMOLOGUE. ;, mRNA sequence.
ACCESSION   AA105531
NID         g1654500
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Mus.
REFERENCE   1  (bases 1 to 340)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:338514
            Possible reversed clone: similarity on wrong strand
            Seq primer: -28M13 rev1 from Amersham
            High quality sequence stop: 157.
FEATURES            Location/Qualifiers
     source         1. .340
                    /organism="Mus musculus"
                    /strain="C57BL/6J"
                    /note="Vector: pCMV-SPORT2; Site_1: SalI; Site_2: NotI;
                    Cloned unidirectionally.  Primer: Oligo dT.  8.5dpc
                    embryos. pCMV-SPORT2 vector."
                    /db_xref="taxon:10090"
                    /clone="557722"
                    /clone_lib="Life Tech mouse embryo 8 5dpc 10664019"
                    /dev_stage="8.5dpc embryos"
                    /lab_host="DH10B"
     mRNA           <1. .>340
BASE COUNT       66 a     98 c    118 g     58 t
ORIGIN
```



```
  Query Match           6.0%;  Score 18;  DB 22;  Length 340;
  Best Local Similarity 100.0%;  Pred. No. 2.03e-05;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    287 CTGGACATTCTGGGGAAC 304
          ||||||||||||||||||
Qy    143 CTGGACATTCTGGGGAAC 160


RESULT   20
LOCUS       AA286324       349 bp    mRNA            EST       09-APR-1997
DEFINITION  va28c05.r1 GuayWoodford Beier mouse kidney day 7 Mus musculus cDNA
            clone 732680 5' similar to TR:G965470 G965470 FX-506 BINDING
            PROTEIN HOMOLOGUE. ;, mRNA sequence.
ACCESSION   AA286324
NID         g1932437
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae;
            Murinae; Mus.
REFERENCE   1  (bases 1 to 349)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:449728
            High quality sequence stop: 279.
FEATURES            Location/Qualifiers
     source         1. .349
                    /organism="Mus musculus"
                    /strain="C57BL/6J"
                    /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                    XhoI; Cloned unidirectionally.  Primer: Oligo dT. Average
                    insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor
                    sequence: 5' GAATTCGGCACGAG 3' -3' adaptor sequence: 5'
                    CTCGAGTTTTTTTTTTTTTTTTT 3' Library provided by Lisa
```

```
                     Guay-Woodford."
                     /db_xref="taxon:10090"
                     /clone="732680"
                     /clone_lib="GuayWoodford Beier mouse kidney day 7"
                     /dev_stage="juvenile (7 days old)"
                     /lab_host="SOLR (kanamycin resistant)"
     mRNA            <1. .>349
BASE COUNT       78 a     98 c    118 g     55 t
ORIGIN

  Query Match           6.0%;  Score 18;  DB 24;  Length 349;
  Best Local Similarity 100.0%;  Pred. No. 2.03e-05;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    122 CTGGACATTCTGGGGAAC 139
          ||||||||||||||||||
Qy    143 CTGGACATTCTGGGGAAC 160


RESULT   21
LOCUS       AA102860      363 bp    mRNA            EST       29-OCT-1996
DEFINITION  mo10e12.r1 Life Tech mouse embryo 10 5dpc 10665016 Mus musculus
            cDNA clone 553198 5' similar to TR:G965470 G965470 FK-506 BINDING
            PROTEIN HOMOLOGUE. ;, mRNA sequence.
ACCESSION   AA102860
NID         g1649142
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Mus.
REFERENCE   1  (bases 1 to 363)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:333990
```



```
            Seq primer: -28M13 rev1 from Amersham
            High quality sequence stop: 294.
FEATURES             Location/Qualifiers
     source          1. .363
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /note="Vector: pCMV-SPORT2; Site_1: SalI; Site_2: NotI;
                     Cloned unidirectionally.  Primer: Oligo dT.  10.5dpc
                     embryos. pCMV-SPORT2 vector."
                     /db_xref="taxon:10090"
                     /clone="553198"
                     /clone_lib="Life Tech mouse embryo 10 5dpc 10665016"
                     /dev_stage="10.5dpc embryos"
                     /lab_host="DH10B"
     mRNA            <1. .>363
BASE COUNT       76 a    111 c    113 g     63 t
ORIGIN

  Query Match           6.0%;  Score 18;  DB 22;  Length 363;
  Best Local Similarity 100.0%;  Pred. No. 2.03e-05;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      2 CTGGACATTCTGGGGAAC 19
          ||||||||||||||||||
Qy    143 CTGGACATTCTGGGGAAC 160


RESULT   22
LOCUS       AA921733      365 bp    mRNA            EST       23-JUN-1998
DEFINITION  om43e01.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
            IMAGE:1543800 3' similar to TR:Q14318 Q14318 FK-506 BINDING PROTEIN
            HOMOLOGUE. ;contains element MER22 repetitive element ;, mRNA
            sequence.
ACCESSION   AA921733
NID         g3069042
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 365)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
```



```
            Insert Length: 1120   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 289.
FEATURES             Location/Qualifiers
     source          1. .365
                     /organism="Homo sapiens"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from three normalized
                     libraries (fetal lung NbHL19W, testis NHT, and B-cell
                     NCI_CGAP_GCB1) were mixed, and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR-amplified cDNAs from pools of 5,000 clones made
                     from the same 3 libraries. The pools consisted of
                     I.M.A.G.E. clones 297480-302087, 682632-687239,
                     726408-728711, and 729096-731399. Subtraction by Bento
                     Soares and M. Fatima Bonaldo."
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1543800"
                     /clone_lib="Soares_NFL_T_GBC_S1"
                     /lab_host="DH10B"
BASE COUNT       75 a    105 c    135 g     50 t
ORIGIN

  Query Match           6.0%;  Score 18;  DB 15;  Length 365;
  Best Local Similarity 100.0%;  Pred. No. 2.03e-05;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    309 CTGGACATTCTGGGGAAC 326
          ||||||||||||||||||
Qy    143 CTGGACATTCTGGGGAAC 160


RESULT   23
LOCUS       AA152541      366 bp    mRNA            EST       11-FEB-1997
DEFINITION  mr85c07.r1 Stratagene mouse heart (#937316) Mus musculus cDNA clone
            604236 5' similar to TR:G965470 G965470 FK-506 BINDING PROTEIN
            HOMOLOGUE. ;, mRNA sequence.
ACCESSION   AA152541
NID         g1724459
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Mus.
REFERENCE   1  (bases 1 to 366)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
```



```
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:369668
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 363.
FEATURES             Location/Qualifiers
     source          1. .366
                     /organism="Mus musculus"
                     /strain="NIH/Swiss"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; Cloned unidirectionally.  Primer: Oligo dT. 93
                     pooled NIH/Swiss 13 day embryo hearts. Average insert
                     size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5'
                     GAATTCGGCACGAG 3' -3' adaptor sequence: 5'
                     CTCGAGTTTTTTTTTTTTTTT 3'"
                     /db_xref="taxon:10090"
                     /clone="604236"
                     /clone_lib="Stratagene mouse heart (#937316)"
                     /sex="pooled"
                     /dev_stage="13 day embryos"
                     /lab_host="SOLR (kanamycin resistant)"
     mRNA            <1. .>366
BASE COUNT       75 a    121 c    112 g     58 t
ORIGIN

  Query Match           6.0%;  Score 18;  DB 23;  Length 366;
  Best Local Similarity 100.0%;  Pred. No. 2.03e-05;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     66 CTGGACATTCTGGGGAAC 83
          ||||||||||||||||||
Qy    143 CTGGACATTCTGGGGAAC 160


RESULT   24
LOCUS       AA760296      368 bp    mRNA            EST       23-JAN-1998
DEFINITION  vv71b06.r1 Stratagene mouse skin (#937313) Mus musculus cDNA clone
            1227827 5', mRNA sequence.
ACCESSION   AA760296
```

```
NID         g2808090
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 368)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:653419
            Seq primer: -28ml3 rev1 ET from Amersham
            High quality sequence stop: 359.
FEATURES              Location/Qualifiers
     source           1. .368
                      /organism="Mus musculus"
                      /strain="C57BL/6"
                      /note="Organ: skin; Vector: pBluescript SK-; Site_1:
                      EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                      Oligo dT. Whole skin from 11 week old C57BL/6 female mice.
                      Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'
                      adaptor sequence: 5' GAATTCGGCACGAG 3' -3' adaptor
                      sequence: 5' CTCGAGTTTTTTTTTTTTTTTTT 3'"
                      /db_xref="taxon:10090"
                      /clone="1227027"
                      /clone_lib="Stratagene mouse skin (#937313)"
                      /sex="females"
                      /tissue_type="whole skin"
                      /dev_stage="11 weeks old"
                      /lab_host="SOLR (kanamycin resistant)"
BASE COUNT       77 a    110 c    120 g     61 t
ORIGIN

  Query Match           6.0%;  Score 18;  DB 12;  Length 368;
  Best Local Similarity 100.0%;  Pred. No. 2.03e-05;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```



```
Db     114 CTGGACATTCTGGGGAAC 131
           ||||||||||||||||||
Qy     143 CTGGACATTCTGGGGAAC 160


RESULT   25
LOCUS       AA656955       373 bp    mRNA            EST       04-NOV-1997
DEFINITION  vr25d02.r1 Barstead mouse myotubes MPLRB5 Mus musculus cDNA clone
            1121667 5' similar to TR:G965470 G965470 FK-506 BINDING PROTEIN
            HOMOLOGUE. ;, mRNA sequence.
ACCESSION   AA656955
NID         g2593109
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 373)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:611003
            Seq primer: -28ml3 rev2 ET from Amersham
            High quality sequence stop: 351.
FEATURES              Location/Qualifiers
     source           1. .373
                      /organism="Mus musculus"
                      /strain="C3H"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA
                      was primed with a Not I - oligo(dT) primer [5'
                      TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTTTT
                      3']; double-stranded cDNA was ligated to Eco RI adaptors
                      [AATTCGGATCCTTG], digested with Not I and cloned into the
                      Not I  and Eco RI sites of the modified pT7T3 vector.
                      Library constructed  by Bob Barstead.  The C2C12 cell line
                      (available from ATCC,  catalog # CRL-1772) differentiates
```



```
                      rapidly, forming contractile  myotubes and producing
                      characteristic muscle proteins."
                      /db_xref="taxon:10090"
                      /clone="1121667"
                      /clone_lib="Barstead mouse myotubes MPLRB5"
                      /cell_line="C2C12"
                      /lab_host="DH10B"
BASE COUNT       73 a    107 c    127 g     65 t      1 others
ORIGIN

  Query Match           6.0%;  Score 18;  DB 10;  Length 373;
  Best Local Similarity 100.0%;  Pred. No. 2.03e-05;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     267 CTGGACATTCTGGGGAAC 284
           ||||||||||||||||||
Qy     143 CTGGACATTCTGGGGAAC 160


RESULT   26
LOCUS       AA998836       375 bp    mRNA            EST       05-JUN-1998
DEFINITION  UI-R-C0-ii-g-01-0-UI.s1 UI-R-C0 Rattus norvegicus cDNA clone
            UI-R-C0-ii-g-01-0-UI 3', mRNA sequence.
ACCESSION   AA998836
NID         g3189487
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 375)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT
            Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            Oligo-dT track not found, Not I site shown in beginning of sequence
            is likely internal to the message. cDNA Library Preparation: M.
            Fatima Bonaldo, Ph.D. Clone distribution: clones will be available
            through Research Genetics
            Seq primer: M13 Forward.
FEATURES              Location/Qualifiers
     source           1. .375
```



```
                      /organism="Rattus norvegicus"
                      /strain="Sprague-Dawley"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-C0
                      library is a subtracted library derived from the UI-R-A1
                      and UI-R-E1 libraries. The UI-R-A1 library consisted of a
                      mixture of individually tagged normalized libraries
                      constructed from rat  placenta, adult lung, brain, liver,
                      kidney, heart, spleen, ovary,  and muscle. The UI-R-E1
                      library  consisted of a mixture of individually tagged
                      normalized libraries constructed from 8, 12 and 18-day
                      embryo. The tag is a string of 3-5 nucleotides present
                      between the Not I  site and the oligo-dT track which
                      allows identification of the library of origin of a clone
                      within the mixture. The subtracted library (UI-R-C0) was
                      constructed as follows: PCR amplified cDNA inserts from a
                      pool of UI-R-A1 and UI-R-E1 clones from which 3' ESTs had
                      been derived was used as a driver in a hybridization with
                      the pooled UI-R-A1 and UI-R-E1 library in the form of
                      single-stranded circles.  The remaining single-stranded
                      circles (subtracted library) was purified by
                      hydroxyapatite column chromatography, converted to
                      double-stranded circles and electroporated into DH10B
                      bacteria (Life Technologies) to generate the UI-R-C0
                      library. This procedure has been previously described
                      (Bonaldo, Lennon and Soares, Genome Research 6: 791-806,
                      1996)"
                      /db_xref="taxon:10116"
                      /clone="UI-R-C0-ii-g-01-0-UI"
                      /clone_lib="UI-R-C0"
                      /dev_stage="adult"
                      /lab_host="DH10B (Life Technologies)"
BASE COUNT       71 a    107 c    136 g     61 t
ORIGIN

  Query Match           6.0%;  Score 18;  DB 15;  Length 375;
  Best Local Similarity 100.0%;  Pred. No. 2.03e-05;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     332 CTGGACATTCTGGGGAAC 349
           ||||||||||||||||||
Qy     143 CTGGACATTCTGGGGAAC 160


RESULT   27
LOCUS       M62106         390 bp    mRNA            EST       26-MAY-1992
DEFINITION  EST00165 Homo sapiens cDNA clone HHCJ35.
ACCESSION   M62106
NID         g272355
KEYWORDS    EST.
SOURCE      human clone=HHCJ35 library=Hippocampus, Stratagene (cat. #936205)
```

```
            vector=lambdaZAP-II primer=M13 Forward Female, 2 years; oligo-dT +
            random primed cDNA synthesis; lambdaZAP-II vector, 1.0kb average
            insert size.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 390)
  AUTHORS   Adams,M.D., Kelley,J.M., Gocayne,J.D., Dubnick,M.,
            Polymeropoulos,M.H., Xiao,H., Merril,C.R., Wu,A., Olde,B.,
            Moreno,R.F. and et al.
  TITLE     Complementary DNA sequencing: expressed sequence tags and human
            genome project
  JOURNAL   Science 252 (5013), 1651-1656 (1991)
  MEDLINE   91262645
COMMENT
            Map: Chromosome 15
            Contact: Kerlavage AR
            The Institute for Genomic Research
            932 Clopper Road,Gaithersburg,MD 20878
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org.
FEATURES              Location/Qualifiers
     source           1. .390
                      /organism="Homo sapiens"
                      /clone="HHCJ35"
     mRNA             <1. .>390
                      /gene="D15S134E"
     gene             <1. .>390
                      /gene="D15S134E"
BASE COUNT       72 a    113 c    140 g     62 t      3 others
ORIGIN

  Query Match           6.0%;  Score 18;  DB 16;  Length 390;
  Best Local Similarity 100.0%;  Pred. No. 2.03e-05;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    151 CTGGACATTCTGGGGAAC 168
          ||||||||||||||||||
Qy    143 CTGGACATTCTGGGGAAC 160


RESULT   28
LOCUS       AA838853      394 bp    mRNA            EST       27-FEB-1998
DEFINITION  vw78a10.r1 Stratagene mouse heart (#937316) Mus musculus cDNA clone
            1261050 5' similar to TR:Q14318 Q14318 FK-506 BINDING PROTEIN
            HOMOLOGUE. ;, mRNA sequence.
ACCESSION   AA838853
NID         g2914948
```



```
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 394)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:663602
            Seq primer: -28m13 revl ET from Amersham
            High quality sequence stop: 359.
FEATURES              Location/Qualifiers
     source           1. .394
                      /organism="Mus musculus"
                      /strain="NIH/Swiss"
                      /note="Organ: heart; Vector: pBluescript SK-; Site_1:
                      EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                      Oligo dT. 93 pooled NIH/Swiss 13 day embryo hearts.
                      Average insert size: 1.0 kb; Uni-EAP XR Vector; -5'
                      adaptor sequence: 5' GAATTCGGCACGAG 3' -3' adaptor
                      sequence: 5' CTCGAGTTTTTTTTTTTTTTTTT 3'"
                      /db_xref="taxon:10090"
                      /clone="1261050"
                      /clone_lib="Stratagene mouse heart (#937316)"
                      /sex="pooled"
                      /tissue_type="heart"
                      /dev_stage="13 day embryos"
                      /lab_host="SOLR (kanamycin resistant)"
BASE COUNT       77 a    114 c    135 g     67 t      1 others
ORIGIN

  Query Match           6.0%;  Score 18;  DB 13;  Length 394;
  Best Local Similarity 100.0%;  Pred. No. 2.03e-05;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    274 CTGGACATTCTGGGGAAC 291
```



```
          ||||||||||||||||||
Qy    143 CTGGACATTCTGGGGAAC 160


RESULT   29
LOCUS       AA599463      395 bp    mRNA            EST       02-MAR-1998
DEFINITION  ag26d01.s1 Jia bone marrow stroma Homo sapiens cDNA clone 1090657
            3', mRNA sequence.
ACCESSION   AA599463
NID         g2433088
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 395)
  AUTHORS   Jia,L., Wilkin,D., Bittner,M., Robey,P., Young,M., Yamada,Y.,
            Krizman,D., Liotta,L., Bonner,R., Schuler,G., Boguski,M.,
            Powell,J., Lennon,G., Roodman,D., Hotchkiss,R., Meltzer,P.,
            Trent,J., Hillier,L., Allen,M., Bowles,L., Geisel,S., Kucaba,T.,
            Marra,M., Martin,J., Steptoe,M., Tan,F., Theising,B., Bowers,Y.,
            Wylie,T., Waterston,R., Wilson,R. and Prancomano,C.
  TITLE     WashU-MGB/NHGRI EST Project
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Wilson RK / Jia L
            WashU-MGB/NHGRI EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 3603  Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 311.
FEATURES              Location/Qualifiers
     source           1. .395
                      /organism="Homo sapiens"
                      /note="Vector: pBluescript; Site_1: EcoRI; Site_2: XhoI;
                      mRNA made from human bone marrow stroma, cDNA made by
                      oligo-dT priming. Directionally cloned. Size-selected for
                      average insert size >0.5 kb. Library supplied by Dr. Libin
                      Jia (NHGRI)."
                      /db_xref="taxon:9606"
                      /clone="1090657"
                      /clone_lib="Jia bone marrow stroma"
                      /sex="mixed"
                      /tissue_type="bone marrow stroma"
                      /dev_stage="mixed"
```



```
                      /lab_host="XL1-Blue MRF'/SOLR"
BASE COUNT       84 a    106 c    124 g     81 t
ORIGIN

  Query Match           6.0%;  Score 18;  DB 13;  Length 395;
  Best Local Similarity 90.0%;  Pred. No. 2.03e-05;
  Matches   18;  Conservative    0;  Mismatches   2;  Indels   0;  Gaps   0;

Db     82 TTGCAGCGTCCTGTGGTTTG 101
          ||||||| |||||||||| |
Cp    266 TTGCAGCNTCCTGTGGTTNG 247


RESULT   30
LOCUS       AI139353      400 bp    mRNA            EST       23-SEP-1998
DEFINITION  qc25e04.x1 Soares_pregnant_uterus_NbHPU Homo sapiens cDNA clone
            IMAGE:1710654 3', mRNA sequence.
ACCESSION   AI139353
NID         g3645325
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 400)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40m13 fwd. ET from Amersham.
FEATURES              Location/Qualifiers
     source           1. .400
                      /organism="Homo sapiens"
                      /note="Organ: uterus; Vector: pT7T3-Pac; Site_1: Not I;
                      Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                      oligo(dT) primer [5'
                      AACTGGAAGAATTCGCGGCCGCCTTTTTTTTTTTTTTTTT 3'],
                      double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Not I and cloned into the Not I
                      and Eco RI sites of the modified pT7T3 vector. Library
                      went through one round of normalization.  Library
                      constructed by M. Fatima Bonaldo."
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1710654"
                      /clone_lib="Soares_pregnant_uterus_NbHPU"
```

```
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
BASE COUNT       84 a    109 c    124 g     83 t
ORIGIN

  Query Match           6.0%;  Score 18;  DB 18;  Length 400;
  Best Local Similarity  90.0%;  Pred. No. 2.03e-05;
  Matches   18;  Conservative    0;  Mismatches   2;  Indels   0;  Gaps   0;

Db      84 TTGCAGCGTCCTGTGGTTTG 103
           ||||||| |||||||||| |
Cp     266 TTGCAGCNTCCTGTGGTTNG 247


RESULT   31
LOCUS       AA755779      403 bp    mRNA            EST       21-JAN-1998
DEFINITION  vv35c01.r1 Stratagene mouse heart (#937316) Mus musculus cDNA clone
            1224384 5', mRNA sequence.
ACCESSION   AA755779
NID         g2802977
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 403)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:649976
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 370.
FEATURES              Location/Qualifiers
     source           1. .403
                      /organism="Mus musculus"
                      /strain="NIH/Swiss"
```



```
                     /note="Organ: heart; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. 93 pooled NIH/Swiss 13 day embryo hearts.
                     Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'
                     adaptor sequence: 5' GAATTCGGCACGAG 3' -3' adaptor
                     sequence: 5' CTCGAGTTTTTTTTTTTTTTTTT 3'"
                     /db_xref="taxon:10090"
                     /clone="1224384"
                     /clone_lib="Stratagene mouse heart (#937316)"
                     /sex="pooled"
                     /tissue_type="heart"
                     /dev_stage="13 day embryos"
                     /lab_host="SOLR (kanamycin resistant)"
BASE COUNT       78 a    118 c    139 g     68 t
ORIGIN

  Query Match           6.0%;  Score 18;  DB 12;  Length 403;
  Best Local Similarity 100.0%;  Pred. No. 2.03e-05;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     281 CTGGACATTCTGGGGAAC 298
           ||||||||||||||||||
Qy     143 CTGGACATTCTGGGGAAC 160


RESULT   32
LOCUS       R11980        411 bp    mRNA            EST       11-APR-1995
DEFINITION  yf51b04.r1 Homo sapiens cDNA clone 25441 5'.
ACCESSION   R11980
NID         g764715
KEYWORDS    EST.
SOURCE      human clone=25441 library=Soares infant brain 1NIB vector=Lafmid BA
            host=DH10B (ampicillin resistant) primer=M13RP1 RsiteI=Not I
            RsiteI=Hind III Whole brain from a 73 days post natal female. 1st
            strand cDNA was primed with a Not I - oligo(dT) primer [5'
            AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTT 3']; double-stranded
            cDNA was ligated to Hind III adaptors (Pharmacia), digested with
            Not I and directionally cloned into the Not I and Hind III sites of
            the Lafmid BA vector. Library went through one round of
            normalization. Library constructed by Bento Soares and M.Fatima
            Bonaldo.
  ORGANISM  Homo sapiens
            Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 411)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
```



```
  JOURNAL   Unpublished (1995)
COMMENT
            GDB: G00-397-788
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 398
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES              Location/Qualifiers
     source           1. .411
                      /organism="Homo sapiens"
                      /clone="25441"
BASE COUNT       77 a    121 c    135 g     70 t      8 others
ORIGIN

  Query Match           6.0%;  Score 18;  DB 5;  Length 411;
  Best Local Similarity 100.0%;  Pred. No. 2.03e-05;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     122 CTGGACATTCTGGGGAAC 139
           ||||||||||||||||||
Qy     143 CTGGACATTCTGGGGAAC 160


RESULT   33
LOCUS       AI123210      416 bp    mRNA            EST       03-SEP-1998
DEFINITION  oo17c01.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone
            IMAGE:1566432 3', mRNA sequence.
ACCESSION   AI123210
NID         g3538976
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 416)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
```



```
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 401.
FEATURES              Location/Qualifiers
     source           1. .416
                      /organism="Homo sapiens"
                      /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                      a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                      Equal amounts of plasmid DNA from five normalized
                      libraries were mixed, and ss circles were made in vitro.
                      Following HAP purification, this DNA was used as tracer in
                      a subtractive hybridization reaction. The driver was
                      PCR-amplified cDNAs from pools of 5,000 clones made from
                      the same 5 libraries. The pools consisted of the following
                      libraries and cloneIDs: Soares NbHSF pool 1:
                      309384-310919, 323208-325895 Soares Nb2HP pool 1:
                      145032-147335, 147720-148103, 148872-149255, 15002 -
                      150407, 151176-152327 Soares Nb2HF8-9W pool 1:
                      758280-760583, 772104-774407 Soares NbHPA pool 1:
                      304776-306311, 320136-322823, 326280-326663 Soares NbHOT
                      pool 1:  723720-726407, 739080-740999 Subtraction by Bento
                      Soares and M. Fatima Bonaldo."
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1566432"
                      /clone_lib="Soares_NSF_F8_9W_OT_PA_P_S1"
                      /lab_host="DH10B"
BASE COUNT       88 a    110 c    131 g     87 t
ORIGIN

  Query Match           6.0%;  Score 18;  DB 17;  Length 416;
  Best Local Similarity  90.0%;  Pred. No. 2.03e-05;
  Matches   18;  Conservative    0;  Mismatches   2;  Indels   0;  Gaps   0;

Db      81 TTGCAGCGTCCTGTGGTTTG 100
           ||||||| |||||||||| |
Cp     266 TTGCAGCNTCCTGTGGTTNG 247


RESULT   34
LOCUS       MUS68E12      421 bp    mRNA            EST       04-OCT-1996
DEFINITION  Mouse embryonal carcinoma F9 cell cDNA, 68E12, mRNA sequence.
ACCESSION   D21541
NID         g618676
KEYWORDS    EST; EST(expressed sequence tag).
SOURCE      Mus musculus cell_line:mouse embryonal carcinoma cell line F9 cDNA
            to mRNA.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae;
            Murinae; Mus.
REFERENCE   1  (bases 1 to 421)
```

```
  AUTHORS   Nishiguchi,S., Joh,T., Horie,K., Zou,Z., Yasunaga,T. and Shimada,K.
  TITLE     A survey of genes expressed in undifferentiated mouse embryonal
            carcinoma F9 cells: characterization of low-abundance mRNAs
  JOURNAL   J. Biochem. 116 (1), 128-139 (1994)
  MEDLINE   95096008
REFERENCE   2  (sites)
  AUTHORS   Nishiguchi,S., Sakuma,R., Nomura,M., Zou,Z., Jearanaisilavong,J.,
            Joh,T., Yasunaga,T. and Shimada,K.
  TITLE     A catalogue of genes in mouse embryonal carcinoma F9 cells
            identified with expressed sequence tags
  JOURNAL   J. Biochem. 119 (4), 749-767 (1996)
  MEDLINE   96337530
REFERENCE   3  (bases 1 to 421)
  AUTHORS   Shimada,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-OCT-1993) to the DDBJ/EMBL/GenBank databases.
            Kazunori Shimada, Research Institute for Microbial Diseases, Osaka
            University, Department of Medical Genetics; 3-1, Yamadaoka, Suita,
            Osaka 565, Japan (Tel:06-877-5111(ex.3293), Fax:06-875-0274)
FEATURES             Location/Qualifiers
     source          1. .421
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /cell_line="mouse embryonal carcinoma cell line F9"
BASE COUNT       93 a    124 c    131 g     71 t      2 others
ORIGIN

  Query Match           6.0%;  Score 18;  DB 7;  Length 421;
  Best Local Similarity 100.0%;  Pred. No. 2.03e-05;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    114 CTGGACATTCTGGGGAAC 131
        ||||||||||||||||||
Qy    143 CTGGACATTCTGGGGAAC 160


RESULT   35
LOCUS       T80918        426 bp    mRNA            EST       15-MAR-1995
DEFINITION  yd25g05.r1 Homo sapiens cDNA clone 109304 5'.
ACCESSION   T80918
NID         g703803
KEYWORDS    EST.
SOURCE      human clone=109304 library=Soares fetal liver spleen 1NFLS
            vector=pT7T3D (Pharmacia) with a modified polylinker host=DH10B
            (ampicillin resistant) primer=M13RP1 Rsitel=Pac I Rsite2=Eco RI
            Liver and spleen from a 20 week-post conception male fetus. 1st
            strand cDNA was primed with a Pac I - oligo(dT) primer [5'
            AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTT 3'], double-stranded
            cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac
            I and cloned into the Pac I and Eco RI sites of the modified pT7T3
            vector.  Library went through one round of normalization. Library
```



```
            constructed by Bento Soares and M.Fatima Bonaldo.
  ORGANISM  Homo sapiens
            Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 426)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     Other_ESTs: yd25g05.s1
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 286
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES             Location/Qualifiers
     source          1. .426
                     /organism="Homo sapiens"
                     /clone="109304"
BASE COUNT       86 a    109 c    111 g    117 t      3 others
ORIGIN

  Query Match           6.0%;  Score 18;  DB 5;  Length 426;
  Best Local Similarity 90.0%;  Pred. No. 2.03e-05;
  Matches   18;  Conservative    0;  Mismatches   2;  Indels   0;  Gaps   0;

Db    401 AAAGAGAAGTGGGTNGGCAC 420
          |||||||||| ||| |||||
Qy    207 AAAGAGAAGTNGGTGGGCAC 226


RESULT   36
LOCUS       AA065589      432 bp    mRNA            EST       03-FEB-1997
DEFINITION  ml71d07.r1 Stratagene mouse kidney (#937315) Mus musculus cDNA
            clone 517453 5' similar to TR:G965470 G965470 FK-506 BINDING
            PROTEIN HOMOLOGUE. ;, mRNA sequence.
ACCESSION   AA065589
NID         g1563013
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
```



```
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Mus.
REFERENCE   1  (bases 1 to 432)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:311301
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 423.
FEATURES             Location/Qualifiers
     source          1. .432
                     /organism="Mus musculus"
                     /strain="C57/B16"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; Cloned unidirectionally.  Primer: Oligo dT. Average
                     insert size: 1.0 kb; Uni-ZAP XR Vector; ~5' adaptor
                     sequence: 5' GAATTCGGCACGAG 3' -3' adaptor sequence: 5'
                     CTCGAGTTTTTTTTTTTTTTTT 3'"
                     /db_xref="taxon:10090"
                     /clone="517453"
                     /clone_lib="Stratagene mouse kidney (#937315)"
                     /sex="females"
                     /dev_stage="4 weeks"
                     /lab_host="SOLR (kanamycin resistant)"
     mRNA            <1. .>432
BASE COUNT       91 a    140 c    134 g     67 t
ORIGIN

  Query Match           6.0%;  Score 18;  DB 23;  Length 432;
  Best Local Similarity 100.0%;  Pred. No. 2.03e-05;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    127 CTGGACATTCTGGGGAAC 144
        ||||||||||||||||||
Qy    143 CTGGACATTCTGGGGAAC 160
```



```
RESULT   37
LOCUS       W51994        434 bp    mRNA            EST       31-MAY-1996
DEFINITION  zc37g04.s1 Soares senescent fibroblasts NbHSF Homo sapiens cDNA
            clone 324534 3', mRNA sequence.
ACCESSION   W51994
NID         g1349871
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 434)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: mob.REGA+ET
            High quality sequence stop: 335.
FEATURES             Location/Qualifiers
     source          1. .434
                     /organism="Homo sapiens"
                     /note="Vector: pT7T3D (Pharmacia) with a modified
                     polylinker V_TYPE: phagemid; Site_1: Not I; Site_2: Eco
                     RI; TGTTACCAATCTGAAGTGGGAGCGGCCGCATTTTTTTTTTTTTTTTTT
                     3'], double-stranded cDNA was size selected, ligated to
                     Eco RI adapters (Pharmacia), digested with Not I and
                     cloned into the Not I and Eco RI sites of a modified pT7T3
                     vector (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by Bento
                     Soares and M.Fatima Bonaldo."
                     /db_xref="taxon:9606"
                     /clone="324534"
                     /clone_lib="Soares senescent fibroblasts NbHSF"
                     /tissue_type="senescent fibroblast"
                     /lab_host="DH10B (ampicillin resistant)"
     mRNA            complement(<1. .>434)
BASE COUNT       93 a    116 c    131 g     93 t      1 others
ORIGIN
```

```
  Query Match           6.0%;  Score 18;  DB 20;  Length 434;
  Best Local Similarity  90.0%;  Pred. No. 2.03e-05;
  Matches   18;  Conservative    0;  Mismatches   2;  Indels   0;  Gaps   0;

Db      86 TTGCAGCGTCCTGTGGTTTG 105
           ||||||| |||||||||| |
Cp     266 TTGCAGCNTCCTGTGGTTNG 247


RESULT   38
LOCUS       AI092100       435 bp    mRNA            EST       18-AUG-1998
DEFINITION  qa57h06.s1 Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGE:1690907
            3', mRNA sequence.
ACCESSION   AI092100
NID         g3431094
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 435)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 412.
FEATURES             Location/Qualifiers
     source          1. .435
                     /organism="Homo sapiens"
                     /note="Organ: mixed (see below); Vector: pT7T3D-Pac
                     (Pharmacia) with a modified polylinker; Site_1: Not I;
                     Site_2: Eco RI; Equal amounts of plasmid DNA from three
                     normalized libraries (melanocyte 2NbHM, pregnant uterus
                     NbHPU, and fetal heart NbHH19W) were mixed, and ss circles
                     were made in vitro. Following HAP purification, this DNA
                     was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from pools of
                     5,000 clones made from the same 3 libraries. The pools
                     consisted of I.M.A.G.E. clones 260232-265223,
                     340488-345479, and 484488-489479."
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1690907"
                     /clone_lib="Soares_NhHMPu_S1"
```



```
                     /tissue_type="Pooled human melanocyte, fetal heart, and
                     pregnant uterus"
                     /lab_host="DH10B"
BASE COUNT       93 a    116 c    134 g     92 t
ORIGIN

  Query Match           6.0%;  Score 18;  DB 17;  Length 435;
  Best Local Similarity  90.0%;  Pred. No. 2.03e-05;
  Matches   18;  Conservative    0;  Mismatches   2;  Indels   0;  Gaps   0;

Db      84 TTGCAGCGTCCTGTGGTTTG 103
           ||||||| |||||||||| |
Cp     266 TTGCAGCNTCCTGTGGTTNG 247


RESULT   39
LOCUS       W70917         439 bp    mRNA            EST       17-JUN-1996
DEFINITION  me23h04.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
            clone 388375 5', mRNA sequence.
ACCESSION   W70917
NID         g1379692
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;
            Murinae; Mus.
REFERENCE   1  (bases 1 to 439)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:240207
            Seq primer: mob.REGA+ET
            High quality sequence stop: 348.
FEATURES             Location/Qualifiers
     source          1. .439
                     /organism="Mus musculus"
```



```
                     /strain="C57BL/6J"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCGGCGAAATTTTTTTTTTTTTTTTTTTT
                     T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
                     14.5dpc embryos [total RNA provided by  Minoru Ko, Wayne
                     State Univ., from 2 ]; double-stranded cDNA was ligated to
                     Eco RI adaptors (Pharmacia), digested with Not I and
                     cloned into the Not I and Eco RI sites of the modified
                     pT7T3 vector. Library went through one round of
                     normalization, and was constructed by Bento Soares and
                     M.Fatima Bonaldo."
                     /db_xref="taxon:10090"
                     /clone="388375"
                     /clone_lib="Soares mouse embryo NbME13.5 14.5"
                     /sex="unknown"
                     /tissue_type="embryo"
                     /dev_stage="13.5-14.5dpc total fetus"
                     /lab_host="DH10B"
     mRNA            <1. .>439
BASE COUNT       92 a    129 c    137 g     81 t
ORIGIN

  Query Match           6.0%;  Score 18;  DB 20;  Length 439;
  Best Local Similarity 100.0%;  Pred. No. 2.03e-05;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      75 CTGGACATTCTGGGGAAC 92
           ||||||||||||||||||
Qy     143 CTGGACATTCTGGGGAAC 160


RESULT   40
LOCUS       AA115038       445 bp    mRNA            EST       14-MAY-1997
DEFINITION  zk89f08.r1 Soares pregnant uterus NbHPU Homo sapiens cDNA clone
            490023 5', mRNA sequence.
ACCESSION   AA115038
NID         g1670291
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 445)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
            Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     WashU-Merck EST Project
```



```
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Putative full length read
            The vector to vector length is 446
            Insert Length: 182    Std Error: 0.00
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 423.
FEATURES             Location/Qualifiers
     source          1. .445
                     /organism="Homo sapiens"
                     /note="Organ: uterus; Vector: pT7T3-Pac; Site_1: Not I;
                     Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                     oligo(dT) primer [5'
                     AACTGGAAGAATTCGCGGCCGCCTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization.  Library
                     constructed by M. Fatima Bonaldo."
                     /db_xref="taxon:9606"
                     /clone="490023"
                     /clone_lib="Soares pregnant uterus NbHPU"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
     mRNA            <1. .>445
                     /db_xref="GDB:3804610"
BASE COUNT       94 a    135 c    118 g     95 t      3 others
ORIGIN

  Query Match           6.0%;  Score 18;  DB 6;  Length 445;
  Best Local Similarity  90.0%;  Pred. No. 2.03e-05;
  Matches   18;  Conservative    0;  Mismatches   2;  Indels   0;  Gaps   0;

Db     342 CAAACCACAGGACGCTGCAA 361
           | |||||||||| |||||||
Qy     247 CNAACCACAGGANGCTGCAA 266


RESULT   41
LOCUS       AQ089629       448 bp    DNA             GSS       25-AUG-1998
DEFINITION  HS_3008_A1_D01_T7 CIT Approved Human Genomic Sperm Library D Homo
```

```
            sapiens genomic clone Plate=3008 Col=1 Row=G, genomic survey
            sequence.
ACCESSION   AQ089629
NID         g3458540
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 448)
  AUTHORS   Mahairas,G.G., Zackrone,K.D., Smith,T., Tipton,S., Schmidt,S.,
            Traicoff,R., Abajian,C., Blanchard,A., West,A. and Hood,L.E.
  TITLE     Construction of a Characterized Clone Resource for Genomic
            Sequencing: Generation and Preliminary Analysis of 20,000 Sequence
            Tagged Connectors
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3008  row: G  column: 1
            Class: BAC ends
            High quality sequence stop: 448.
FEATURES              Location/Qualifiers
     source           1. .448
                      /organism="Homo sapiens"
                      /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                      E-Coli DH10B"
                      /db_xref="taxon:9606"
                      /clone="Plate=3008 Col=1 Row=G"
                      /clone_lib="CIT Approved Human Genomic Sperm Library D"
                      /sex="male"
BASE COUNT       91 a    114 c    118 g    122 t      3 others
ORIGIN

  Query Match           6.0%;  Score 18;  DB 27;  Length 448;
  Best Local Similarity  90.0%;  Pred. No. 2.03e-05;
  Matches   18;  Conservative    0;  Mismatches   2;  Indels   0;  Gaps   0;

Db    141 CAAACCACAGGACGCTGCAA 160
          | ||||||||||| |||||||
Qy    247 CNAACCACAGGANGCTGCAA 266


RESULT   42
LOCUS       AA114966      458 bp    mRNA            EST       14-MAY-1997
```



```
DEFINITION  zk89f08.s1 Soares pregnant uterus NbHPU Homo sapiens cDNA clone
            490023 3', mRNA sequence.
ACCESSION   AA114966
NID         g1670181
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 458)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
            Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 182   Std Error: 0.00
            Seq primer: -40M13 fwd. from Amersham
            High quality sequence stop: 352.
FEATURES              Location/Qualifiers
     source           1. .458
                      /organism="Homo sapiens"
                      /note="Organ: uterus; Vector: pT7T3-Pac; Site_1: Not I;
                      Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                      oligo(dT) primer [5'
                      AACTGGAAGAATTCGCGGCCGCCCTTTTTTTTTTTTTTTT 3'],
                      double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Not I and cloned into the Not I
                      and Eco RI sites of the modified pT7T3 vector. Library
                      went through one round of normalization.  Library
                      constructed by M. Fatima Bonaldo."
                      /db_xref="taxon:9606"
                      /clone="490023"
                      /clone_lib="Soares pregnant uterus NbHPU"
                      /sex="female"
                      /dev_stage="adult"
                      /lab_host="DH10B"
     mRNA             complement(<1. .>458)
                      /db_xref="GDB:3804610"
BASE COUNT       96 a    124 c    138 g     99 t      1 others
```



```
ORIGIN

  Query Match           6.0%;  Score 18;  DB 6;  Length 458;
  Best Local Similarity  90.0%;  Pred. No. 2.03e-05;
  Matches   18;  Conservative    0;  Mismatches   2;  Indels   0;  Gaps   0;

Db     86 TTGCAGCGTCCTGTGGTTTG 105
          ||||||| |||||||||| |
Cp    266 TTGCAGCNTCCTGTGGTTNG 247


RESULT   43
LOCUS       AA669796      487 bp    mRNA            EST       20-NOV-1997
DEFINITION  ag36b12.s1 Jia bone marrow stroma Homo sapiens cDNA clone 1118879
            3', mRNA sequence.
ACCESSION   AA669796
NID         g2631295
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 487)
  AUTHORS   Jia,L., Wilkin,D., Bittner,M., Robey,P., Young,M., Yamada,Y.,
            Krizman,D., Liotta,L., Bonner,R., Schuler,G., Boguski,M.,
            Powell,J., Lennon,G., Roodman,D., Hotchkiss,R., Meltzer,P.,
            Trent,J., Hillier,L., Allen,M., Bowles,L., Geisel,S., Kucaba,T.,
            Marra,M., Martin,J., Steptoe,M., Tan,F., Theising,B., Bowers,Y.,
            Wylie,T., Waterston,R., Wilson,R. and Francomano,C.
  TITLE     WashU-MGB/NHGRI EST Project
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Wilson RK / Jia L
            WashU-MGB/NHGRI EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 446.
FEATURES              Location/Qualifiers
     source           1. .487
                      /organism="Homo sapiens"
                      /note="Vector: pBluescript; Site_1: EcoRI; Site_2: XhoI;
                      mRNA made from human bone marrow stroma, cDNA made by
                      oligo-dT priming. Directionally cloned. Size-selected for
                      average insert size >0.5 kb. Library supplied by Dr. Libin
                      Jia (NHGRI)."
```



```
                      /db_xref="taxon:9606"
                      /clone="1118879"
                      /clone_lib="Jia bone marrow stroma"
                      /sex="mixed"
                      /tissue_type="bone marrow stroma"
                      /dev_stage="mixed"
                      /lab_host="XL1-Blue MRF'/SOLR"
BASE COUNT      102 a    122 c    149 g    114 t
ORIGIN

  Query Match           6.0%;  Score 18;  DB 11;  Length 487;
  Best Local Similarity  90.0%;  Pred. No. 2.03e-05;
  Matches   18;  Conservative    0;  Mismatches   2;  Indels   0;  Gaps   0;

Db    107 TTGCAGCGTCCTGTGGTTTG 126
          ||||||| |||||||||| |
Cp    266 TTGCAGCNTCCTGTGGTTNG 247


RESULT   44
LOCUS       AA845481      499 bp    mRNA            EST       04-MAR-1998
DEFINITION  ai89d04.s1 Soares parathyroid tumor NbHPA Homo sapiens cDNA clone
            IMAGE:1387975 3' similar to TR:Q14318 Q14318 FK-506 BINDING PROTEIN
            HOMOLOGUE. ;contains element MER22 repetitive element ;, mRNA
            sequence.
ACCESSION   AA845481
NID         g2933240
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 499)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Possible reversed clone: similarity on wrong strand
            Possible reversed clone: polyT not found
```

```
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 451.
FEATURES              Location/Qualifiers
     source           1. .499
                      /organism="Homo sapiens"
                      /note="Organ: parathyroid gland; Vector: pT7T3D
                      (Pharmacia) with a modified polylinker; Site_1: Not I;
                      Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                      oligo(dT) primer
                      [5'-
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCACCAATTTTTTTTTTTTTTTTTTTTTT
                      T-3'], double-stranded cDNA was size selected, ligated to
                      Eco RI adapters (Pharmacia), digested with Not I and
                      cloned into the Not I and Eco RI sites of a modified pT7T3
                      vector (Pharmacia). Library went through one round of
                      normalisation to a Cot = 5. Library constructed by Bento
                      Soares and M.Fatima Bonaldo. RNA from sporadic parathyroid
                      adenomas was kindly provided by Dr. Stephen Marx, National
                      Institute of Diabetes and Digestive and Kidney Diseases,
                      NIH."
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1387975"
                      /clone_lib="Soares parathyroid tumor NbHPA"
                      /tissue_type="parathyroid tumor"
                      /dev_stage="adult"
                      /lab_host="DH10B (ampicillin resistant)"
BASE COUNT       94 a    144 c    183 g     78 t
ORIGIN

  Query Match             6.0%;  Score 18;  DB 13;  Length 499;
  Best Local Similarity 100.0%;  Pred. No. 2.03e-05;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    299 CTGGACATTCTGGGGAAC 316
          ||||||||||||||||||
Qy    143 CTGGACATTCTGGGGAAC 160


RESULT   45
LOCUS       AA848340      501 bp    mRNA            EST       30-APR-1998
DEFINITION  EST191100 Normalized rat kidney, Bento Soares Rattus sp. cDNA clone
            RKIAB15 5' end, mRNA sequence.
ACCESSION   AA848340
NID         g2935880
KEYWORDS    EST.
SOURCE      Rattus sp.
  ORGANISM  Rattus sp.
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 501)
  AUTHORS   Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
```



```
            Kerlavage,A.R. and Adams,M.D.
  TITLE     Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
            Gene Index
  JOURNAL   Unpublished (1998)
COMMENT
            Contact: Lee, NH
            ATCC
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
            Seq primer: M13 Reverse.
FEATURES              Location/Qualifiers
     source           1. .501
                      /organism="Rattus sp."
                      /note="Organ: kidney; Vector: pT7T3Pac; Site_1: EcoRI;
                      Site_2: NotI"
                      /db_xref="ATCC (inhost):2007328"
                      /db_xref="taxon:10118"
                      /clone="RKIAB15"
                      /clone_lib="Normalized rat kidney, Bento Soares"
BASE COUNT       90 a    143 c    174 g     94 t
ORIGIN

  Query Match             6.0%;  Score 18;  DB 14;  Length 501;
  Best Local Similarity 100.0%;  Pred. No. 2.03e-05;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    320 CTGGACATTCTGGGGAAC 337
          ||||||||||||||||||
Qy    143 CTGGACATTCTGGGGAAC 160


RESULT   46
LOCUS       AQ180720      506 bp    DNA             GSS       09-SEP-1998
DEFINITION  HS_3220_A2_B04_T7 CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3220 Col=8 Row=C, genomic survey
            sequence.
ACCESSION   AQ180720
NID         g3578087
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 506)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Furlong,J. and
            Shaker,R.
  TITLE     Construction of a Characterized Clone Resource for Genomic
            Sequencing
```



```
  JOURNAL   Unpublished (1998)
COMMENT
            Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3220  row: C  column: 8
            Class: BAC ends
            High quality sequence stop: 506.
FEATURES              Location/Qualifiers
     source           1. .506
                      /organism="Homo sapiens"
                      /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                      E-Coli DH10B"
                      /db_xref="taxon:9606"
                      /clone="Plate=3220 Col=8 Row=C"
                      /clone_lib="CIT Approved Human Genomic Sperm Library D"
                      /sex="male"
BASE COUNT      136 a    117 c     99 g    152 t      2 others
ORIGIN

  Query Match             6.0%;  Score 18;  DB 28;  Length 506;
  Best Local Similarity 100.0%;  Pred. No. 2.03e-05;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    160 ACTCTTAGGCCTGGCCCA 177
          ||||||||||||||||||
Cp     71 ACTCTTAGGCCTGGCCCA 54


RESULT   47
LOCUS       AI116452      507 bp    mRNA            EST       02-SEP-1998
DEFINITION  ue98d02.y1 Sugano mouse embryo mewa Mus musculus cDNA clone 1499139
            5' similar to TR:Q14318 Q14318 FK-506 BINDING PROTEIN HOMOLOGUE. ;,
            mRNA sequence.
ACCESSION   AI116452
NID         g3516776
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 507)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
```



```
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:936743
            Seq primer: custom primer used.
FEATURES              Location/Qualifiers
     source           1. .507
                      /organism="Mus musculus"
                      /strain="C57BL"
                      /note="Vector: pME18S-FL3; Site_1: DraIII (CACTGTGTG);
                      Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed
                      with an oligo(dT) primer  [ATGTGGCCTTTTTTTTTTTTTTTT];
                      double-stranded cDNA was ligated to a  DraIII adaptor
                      [TGTTGGCCTACTGG], digested and cloned into distinct DraIII
                      sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site
                      CACCATGTG). XhoI should be used to isolate the cDNA
                      insert.  Size selection was performed to exclude fragments
                      <1.5kb.  Library constructed by Dr. Sumio Sugano
                      (University of Tokyo Institute of Medical Science).
                      Custom primers  for sequencing: 5' end primer
                      CTTCTGCTCTAAAAGCTGCG and 3' end primer
                      CGACCTGCAGCTCGAGCACA."
                      /db_xref="taxon:10090"
                      /clone="1499139"
                      /clone_lib="Sugano mouse embryo mewa"
                      /dev_stage="embryo, 14 dpc"
                      /lab_host="DH10B"
BASE COUNT       97 a    152 c    173 g     84 t      1 others
ORIGIN

  Query Match             6.0%;  Score 18;  DB 17;  Length 507;
  Best Local Similarity 100.0%;  Pred. No. 2.03e-05;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    304 CTGGACATTCTGGGGAAC 321
          ||||||||||||||||||
Qy    143 CTGGACATTCTGGGGAAC 160


RESULT   48
LOCUS       W52834        530 bp    mRNA            EST       11-OCT-1996
```

```
DEFINITION  zc55e06.r1 Soares senescent fibroblasts NbHSF Homo sapiens cDNA
            clone 326242 5', mRNA sequence.
ACCESSION   W52834
NID         g1350258
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 530)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 398   Std Error: 0.00
            Seq primer: mob.REGA+ET
            High quality sequence stop: 457.
FEATURES             Location/Qualifiers
     source          1. .530
                     /organism="Homo sapiens"
                     /note="Vector: pT7T3D (Pharmacia) with a modified
                     polylinker V_TYPE: phagemid; Site_1: Not I; Site_2: Eco
                     RI; TGTTACCAATCTGAAGTGGGAGCGGCCGCATTTTTTTTTTTTTTTTTTT
                     3'], double-stranded cDNA was size selected, ligated to
                     Eco RI adapters (Pharmacia), digested with Not I and
                     cloned into the Not I and Eco RI sites of a modified pT7T3
                     vector (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by Bento
                     Soares and M.Fatima Bonaldo."
                     /db_xref="taxon:9606"
                     /clone="326242"
                     /clone_lib="Soares senescent fibroblasts NbHSF"
                     /tissue_type="senescent fibroblast"
                     /lab_host="DH10B (ampicillin resistant)"
     mRNA            <1. .>530
BASE COUNT      117 a    153 c    138 g    119 t      3 others
ORIGIN
```



```
  Query Match           6.0%;  Score 18;  DB 21;  Length 530;
  Best Local Similarity  90.0%;  Pred. No. 2.03e-05;
  Matches   18;  Conservative    0;  Mismatches   2;  Indels   0;  Gaps   0;

Db     414 CAAACCACAGGACGCTGCAA 433
           | |||||||||| |||||||
Qy     247 CNAACCACAGGANGCTGCAA 266


RESULT   49
LOCUS       W58425         540 bp    mRNA            EST       15-OCT-1996
DEFINITION  zd25e08.r1 Soares fetal heart NbHH19W Homo sapiens cDNA clone
            341702 5', mRNA sequence.
ACCESSION   W58425
NID         g1365207
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 540)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 752   Std Error: 0.00
            Seq primer: mob.REGA+ET
            High quality sequence stop: 399.
FEATURES             Location/Qualifiers
     source          1. .540
                     /organism="Homo sapiens"
                     /note="Organ: heart; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATCTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
```



```
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by
                     M.Fatima Bonaldo. This library was constructed from the
                     same fetus as the fetal lung library, Soares fetal lung
                     NbHL19W."
                     /db_xref="taxon:9606"
                     /clone="341702"
                     /clone_lib="Soares fetal heart NbHH19W"
                     /sex="unknown"
                     /dev_stage="19 weeks"
                     /lab_host="DH10B (ampicillin resistant)"
     mRNA            <1. .>540
BASE COUNT       88 a    180 c    152 g    111 t      9 others
ORIGIN

  Query Match           6.0%;  Score 18;  DB 21;  Length 540;
  Best Local Similarity 100.0%;  Pred. No. 2.03e-05;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     315 GTTCCCCAGAATGTCCAG 332
           ||||||||||||||||||
Cp     160 GTTCCCCAGAATGTCCAG 143


RESULT   50
LOCUS       AA523420       565 bp    mRNA            EST       20-AUG-1997
DEFINITION  ng30b04.s1 NCI_CGAP_Co3 Homo sapiens cDNA clone IMAGE:936271, mRNA
            sequence.
ACCESSION   AA523420
NID         g2264132
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 565)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck,
            M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arraying: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
```



```
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 466.
FEATURES             Location/Qualifiers
     source          1. .565
                     /organism="Homo sapiens"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from 12 pooled bulk tumor samples and primed
                     with a Not I - oligo(dT) primer. Double-stranded cDNA was
                     ligated to Eco RI adaptors (Pharmacia), digested with Not
                     I and cloned into the Not I and Eco RI sites of the
                     modified pT7T3 vector. Library went through one round of
                     normalization."
                     /db_xref="taxon:9606"
                     /clone="IMAGE:936271"
                     /clone_lib="NCI_CGAP_Co3"
                     /sex="pooled"
                     /tissue_type="colon"
                     /lab_host="DH10B"
     mRNA            <1. .>565
BASE COUNT      128 a    150 c    166 g    121 t
ORIGIN

  Query Match           6.0%;  Score 18;  DB 9;  Length 565;
  Best Local Similarity  90.0%;  Pred. No. 2.03e-05;
  Matches   18;  Conservative    0;  Mismatches   2;  Indels   0;  Gaps   0;

Db      91 TTGCAGGGTCCTGTGGTTTG 110
           |||||| |||||||||| |
Cp     266 TTGCAGCNTCCTGTGGTTNG 247


Search completed: Tue Dec 22 22:34:30 1998
Job time : 566 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Tue Dec 22 22:12:19 1998; MasPar time 482.64 Seconds
1433.772 Million cell updates/sec

Tabular output not generated.

Title: >US-08-530-112A-4937
Description: (1-314) from US08530112A.seq
Perfect Score: 298
N.A. Sequence: 1 GATCCTCTCTCCCAGGGAAT.....CCTNCNTGTGGCCCCTNTTN 314
Comp: CTAGGAGAGAGGGTCCCTTA.....GGANGNACACCGGGGANAAN

Scoring table: TABLE jmetric
Gap 60

Nmatch STD : Dbase 0; Query 0

Searched: 567134 seqs, 1101898692 bases x 2

Post-processing: Minimum Match 0%
Listing first 1000 summaries

Database: embl55
1:em_ba 2:em_fun 3:em_htg 4:em_hum1 5:em_hum2 6:em_in
7:em_om 8:em_or 9:em_ov 10:em_pat 11:em_ph 12:em_pl
13:em_ro 14:em_vi
Database: genbank107
15:gb_ba1 16:gb_ba2 17:gb_htg 18:gb_in 19:gb_om 20:gb_ov
21:gb_pat 22:gb_ph 23:gb_pl1 24:gb_pl2 25:gb_pr1
26:gb_pr2 27:gb_pr3 28:gb_ro 29:gb_st 30:gb_sts 31:gb_sy
32:gb_un 33:gb_vi

Statistics: Mean 8.211; Variance 2.717; scale 3.022

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
	1	223	74.8	5635	26	AF051782	Homo sapiens diaphanou	1.52e-243
	2	223	74.8	68067	26	AC005366	Homo sapiens chromosom	1.52e-243
c	3	19	6.4	27207	18	PFSC03017	Plasmodium falciparum	4.97e-01
c	4	19	6.4	35955	18	CEF53B7	Caenorhabditis elegans	4.97e-01
c	5	19	6.4	147476	17	AC004157	*** SEQUENCING IN PROG	4.97e-01
c	6	19	6.4	178615	17	HS742C19	Human DNA sequence ***	4.97e-01
	7	18	6.0	600	15	MTCMXAFAA	Methylobacter albus me	2.92e+00
	8	18	6.0	1514	28	AF030635	Mus musculus FK-506 bi	2.92e+00
	9	18	6.0	1641	25	HUMFKBPA	Human FK-506 binding p	2.92e+00
	10	18	6.0	8717	25	HSL226F1	Human DNA sequence fro	2.92e+00
	11	18	6.0	41414	27	AC005758	Homo sapiens chromosom	2.92e+00
	12	18	6.0	44219	27	AC005512	Homo sapiens chromosom	2.92e+00
	13	18	6.0	44897	26	AC004461	Homo sapiens Chromosom	2.92e+00
	14	18	6.0	49496	17	HS313F4	Human DNA sequence ***	2.92e+00
c	15	18	6.0	73864	17	HS496H19	Human DNA sequence ***	2.92e+00
	16	18	6.0	89350	23	ATT12H17	Arabidopsis thaliana D	2.92e+00
	17	18	6.0	90832	27	AC003065	Homo sapiens Chromosom	2.92e+00
c	18	18	6.0	108400	27	HUMDGCRCEN	Homo sapiens DiGeorge	2.92e+00
	19	18	6.0	139316	24	ATF7H19	Arabidopsis thaliana D	2.92e+00
	20	18	6.0	153460	27	HSU52111	Homo sapiens Xq28 geno	2.92e+00
	21	18	6.0	158537	26	HSAC002069	Human BAC clone RG326K	2.92e+00
c	22	18	6.0	167468	17	AC005033	*** SEQUENCING IN PROG	2.92e+00
c	23	18	6.0	175162	17	AC005182	*** SEQUENCING IN PROG	2.92e+00
	24	17	5.7	130	25	HS191F10R	H.sapiens CpG island D	1.61e+01
c	25	17	5.7	130	25	HS191F9F	H.sapiens CpG island D	1.61e+01
	26	17	5.7	263	24	AF081430	Oidiodendron chlamydos	1.61e+01
	27	17	5.7	264	24	AF081429	Oidiodendron scytaloid	1.61e+01
	28	17	5.7	513	24	AF062804	Oidiodendron scytaloid	1.61e+01
	29	17	5.7	522	24	AF062789	Oidiodendron chlamydos	1.61e+01
	30	17	5.7	1462	23	CAHIS1GEN	C.albicans HIS1 gene f	1.61e+01
	31	17	5.7	2881	26	AC002182	Homo sapiens (subclone	1.61e+01
	32	17	5.7	2900	20	XLU75996	Xenopus laevis eomesod	1.61e+01
	33	17	5.7	3195	33	PPB2LR	Bovine papillomavirus	1.61e+01
	34	17	5.7	3195	33	PABPV2L	Bovine papillomavirus	1.61e+01
	35	17	5.7	3547	26	HSAC001477	Homo sapiens (subclone	1.61e+01
	36	17	5.7	5127	18	AF029395	Drosophila melanogaste	1.61e+01
	37	17	5.7	7937	33	PPB2CG	Bovine papillomavirus	1.61e+01
c	38	17	5.7	8595	18	AF022713	Drosophila melanogaste	1.61e+01
c	39	17	5.7	9208	23	ATT3T7POL	Arabidopsis thaliana r	1.61e+01
c	40	17	5.7	9764	23	ATSSDDRPL	A.thaliana genes encod	1.61e+01
c	41	17	5.7	27008	18	CEC34E7	Caenorhabditis elegans	1.61e+01
	42	17	5.7	27013	18	CEM01B2	Caenorhabditis elegans	1.61e+01

	43	17	5.7	27470	20	FRHDGENE	F.rubripes Huntington'	1.61e+01
	44	17	5.7	30288	33	CAU55001	Canine adenovirus type	1.61e+01
	45	17	5.7	30536	33	CAV1GENOM	Canine adenovirus type	1.61e+01
c	46	17	5.7	32601	18	CET18D3	Caenorhabditis elegans	1.61e+01
	47	17	5.7	32828	18	CEK08E5	Caenorhabditis elegans	1.61e+01
	48	17	5.7	33448	18	CELC34B2	Caenorhabditis elegans	1.61e+01
	49	17	5.7	37386	18	CELT21P2	Caenorhabditis elegans	1.61e+01
c	50	17	5.7	38587	17	AC004177	*** SEQUENCING IN PROG	1.61e+01
	51	17	5.7	43236	18	CELC33H5	Caenorhabditis elegans	1.61e+01
c	52	17	5.7	59411	26	AC004040	Homo sapiens chromosom	1.61e+01
	53	17	5.7	75547	27	AC004919	Homo sapiens PAC clone	1.61e+01
	54	17	5.7	77447	18	L31790	Drosophila melanogaste	1.61e+01
c	55	17	5.7	79976	23	AB008264	Arabidopsis thaliana g	1.61e+01
	56	17	5.7	82512	25	HUM133K23	Human BAC clone 133K23	1.61e+01
	57	17	5.7	87585	18	AC004299	Drosophila melanogaste	1.61e+01
c	58	17	5.7	96642	23	AC002291	Arabidopsis thaliana c	1.61e+01
c	59	17	5.7	99682	26	HS452H17	Human DNA sequence fro	1.61e+01
c	60	17	5.7	100000	26	AP000032	Homo sapiens genomic D	1.61e+01
	61	17	5.7	109290	17	HS838L14	*** SEQUENCING IN PROG	1.61e+01
c	62	17	5.7	109741	24	ATAC004005	Arabidopsis thaliana c	1.61e+01
c	63	17	5.7	122940	17	AC004709	*** SEQUENCING IN PROG	1.61e+01
	64	17	5.7	123556	26	HS507I15	Human DNA sequence fro	1.61e+01
	65	17	5.7	130020	26	HUAC004525	Homo sapiens Chromosom	1.61e+01
c	66	17	5.7	131530	26	HUAC002303	Homo sapiens Chromosom	1.61e+01
	67	17	5.7	144838	17	HS376H23	Human DNA sequence ***	1.61e+01
	68	17	5.7	152184	17	CEY70G10	Caenorhabditis elegans	1.61e+01
c	69	17	5.7	161475	26	HS154C20	Homo sapiens DNA seque	1.61e+01
c	70	17	5.7	163216	26	HUAC002289	Homo sapiens Chromosom	1.61e+01
c	71	17	5.7	165507	17	AC004903	*** SEQUENCING IN PROG	1.61e+01
	72	17	5.7	168238	17	AC005014	*** SEQUENCING IN PROG	1.61e+01
c	73	17	5.7	177455	17	AC004980	*** SEQUENCING IN PROG	1.61e+01
c	74	17	5.7	178844	17	HS1054A22	Human DNA sequence ***	1.61e+01
c	75	17	5.7	197391	17	AC005140	*** SEQUENCING IN PROG	1.61e+01
	76	17	5.7	198326	17	AC004963	*** SEQUENCING IN PROG	1.61e+01
c	77	17	5.7	200611	23	ATFCA0	Arabidopsis thaliana D	1.61e+01
	78	17	5.7	204082	27	AC005323	Homo sapiens chromosom	1.61e+01
c	79	17	5.7	207931	17	HS1016G11	Human DNA sequence ***	1.61e+01
	80	17	5.7	212107	17	AC000134	*** SEQUENCING IN PROG	1.61e+01
	81	17	5.7	243820	17	CEY75B12	Caenorhabditis elegans	1.61e+01
	82	17	5.7	326569	17	AC004912	*** SEQUENCING IN PROG	1.61e+01
	83	17	5.7	338234	18	DMU31961	Drosophila melanogaste	1.61e+01
c	84	16	5.4	214	30	G13035	sWSS2929 Eric D. Green	8.18e+01
c	85	16	5.4	218	28	S77491	growth hormone recepto	8.18e+01
	86	16	5.4	246	30	G34250	human STS SHGC-51167,	8.18e+01
	87	16	5.4	282	30	G22942	human STS WI-11285.	8.18e+01
c	88	16	5.4	310	30	G04168	human STS WI-3079.	8.18e+01
c	89	16	5.4	310	30	G34679	human STS SHGC-52067,	8.18e+01
	90	16	5.4	333	30	MMSTS116	M.musculus (D11Bhm116)	8.18e+01
c	91	16	5.4	343	24	HVJ000236	Hordeum vulgare mRNA f	8.18e+01
	92	16	5.4	350	30	HSPE40F12	H.sapiens flow-sorted	8.18e+01
c	93	16	5.4	400	30	G17400	human STS SHGC-15020 c	8.18e+01

c	94	16	5.4	445	30	G24156	human STS WI-14099.	8.18e+01
c	95	16	5.4	459	23	BNPBN4RPT	B.nigra pBN-4 repetiti	8.18e+01
	96	16	5.4	462	25	HSPA29H1	H.sapiens flow-sorted	8.18e+01
c	97	16	5.4	470	20	S77208	Leu8 phyllolitorin=bom	8.18e+01
c	98	16	5.4	470	20	S77226	Phe8 phyllolitorin=bom	8.18e+01
	99	16	5.4	634	33	HLTSPLYSAD	Haliotis tuberculata l	8.18e+01
	100	16	5.4	651	18	HLTLYSINK	Abalone sperm lysin mR	8.18e+01
	101	16	5.4	695	23	BPU30357	Botryocladia pseudodic	8.18e+01
c	102	16	5.4	770	18	DMATDNA	D.melanogaster DNA, AT	8.18e+01
	103	16	5.4	890	28	RNEDN2S01	Rattus norvegicus endo	8.18e+01
c	104	16	5.4	1101	20	ZEFRARGII	Zebrafish retinoic aci	8.18e+01
c	105	16	5.4	1145	28	AF013288	Mus musculus 9-cis ret	8.18e+01
c	106	16	5.4	1175	16	LGU91638	Legionella gormanii ma	8.18e+01
c	107	16	5.4	1195	28	AF033195	Mus musculus IMAGE 776	8.18e+01
	108	16	5.4	1243	25	HUMIGHATC	Human immunoglobulin g	8.18e+01
c	109	16	5.4	1278	28	AF033196	Mus musculus IMAGE 693	8.18e+01
	110	16	5.4	1320	33	YFU52398	Yellow fever virus Ecu	8.18e+01
c	111	16	5.4	1623	23	SOYSACPD	Glycine max stearoyl-a	8.18e+01
c	112	16	5.4	1649	21	I22494	Sequence 35 from paten	8.18e+01
	113	16	5.4	1675	26	AF070591	Homo sapiens clone 246	8.18e+01
c	114	16	5.4	1687	25	S74134	J chain region: {pseud	8.18e+01
c	115	16	5.4	1688	25	HSU09955	Human ribosomal protei	8.18e+01
	116	16	5.4	1749	24	AF052401	Conidiobolus thromboid	8.18e+01
	117	16	5.4	1763	25	HUMMETIE	Human metallothionein-	8.18e+01
c	118	16	5.4	1837	23	YSCPOP2S	Saccharomyces cerevisi	8.18e+01
c	119	16	5.4	1870	23	YSCPOP2A	Saccharomyces cerevisi	8.18e+01
	120	16	5.4	1891	21	I08206	Sequence 15 from Paten	8.18e+01
	121	16	5.4	1891	23	YSCLAP4A	Saccharomyces cerevisi	8.18e+01
	122	16	5.4	1891	21	I06495	Sequence 4 from Patent	8.18e+01
c	123	16	5.4	1897	26	AC002250	Homo sapiens (subclone	8.18e+01
	124	16	5.4	2053	23	SCAPE1	Saccharomyces cerevisi	8.18e+01
	125	16	5.4	2058	21	A63891	Sequence 16 from Paten	8.18e+01
c	126	16	5.4	2090	23	OSRAMY3C	O.sativa RAmy3C gene f	8.18e+01
c	127	16	5.4	2109	20	S74156	zRAR gamma =retinoic a	8.18e+01
c	128	16	5.4	2163	26	HSU85048	Homo sapiens cyclic AM	8.18e+01
c	129	16	5.4	2201	23	SCYNR032W	S.cerevisiae chromosom	8.18e+01
	130	16	5.4	2207	26	U87864	Homo sapiens neuralize	8.18e+01
	131	16	5.4	2226	21	E07930	cDNA encoding feline z	8.18e+01
c	132	16	5.4	2284	26	HUMHA1551B	Human mRNA for KIAA004	8.18e+01
c	133	16	5.4	2337	23	YSPFRP1A	Schizosaccharomyces po	8.18e+01
	134	16	5.4	2461	26	AF017061	Homo sapiens vasopress	8.18e+01
c	135	16	5.4	2474	27	HSU58048	Human metallopeptidase	8.18e+01
c	136	16	5.4	2547	18	DMCPO612G	Drosophila melanogaste	8.18e+01
c	137	16	5.4	2631	18	DROSHPSHP	Drosophila melanogaste	8.18e+01
	138	16	5.4	2666	27	HSU52965	Human putative transcr	8.18e+01
c	139	16	5.4	2688	23	SCYKL103C	S.cerevisiae chromosom	8.18e+01
	140	16	5.4	2707	23	SCYNR053C	S.cerevisiae chromosom	8.18e+01
	141	16	5.4	2792	23	SCYNR052C	S.cerevisiae chromosom	8.18e+01
c	142	16	5.4	3035	18	DMCPO17G	D.melanogaster Cpo 17	8.18e+01
c	143	16	5.4	3186	25	HUMPDEB	Human phosphodiesteras	8.18e+01
c	144	16	5.4	3186	21	I22487	Sequence 23 from paten	8.18e+01

	145	16	5.4	3200	19	OCU30380	Oryctolagus cuniculus 8.18e+01
	146	16	5.4	3200	19	S78157	VACM-1=vasopressin-act 8.18e+01
	147	16	5.4	3325	18	CEU52515	Caenorhabditis elegans 8.18e+01
	148	16	5.4	3426	18	SPU17377	Strongylocentrotus pur 8.18e+01
	149	16	5.4	3601	26	HSVACM1	H.sapiens mRNA for for 8.18e+01
	150	16	5.4	3643	25	HUMPTYPH	Human protein-tyrosine 8.18e+01
c	151	16	5.4	3787	25	HUMCAMPB	Homo sapiens 3',5'-cyc 8.18e+01
	152	16	5.4	3794	26	AF027208	Homo sapiens AC133 ant 8.18e+01
c	153	16	5.4	3871	25	HUMPDE2A	Human rolipram-sebsiti 8.18e+01
c	154	16	5.4	4068	25	HUMPDEG	Human phosphodiesteras 8.18e+01
	155	16	5.4	4074	23	PRCRTI	Phaffia rhodozyma crtI 8.18e+01
c	156	16	5.4	4134	18	DMCPO611G	D.melanogaster Cpo 61. 8.18e+01
	157	16	5.4	4386	16	ECODADAX	E. coli D-amino acid d 8.18e+01
c	158	16	5.4	4489	28	RNU17096	Rattus norvegicus gamm 8.18e+01
	159	16	5.4	5112	5	HSUUSTLS4	Human FUS/TLS gene, ex 8.18e+01
	160	16	5.4	5521	18	PFSC04008	Plasmodium falciparum 8.18e+01
c	161	16	5.4	5555	23	SCYNR031C	S.cerevisiae chromosom 8.18e+01
	162	16	5.4	5565	16	AP005098	Lactococcus lactis RNA 8.18e+01
c	163	16	5.4	5792	20	XLU7SNRG	X.laevis U7 snRNA gene 8.18e+01
	164	16	5.4	6081	15	BPU53767	Bacillus pumilus plasm 8.18e+01
	165	16	5.4	6253	15	LMU38181	Leuconostoc mesenteroi 8.18e+01
	166	16	5.4	6294	28	RATGSTY	Rattus norvegicus glut 8.18e+01
	167	16	5.4	6465	18	CELEGAP3	Caenorhabditis elegans 8.18e+01
c	168	16	5.4	6520	25	HUMKER18	Human keratin 18 (K18) 8.18e+01
	169	16	5.4	7950	24	YSCSSK2	Saccharomyces cerevisi 8.18e+01
c	170	16	5.4	7963	27	HSE75B8	Human DNA sequence fro 8.18e+01
c	171	16	5.4	8283	28	MMINT2	Mouse int-2 gene. 8.18e+01
c	172	16	5.4	9347	28	MMRETDEHY	M.musculus gene rdh fo 8.18e+01
c	173	16	5.4	9580	15	U67471	Methanococcus jannasch 8.18e+01
c	174	16	5.4	10144	25	HUMHMGIY	Human high mobility gr 8.18e+01
c	175	16	5.4	10284	16	AE001280	Chlamydia trachomatis 8.18e+01
	176	16	5.4	10391	15	AE000996	Archaeoglobus fulgidus 8.18e+01
	177	16	5.4	11582	15	ECAE000217	Escherichia coli K-12 8.18e+01
c	178	16	5.4	11922	23	SCZ86109	S.carlsbergensis 12 kb 8.18e+01
	179	16	5.4	12321	27	AF071213	Homo sapiens FUS/TLS p 8.18e+01
	180	16	5.4	12887	16	AF077006	Helicobacter pylori pl 8.18e+01
c	181	16	5.4	13958	15	AE000901	Methanobacterium therm 8.18e+01
c	182	16	5.4	13990	15	AE001166	Borrelia burgdorferi (8.18e+01
	183	16	5.4	14661	15	D90753	Escherichia coli genom 8.18e+01
c	184	16	5.4	22429	18	CELF58F12	Caenorhabditis elegans 8.18e+01
	185	16	5.4	22559	18	CET27D1	Caenorhabditis elegans 8.18e+01
c	186	16	5.4	23140	18	CELF18E3	Caenorhabditis elegans 8.18e+01
c	187	16	5.4	23952	18	CEF29D10	Caenorhabditis elegans 8.18e+01
c	188	16	5.4	26857	26	AF048728	Homo sapiens minisatel 8.18e+01
c	189	16	5.4	27779	15	BSY13937	Bacillus subtilis geno 8.18e+01
	190	16	5.4	29156	25	HSL98A6B	Human DNA sequence fro 8.18e+01
	191	16	5.4	32472	18	CELT08B1	Caenorhabditis elegans 8.18e+01
	192	16	5.4	32547	18	CEK02A4	Caenorhabditis elegans 8.18e+01
c	193	16	5.4	33737	25	HSIGLAMB	Human lambda-immunoglo 8.18e+01
c	194	16	5.4	33977	18	CEF21G4	Caenorhabditis elegans 8.18e+01
c	195	16	5.4	35757	23	SCHAPLAP	S.cerevisiae HAP4, AAT 8.18e+01

	196	16	5.4	36133	27	HSA007973	Homo sapiens LGMD2B ge 8.18e+01
	197	16	5.4	36368	26	AF036876	Homo sapiens cosmid 12 8.18e+01
	198	16	5.4	37596	23	SPU33008	Schizosaccharomyces po 8.18e+01
	199	16	5.4	38801	18	CEK09B11	Caenorhabditis elegans 8.18e+01
c	200	16	5.4	39283	18	CET04A11	Caenorhabditis elegans 8.18e+01
	201	16	5.4	39305	23	SPU33009	Schizosaccharomyces po 8.18e+01
c	202	16	5.4	40152	26	D87017	Human (lambda) DNA for 8.18e+01
c	203	16	5.4	40392	26	D87023	Human (lambda) DNA for 8.18e+01
	204	16	5.4	40897	18	CELC39D10	Caenorhabditis elegans 8.18e+01
	205	16	5.4	41339	27	AC005490	Homo sapiens UWGC:g139 8.18e+01
	206	16	5.4	42793	23	SC9168	S.cerevisiae chromosom 8.18e+01
c	207	16	5.4	43236	18	CELC33H5	Caenorhabditis elegans 8.18e+01
	208	16	5.4	43311	17	HSNS3F10	Human DNA sequence *** 8.18e+01
	209	16	5.4	44427	18	CELF41E6	Caenorhabditis elegans 8.18e+01
c	210	16	5.4	44674	17	AC005652	*** SEQUENCING IN PROG 8.18e+01
c	211	16	5.4	46249	18	CELF53G2	Caenorhabditis elegans 8.18e+01
c	212	16	5.4	49501	17	AC005423	*** SEQUENCING IN PROG 8.18e+01
	213	16	5.4	51476	26	HS503P6	Homo sapiens DNA seque 8.18e+01
c	214	16	5.4	51498	26	HSU82694	Human cosmid Xq28_IA64 8.18e+01
	215	16	5.4	52173	26	HSG6PDGEN	H.sapiens G6PD gene fo 8.18e+01
	216	16	5.4	53367	17	HSLUCA11	Human DNA sequence *** 8.18e+01
	217	16	5.4	59641	28	D78344	Mouse DNA for Ig gamma 8.18e+01
c	218	16	5.4	64044	23	SCE9163	Saccharomyces cerevisi 8.18e+01
c	219	16	5.4	71482	17	HSNS6EB	Human DNA sequence *** 8.18e+01
	220	16	5.4	72698	23	AB007645	Arabidopsis thaliana g 8.18e+01
	221	16	5.4	77022	26	HS983L19	Human DNA sequence fro 8.18e+01
c	222	16	5.4	80507	27	AC004024	Homo sapiens 12q24.2 P 8.18e+01
c	223	16	5.4	80568	26	AC002420	Human Chromosome X, co 8.18e+01
c	224	16	5.4	81741	26	AC005367	Homo sapiens chromosom 8.18e+01
c	225	16	5.4	81850	23	ATFCA9	Arabidopsis thaliana D 8.18e+01
c	226	16	5.4	82503	24	AB016886	Arabidopsis thaliana g 8.18e+01
	227	16	5.4	82668	17	AC002559	*** SEQUENCING IN PROG 8.18e+01
c	228	16	5.4	85372	17	AC005508	*** SEQUENCING IN PROG 8.18e+01
c	229	16	5.4	85463	24	ATAC004483	Arabidopsis thaliana c 8.18e+01
	230	16	5.4	85837	23	SPU33010	Schizosaccharomyces po 8.18e+01
c	231	16	5.4	86263	23	AB006702	Arabidopsis thaliana g 8.18e+01
	232	16	5.4	86897	17	HSAJ9616	Homo sapiens *** SEQUE 8.18e+01
c	233	16	5.4	87246	26	AC002539	Homo sapiens chromosom 8.18e+01
c	234	16	5.4	89219	24	ATT6K22	Arabidopsis thaliana D 8.18e+01
	235	16	5.4	89603	27	HSU63834	Human KIT protein and 8.18e+01
c	236	16	5.4	90551	23	ATT5K18	Arabidopsis thaliana D 8.18e+01
	237	16	5.4	90665	24	ATAC005496	Arabidopsis thaliana c 8.18e+01
c	238	16	5.4	91184	17	HS633019	Human DNA sequence *** 8.18e+01
	239	16	5.4	95643	24	F6N15	Arabidopsis thaliana B 8.18e+01
c	240	16	5.4	98517	26	HS366N23	Human DNA sequence fro 8.18e+01
	241	16	5.4	100000	26	AP000010	Homo sapiens genomic D 8.18e+01
c	242	16	5.4	100000	26	AP000010	Homo sapiens genomic D 8.18e+01
	243	16	5.4	106760	26	HUAC002990	Human Chromosome 16 BA 8.18e+01
c	244	16	5.4	106793	26	AC002538	Homo sapiens chromosom 8.18e+01
c	245	16	5.4	109821	27	HS10C16	Human DNA sequence fro 8.18e+01
	246	16	5.4	110041	26	HS175E3	Human DNA sequence fro 8.18e+01

	247	16	5.4	110174	26	D83253	Homo sapiens genomic D 8.18e+01
c	248	16	5.4	110174	26	D83253	Homo sapiens genomic D 8.18e+01
	249	16	5.4	111368	26	AP000025	Homo sapiens genomic D 8.18e+01
c	250	16	5.4	111994	26	HS296K21	Human DNA sequence fro 8.18e+01
c	251	16	5.4	112776	17	AC005660	*** SEQUENCING IN PROG 8.18e+01
c	252	16	5.4	114173	26	HS884M20	Homo sapiens DNA seque 8.18e+01
c	253	16	5.4	114659	17	AC004918	*** SEQUENCING IN PROG 8.18e+01
c	254	16	5.4	114827	26	AC004413	Homo sapiens PAC clone 8.18e+01
c	255	16	5.4	116218	27	AC005090	Homo sapiens BAC clone 8.18e+01
	256	16	5.4	116884	17	PFMAL1P1	Plasmodium falciparum 8.18e+01
c	257	16	5.4	118387	17	AC005628	*** SEQUENCING IN PROG 8.18e+01
c	258	16	5.4	119044	17	HS355C18	Human DNA sequence *** 8.18e+01
c	259	16	5.4	126198	26	AC003089	Human BAC clone RG180F 8.18e+01
	260	16	5.4	127917	26	HS257A7	Human DNA sequence fro 8.18e+01
c	261	16	5.4	128136	17	HS889N15	Human DNA sequence *** 8.18e+01
	262	16	5.4	129295	17	HS48A11	Human DNA sequence *** 8.18e+01
	263	16	5.4	129969	26	HS204E5	Human DNA sequence fro 8.18e+01
c	264	16	5.4	130513	26	AC004592	Homo sapiens PAC clone 8.18e+01
c	265	16	5.4	130931	17	HS69M21	Human DNA sequence *** 8.18e+01
c	266	16	5.4	136037	26	AC004104	Homo sapiens 12q24 PAC 8.18e+01
c	267	16	5.4	136150	27	AC002081	Homo sapiens BAC clone 8.18e+01
	268	16	5.4	136805	17	AC005507	*** SEQUENCING IN PROG 8.18e+01
c	269	16	5.4	140856	26	AC002430	Human BAC clone RG066D 8.18e+01
	270	16	5.4	141120	26	AC002528	Human BAC clone GS542D 8.18e+01
	271	16	5.4	144201	27	HS230I19	Human DNA sequence fro 8.18e+01
c	272	16	5.4	145618	26	AC005138	Homo sapiens chromosom 8.18e+01
	273	16	5.4	149140	26	HS256G22	Human DNA sequence fro 8.18e+01
c	274	16	5.4	149752	26	AC004616	Homo sapiens Xp22 BAC 8.18e+01
	275	16	5.4	150223	33	HSECOMGEN	Equine herpesvirus 1 c 8.18e+01
	276	16	5.4	151152	26	HS454M7	Homo sapiens DNA seque 8.18e+01
	277	16	5.4	151630	26	AC005154	Homo sapiens PAC clone 8.18e+01
	278	16	5.4	152585	17	HS73M23	Human DNA sequence *** 8.18e+01
	279	16	5.4	152936	17	AC004229	*** SEQUENCING IN PROG 8.18e+01
	280	16	5.4	153568	26	AC002981	Homo sapiens Xp22 BAC 8.18e+01
c	281	16	5.4	155952	26	AC004534	Homo sapiens PAC clone 8.18e+01
c	282	16	5.4	157653	27	AC005146	Homo sapiens 12q24.2 P 8.18e+01
	283	16	5.4	159387	26	HS455H14	Homo sapiens DNA seque 8.18e+01
c	284	16	5.4	160850	17	AC004957	*** SEQUENCING IN PROG 8.18e+01
	285	16	5.4	165382	26	AF064860	Homo sapiens chromosom 8.18e+01
	286	16	5.4	167234	24	AC004708	CIC5B11.1 check: 4870 8.18e+01
	287	16	5.4	168636	17	AC002060	*** SEQUENCING IN PROG 8.18e+01
	288	16	5.4	170000	17	AC004579	*** SEQUENCING IN PROG 8.18e+01
	289	16	5.4	170000	17	AC004394	*** SEQUENCING IN PROG 8.18e+01
c	290	16	5.4	170000	17	AC005509	*** SEQUENCING IN PROG 8.18e+01
	291	16	5.4	171801	17	HS135L22	Human DNA sequence *** 8.18e+01
c	292	16	5.4	171905	26	HSAC002080	Human BAC clone RG253B 8.18e+01
	293	16	5.4	172984	17	HS862K6	Human DNA sequence *** 8.18e+01
	294	16	5.4	173494	17	HS662N3	Human DNA sequence *** 8.18e+01
c	295	16	5.4	173513	26	HS125H2	Human DNA sequence fro 8.18e+01
	296	16	5.4	176433	26	HS71L16	Human DNA sequence fro 8.18e+01
	297	16	5.4	177137	17	AC005089	*** SEQUENCING IN PROG 8.18e+01

c	298	16	5.4	180565	17	HS149A16	Human DNA sequence *** 8.18e+01
c	299	16	5.4	187051	17	AC004688	*** SEQUENCING IN PROG 8.18e+01
	300	16	5.4	190000	17	AC004580	*** SEQUENCING IN PROG 8.18e+01
c	301	16	5.4	190000	17	AC005598	*** SEQUENCING IN PROG 8.18e+01
c	302	16	5.4	190482	26	AC004106	Homo sapiens Xp22 BAC 8.18e+01
	303	16	5.4	192078	17	AC004885	*** SEQUENCING IN PROG 8.18e+01
	304	16	5.4	192609	17	AC005345	*** SEQUENCING IN PROG 8.18e+01
	305	16	5.4	193374	17	AC005139	*** SEQUENCING IN PROG 8.18e+01
	306	16	5.4	197391	17	AC005140	*** SEQUENCING IN PROG 8.18e+01
	307	16	5.4	200134	17	AC004806	*** SEQUENCING IN PROG 8.18e+01
c	308	16	5.4	204877	17	AC005375	*** SEQUENCING IN PROG 8.18e+01
c	309	16	5.4	208780	15	BSUB0009	Bacillus subtilis comp 8.18e+01
c	310	16	5.4	215609	26	AC002485	Human PAC clone RG212D 8.18e+01
c	311	16	5.4	215640	15	BSUB0016	Bacillus subtilis comp 8.18e+01
c	312	16	5.4	218470	15	BSUB0013	Bacillus subtilis comp 8.18e+01
c	313	16	5.4	219447	25	HUMFLNG6PD	Homo sapiens chromosom 8.18e+01
	314	16	5.4	220060	15	AP008220	Bacillus subtilis rrnB 8.18e+01
	315	16	5.4	244649	17	AC005308	*** SEQUENCING IN PROG 8.18e+01
c	316	16	5.4	250685	17	CET62E10	Caenorhabditis elegans 8.18e+01
c	317	16	5.4	259202	26	AC002366	Human Xp22 BAC CT-285I 8.18e+01
	318	16	5.4	282700	15	BACJH642	Bacillus subtilis DNA, 8.18e+01
c	319	16	5.4	315339	23	SCCHRIII	S.cerevisiae chromosom 8.18e+01
	320	15	5.0	133	28	MUSAFP05	Mouse alpha-fetoprotei 3.83e+02
	321	15	5.0	149	28	MUSMAT04	Mus musculus metallopr 3.83e+02
	322	15	5.0	161	25	HS45D10F	H.sapiens CpG island D 3.83e+02
c	323	15	5.0	161	25	HS45D10R	H.sapiens CpG island D 3.83e+02
	324	15	5.0	167	30	AF045310	Rattus norvegicus D5Uw 3.83e+02
c	325	15	5.0	182	28	MUSCROS2	Mus musculus protein k 3.83e+02
c	326	15	5.0	243	23	PINPOL1	Pinus thunbergii pol g 3.83e+02
	327	15	5.0	279	25	HUMCD1AA3C	Human differentiation 3.83e+02
c	328	15	5.0	285	28	MMIGFBPS3	Mus musculus insulin-l 3.83e+02
	329	15	5.0	299	25	HUMCD1B4	Human cortical thymocy 3.83e+02
c	330	15	5.0	327	26	UMCACN4S03	Human DNA for voltage- 3.83e+02
c	331	15	5.0	336	20	PSU90770	Pseudotropheus sp. 'go 3.83e+02
	332	15	5.0	346	30	HS344TE5	H.sapiens (D15S1035) D 3.83e+02
c	333	15	5.0	352	30	G25858	human STS EST276624. 3.83e+02
c	334	15	5.0	362	30	G38252	RPCI-5-937G1T7 Human 3.83e+02
c	335	15	5.0	364	30	HS203WE1	H.sapiens DNA segment 3.83e+02
	336	15	5.0	384	30	G06958	human STS WI-8657. 3.83e+02
c	337	15	5.0	387	25	HSNL1291S	H.sapiens genomic DNA 3.83e+02
	338	15	5.0	400	30	G19162	human STS SHGC-18171 c 3.83e+02
	339	15	5.0	400	30	G11030	human STS SHGC-8799 cl 3.83e+02
c	340	15	5.0	422	28	MUSCRRY05	Mouse complement recep 3.83e+02
c	341	15	5.0	432	18	PRZ97971	Polystenus rugosus 28S 3.83e+02
	342	15	5.0	437	28	MMU50955	Mus musculus clone TSA 3.83e+02
c	343	15	5.0	463	30	G27522	human STS SHGC-32407. 3.83e+02
	344	15	5.0	465	28	MUSS70P03	Mouse gene for mitocho 3.83e+02
c	345	15	5.0	486	30	MASTMS436	M.acuminata DNA for se 3.83e+02
c	346	15	5.0	518	30	MSZ85964	Musa sp. DNA for seque 3.83e+02
	347	15	5.0	534	21	A47593	Sequence 6 from Patent 3.83e+02
	348	15	5.0	537	25	HSPA14H9	H.sapiens flow-sorted 3.83e+02

	349	15	5.0	580	28	MMU63546	Mus musculus T cell re 3.83e+02
	350	15	5.0	587	30	G38194	RPCI-3-335J1T7 Human H 3.83e+02
	351	15	5.0	598	26	HSSUR2G22	Homo sapiens sulfonylu 3.83e+02
c	352	15	5.0	598	20	XLGLB5	Xenopus laevis globin 3.83e+02
c	353	15	5.0	600	30	DM171A7S	Drosophila melanogaste 3.83e+02
c	354	15	5.0	604	15	CLOFDX	C.pasteurianum ferredo 3.83e+02
c	355	15	5.0	615	23	BMVVP	Belladonna mottle viru 3.83e+02
c	356	15	5.0	622	27	HSY17975	Homo sapiens mRNA for 3.83e+02
	357	15	5.0	633	15	STAGYRAB	S.aureus gyrase (gyrB 3.83e+02
c	358	15	5.0	682	28	RATIL10X	Rat interleukin 10 (IL 3.83e+02
	359	15	5.0	683	18	AF009907	Caenorhabditis elegans 3.83e+02
	360	15	5.0	687	30	HUMUT7572	Human STS UT7572. 3.83e+02
	361	15	5.0	694	28	MUSTCRAVM	Mus musculus TCRAV2.7 3.83e+02
	362	15	5.0	694	28	MUSTCRAVK	Mus musculus TCRAV2.5 3.83e+02
	363	15	5.0	703	30	AF007199	Apodemus agrarius GTT- 3.83e+02
	364	15	5.0	708	30	G34933	Atlantic walrus STS Or 3.83e+02
	365	15	5.0	831	25	PTU19114	Pan troglodytes HOXB6 3.83e+02
	366	15	5.0	832	25	GGU19115	Gorilla gorilla HOXB6 3.83e+02
	367	15	5.0	832	25	PPU19116	Pongo pygmaeus HOXB6 g 3.83e+02
	368	15	5.0	858	25	SSU19113	Hylobates syndactylus 3.83e+02
	369	15	5.0	858	25	HKU19112	Hylobates klossi HOXB6 3.83e+02
	370	15	5.0	858	25	HLU19117	Hylobates lar HOXB6 ge 3.83e+02
	371	15	5.0	886	28	MMU04624	Mus musculus BALB/c T 3.83e+02
	372	15	5.0	887	28	MMU04626	Mus musculus BALB/c T 3.83e+02
c	373	15	5.0	901	15	RP282424	R.prowazekii genomic D 3.83e+02
	374	15	5.0	925	30	AF032441	Canis familiaris Y chr 3.83e+02
c	375	15	5.0	963	32	MLVRT4	Murine leukemia relate 3.83e+02
c	376	15	5.0	990	28	MUSXYG	Mouse X/Y protein mRNA 3.83e+02
	377	15	5.0	1004	25	HSLCK2	Human lymphocyte-speci 3.83e+02
c	378	15	5.0	1008	19	BRENDOG	B.taurus mRNA for endo 3.83e+02
c	379	15	5.0	1026	16	NSI223916	Neisseria sicca rho ge 3.83e+02
c	380	15	5.0	1031	25	S77906	Pax (PAX8e)=paired box 3.83e+02
	381	15	5.0	1072	28	MUSMAT	Mus musculus metallopr 3.83e+02
c	382	15	5.0	1095	18	MILMND1	Locusta migratoria mit 3.83e+02
	383	15	5.0	1131	21	I11620	Sequence 2 from Patent 3.83e+02
c	384	15	5.0	1133	23	MEU51851	Morchella esculenta 17 3.83e+02
c	385	15	5.0	1164	16	PFU37103	Pseudomonas fluorescen 3.83e+02
c	386	15	5.0	1193	18	DDIHTATP	Dictyostelium discoide 3.83e+02
c	387	15	5.0	1209	28	RATDBP07	Rat vitamin D binding 3.83e+02
c	388	15	5.0	1220	25	S77905	Pax (PAX8d)=paired box 3.83e+02
c	389	15	5.0	1238	23	PVU72663	Phaseolus vulgaris uri 3.83e+02
	390	15	5.0	1241	25	HUMIGHAOC	Human immunoglobulin g 3.83e+02
	391	15	5.0	1295	25	HUMCD1B	Human thymocyte antige 3.83e+02
c	392	15	5.0	1307	28	RNIL10	R.norvegicus mRNA for 3.83e+02
	393	15	5.0	1309	23	YSAZNF1	Candida albicans zinc 3.83e+02
c	394	15	5.0	1343	16	CBCBM17	C.bifermentans cbm17.1 3.83e+02
c	395	15	5.0	1372	25	HUMPAX8A	Human paired box homeo 3.83e+02
c	396	15	5.0	1451	25	S77904	Pax (PAX8c)=paired box 3.83e+02
	397	15	5.0	1464	28	RATBTG1H	Rattus norvegicus anti 3.83e+02
	398	15	5.0	1473	21	I01816	Sequence 1 from Patent 3.83e+02
	399	15	5.0	1474	15	BACFLAGA	B.subtilis flagellin g 3.83e+02

c	400	15	5.0	1507	20	CHKCYB5	Chicken cytochrome b5 3.83e+02
c	401	15	5.0	1568	20	XLU95098	Xenopus laevis mitotic 3.83e+02
	402	15	5.0	1578	21	E07867	cDNA of human alpha-ph 3.83e+02
	403	15	5.0	1602	26	HSY13927	Homo sapiens PRKY exon 3.83e+02
c	404	15	5.0	1663	23	PERETREL	P.elliotii DNA for rep 3.83e+02
	405	15	5.0	1672	25	HSGPCRCO	H.sapiens mRNA for G p 3.83e+02
c	406	15	5.0	1716	15	AB009289	Streptomyces griseus g 3.83e+02
c	407	15	5.0	1739	20	CFU82688	Chamaeleo fischeri CAS 3.83e+02
c	408	15	5.0	1757	25	HUMMFAPA	Human microfibril-asso 3.83e+02
c	409	15	5.0	1781	28	MUSCRYL	Mus musculus complemen 3.83e+02
	410	15	5.0	1788	25	MFALPBLDL	Macaca fascicularis mRN 3.83e+02
	411	15	5.0	1799	28	MUSMS70	Mouse pseudogene for m 3.83e+02
	412	15	5.0	1800	28	MUSBTG1X	Mouse BTG1 mRNA, compl 3.83e+02
c	413	15	5.0	1820	25	HSU18300	Human damage-specific 3.83e+02
c	414	15	5.0	1846	26	HSO19AO14	Human DNA sequence fro 3.83e+02
c	415	15	5.0	1864	28	RATALBA1	Rat transthyretin gene 3.83e+02
c	416	15	5.0	1869	23	ATU18675	Arabidopsis thaliana 4 3.83e+02
	417	15	5.0	1875	20	GGU83593	Gallus gallus aggrecan 3.83e+02
	418	15	5.0	1894	23	SCYBR177C	S.cerevisiae chromosom 3.83e+02
c	419	15	5.0	1915	24	AF014802	Eschscholzia californi 3.83e+02
c	420	15	5.0	1929	15	PLLUXCG	P.leiognathi luxC gene 3.83e+02
	421	15	5.0	1937	15	ERWIAAM	Erwinia herbicola Y015 3.83e+02
	422	15	5.0	1979	27	AF038392	Homo sapiens pre-mRNA 3.83e+02
	423	15	5.0	2007	25	HSU19111	Human HOXB6 gene, Hoxb 3.83e+02
	424	15	5.0	2009	28	MMFETO	Mouse mRNA encoding al 3.83e+02
c	425	15	5.0	2027	20	XELBGAI	X.laevis adult beta-gl 3.83e+02
c	426	15	5.0	2027	20	XELHBBBAI	X.laevis beta-A-I glob 3.83e+02
	427	15	5.0	2074	28	RATNT3H3MG	Rattus norvegicus mito 3.83e+02
c	428	15	5.0	2078	19	BTKER19B	B.taurus cytokeratin 1 3.83e+02
	429	15	5.0	2090	28	S75280	pre-mtHSP70=70 kda hea 3.83e+02
	430	15	5.0	2155	28	MUSP66	Mouse mRNA for p66 mot 3.83e+02
	431	15	5.0	2155	21	I48938	Sequence 1 from patent 3.83e+02
	432	15	5.0	2203	26	HSU96131	Homo sapiens HPV16 E1 3.83e+02
	433	15	5.0	2228	19	BTUPA	B.taurus mRNA for urok 3.83e+02
	434	15	5.0	2248	15	CJDNATIG	C.jejuni tig, folE & c 3.83e+02
	435	15	5.0	2292	16	AFU44893	Butyrivibrio fibrisolv 3.83e+02
	436	15	5.0	2314	15	BHU63319	Borrelia hermsii plasm 3.83e+02
	437	15	5.0	2336	28	MUSPBP	Mouse mRNA sequence. 3.83e+02
	438	15	5.0	2355	19	BOVUKPA	Bos taurus urokinase-t 3.83e+02
	439	15	5.0	2360	15	BACHAGFA	B.subtilis flagellin g 3.83e+02
	440	15	5.0	2366	26	AF044896	Homo sapiens ICB-1 mRN 3.83e+02
	441	15	5.0	2369	28	AF012104	Mus musculus cytoplasm 3.83e+02
	442	15	5.0	2513	18	PFAACTIIA	P.falciparum actin II 3.83e+02
	443	15	5.0	2580	19	ERD97339	Elephantulus rufescens 3.83e+02
	444	15	5.0	2608	15	AAQUINOX	A.ambivalens doxD gene 3.83e+02
	445	15	5.0	2609	28	MMU88091	Mus musculus protein t 3.83e+02
c	446	15	5.0	2692	20	CMBGA2B2	Duck rearranged beta-g 3.83e+02
c	447	15	5.0	2711	25	HSPAX8A	H.sapiens Pax8 mRNA. 3.83e+02
c	448	15	5.0	2760	23	U10669	Arabidopsis thaliana p 3.83e+02
c	449	15	5.0	2766	25	HSU15782	Human cleavage stimula 3.83e+02
c	450	15	5.0	2773	15	COWIMMDOM	Cowdria ruminantium (c 3.83e+02

c	451	15	5.0	2806	28	MMAJ2364	Mus musculus interleuk 3.83e+02
	452	15	5.0	2820	27	AB010153	Homo sapiens mRNA for 3.83e+02
c	453	15	5.0	2857	20	AF030997	Agelaius phoeniceus MH 3.83e+02
	454	15	5.0	2987	23	AF004396	Arabidopsis thaliana s 3.83e+02
	455	15	5.0	2997	28	MUSPBP74	Mouse mRNA for mitocho 3.83e+02
	456	15	5.0	3001	28	S78556	grp75=75 kda glucose r 3.83e+02
	457	15	5.0	3019	23	U94783	Helianthus annuus unco 3.83e+02
	458	15	5.0	3068	26	HSL81088	Homo sapiens (subclone 3.83e+02
	459	15	5.0	3245	26	HOS223093	Homo sapiens LAGE-1 ge 3.83e+02
c	460	15	5.0	3278	26	MMU89918	Macaca mulatta homeobo 3.83e+02
	461	15	5.0	3295	20	CCEP2GEN	C.carpio EP2 gene. 3.83e+02
c	462	15	5.0	3297	18	DMFPS85D	Drosophila melanogaste 3.83e+02
	463	15	5.0	3299	26	HSAJ3324	Homo sapiens mRNA for 3.83e+02
c	464	15	5.0	3320	18	GIU93353	Giardia intestinalis a 3.83e+02
c	465	15	5.0	3331	23	SCYKL179C	S.cerevisiae chromosom 3.83e+02
c	466	15	5.0	3402	15	FMUB1184	Fusobacterium mortifer 3.83e+02
	467	15	5.0	3409	26	HSL81776	Homo sapiens (subclone 3.83e+02
	468	15	5.0	3468	25	HUMSCNA22	Human type IV sodium c 3.83e+02
	469	15	5.0	3510	25	HSINTA6	H.sapiens mRNA for int 3.83e+02
	470	15	5.0	3510	27	HSU89995	Human DNA binding prot 3.83e+02
c	471	15	5.0	3527	23	GMGY1	Soybean Gy1 gene for g 3.83e+02
	472	15	5.0	3615	27	HS717AX	Homo sapiens mRNA for 3.83e+02
	473	15	5.0	3620	18	DMU04692	Drosophila melanogaste 3.83e+02
	474	15	5.0	3741	27	AB012942S1	Homo sapiens AOC2 gene 3.83e+02
	475	15	5.0	3851	23	ATATM	A.thaliana ATM mRNA fo 3.83e+02
c	476	15	5.0	4042	27	AF087142	Homo sapiens TED prote 3.83e+02
	477	15	5.0	4192	15	BSFLIDST	B.subtilis (EB2058) ge 3.83e+02
	478	15	5.0	4247	15	VIBICD	Vibrio sp. genes for i 3.83e+02
	479	15	5.0	4278	27	HS717AXAS	Homo sapiens mRNA for 3.83e+02
c	480	15	5.0	4316	15	BBTYCA	Bacillus brevis tycA g 3.83e+02
	481	15	5.0	4317	23	SOYACCAA	Glycine max (clone 91) 3.83e+02
	482	15	5.0	4465	28	RNAPOBEC2	Rattus norvegicus apol 3.83e+02
c	483	15	5.0	4483	15	BSU13979	Bacillus subtilis mino 3.83e+02
c	484	15	5.0	4527	33	SSHM	Snowshoe hare virus, c 3.83e+02
	485	15	5.0	4588	28	MMU56734	Mus musculus lectin la 3.83e+02
	486	15	5.0	4634	25	HUMAPOB	Human apolipoprotein B 3.83e+02
c	487	15	5.0	4638	26	AF035180	Homo sapiens chromosom 3.83e+02
c	488	15	5.0	4641	26	AF035179	Homo sapiens chromosom 3.83e+02
	489	15	5.0	4674	19	AF020710	Oryctolagus cuniculus 3.83e+02
	490	15	5.0	4756	27	AF091627	Homo sapiens CUSP mRNA 3.83e+02
	491	15	5.0	4779	18	AF035549	Drosophila melanogaste 3.83e+02
c	492	15	5.0	4848	18	AF025312	Tetrahymena thermophil 3.83e+02
	493	15	5.0	4950	15	STAGYRABA	Staphylococcus aureus 3.83e+02
	494	15	5.0	5031	23	ATHRPB3A	Arabidopsis thaliana R 3.83e+02
c	495	15	5.0	5135	15	TTNUSAINF	T.thermophilus nusA/in 3.83e+02
c	496	15	5.0	5173	18	DDG14D5	D. discoideum plasmid 3.83e+02
c	497	15	5.0	5282	18	PFP190	P.falciparum gp190 (MS 3.83e+02
	498	15	5.0	5323	26	HUMORFLA	Human mRNA for KIAA001 3.83e+02
	499	15	5.0	5348	15	STARECF	Staphylococcus aureus 3.83e+02
c	500	15	5.0	5431	19	OAPLE1	Ovis aries placental l 3.83e+02
	501	15	5.0	5499	21	I32960	Sequence 4 from patent 3.83e+02

c	502	15	5.0	5576	28	MUSMHCT2A	Mouse MHC class I Tla- 3.83e+02
c	503	15	5.0	5600	18	PFSC03005	Plasmodium falciparum 3.83e+02
	504	15	5.0	5629	21	I32959	Sequence 2 from patent 3.83e+02
	505	15	5.0	5629	26	HSINTA6R	Human mRNA for integri 3.83e+02
c	506	15	5.0	5754	18	PFAMMSAA	P.falciparum major mer 3.83e+02
c	507	15	5.0	5760	21	A04562	P.falciparum P.195 gen 3.83e+02
c	508	15	5.0	5760	21	E00656	cDNA encoding P.195 pr 3.83e+02
	509	15	5.0	5848	18	PFSC04081	Plasmodium falciparum 3.83e+02
	510	15	5.0	5909	15	SAGYRREC	S.aureus genes gyrB, g 3.83e+02
c	511	15	5.0	5917	18	PFANT195	Plasmodium falciparum 3.83e+02
	512	15	5.0	6597	20	GGU78555	Gallus gallus brain ch 3.83e+02
	513	15	5.0	6748	20	CPU35006	Carcharhinus plumbeus 3.83e+02
c	514	15	5.0	7000	15	LIU61226	Leptospira interrogans 3.83e+02
c	515	15	5.0	7038	18	PFGP195A	Plasmodium falciparum 3.83e+02
c	516	15	5.0	7200	18	AF030197	Drosophila melanogaste 3.83e+02
c	517	15	5.0	7238	15	AF030941	Neisseria meningitidis 3.83e+02
c	518	15	5.0	7288	28	MMCROSP	M.musculus mRNA for c- 3.83e+02
c	519	15	5.0	7407	28	MMU15443	Mus musculus proto-onc 3.83e+02
c	520	15	5.0	7439	18	DMIS176	Drosophila melanogaste 3.83e+02
c	521	15	5.0	7462	23	CCPLAS	C.caldarium plastid ge 3.83e+02
	522	15	5.0	7485	15	SEU71377	Staphylococcus epiderm 3.83e+02
	523	15	5.0	7598	15	AAPDL10	Acidianus ambivalens p 3.83e+02
c	524	15	5.0	7791	16	SAY14370	Staphylococcus aureus 3.83e+02
	525	15	5.0	7823	25	HUMSKM1A	Homo sapiens skeletal 3.83e+02
c	526	15	5.0	8039	27	AF009427	Homo sapiens clone 22 3.83e+02
c	527	15	5.0	8093	27	AF009426	Homo sapiens clone 22 3.83e+02
	528	15	5.0	8405	33	AB010302	Garlic virus C genomic 3.83e+02
	529	15	5.0	8421	18	DVU39746	Drosophila virilis gla 3.83e+02
c	530	15	5.0	8440	27	AF009425	Homo sapiens clone 22 3.83e+02
c	531	15	5.0	8494	27	AF009424	Homo sapiens clone 22 3.83e+02
c	532	15	5.0	8670	26	HSS100A2	H.sapiens S100A2 gene, 3.83e+02
c	533	15	5.0	8930	25	HUMBPAG1A	Human bullous pemphigo 3.83e+02
c	534	15	5.0	8956	23	MTPLCOX1	P.littoralis mitochond 3.83e+02
	535	15	5.0	9367	20	U00147	Gallus domesticus fila 3.83e+02
c	536	15	5.0	9531	27	AF055475	Homo sapiens GAGE-8 ge 3.83e+02
c	537	15	5.0	9983	18	DMU96440	Drosophila melanogaste 3.83e+02
	538	15	5.0	10316	15	SERFBC1	Salmonella enterica gr 3.83e+02
c	539	15	5.0	10532	15	HPAE000626	Helicobacter pylori se 3.83e+02
	540	15	5.0	10901	16	AE001305	Chlamydia trachomatis 3.83e+02
c	541	15	5.0	11170	18	CVULET23	C. vulgaris let-23 hom 3.83e+02
c	542	15	5.0	12507	28	RNY16563	Rattus norvegicus mRNA 3.83e+02
c	543	15	5.0	12674	18	AF053496	Caenorhabditis elegans 3.83e+02
c	544	15	5.0	12993	18	CELPOLII	C.elegans RNA polymera 3.83e+02
c	545	15	5.0	13674	18	U00691	Dictyostelium discoide 3.83e+02
c	546	15	5.0	13729	18	U00796	Dictyostelium discoide 3.83e+02
	547	15	5.0	13993	25	HUMAPOBB	Human apolipoprotein B 3.83e+02
	548	15	5.0	14041	26	HSKI67	H.sapiens DNA for Ki-6 3.83e+02
	549	15	5.0	14070	25	HUMAPOBA	Human apolipoprotein B 3.83e+02
c	550	15	5.0	14096	16	AF084104	Bacillus firmus AcsA (3.83e+02
	551	15	5.0	14107	20	FRBB6GAP01	Fugu rubripes b8b6gap 3.83e+02
	552	15	5.0	14112	26	HSAPOB10	Human mRNA for apolipo 3.83e+02

	553	15	5.0	14121	25	HSLDL100	Human mRNA for apolipo 3.83e+02
	554	15	5.0	14635	25	HUMAPB21	Human apolipoprotein B 3.83e+02
	555	15	5.0	14882	23	SOYACCASEA	Glycine max (clones 51 3.83e+02
c	556	15	5.0	14993	18	CEF57G9	Caenorhabditis elegans 3.83e+02
	557	15	5.0	15722	18	MTLMCOMP	L.migratoria mitochond 3.83e+02
c	558	15	5.0	15999	15	AE000750	Aquifex aeolicus secti 3.83e+02
	559	15	5.0	18875	15	BSU93874	Bacillus subtilis cyst 3.83e+02
c	560	15	5.0	19188	18	CELC45E5	Caenorhabditis elegans 3.83e+02
c	561	15	5.0	20279	18	CER31	Caenorhabditis elegans 3.83e+02
	562	15	5.0	20320	16	BSU56901	Bacillus subtilis puta 3.83e+02
	563	15	5.0	20803	18	CELF57H12	Caenorhabditis elegans 3.83e+02
c	564	15	5.0	22947	16	MTCY4C12	Mycobacterium tubercul 3.83e+02
	565	15	5.0	23140	18	CELF18E3	Caenorhabditis elegans 3.83e+02
	566	15	5.0	26624	23	AB011481	Arabidopsis thaliana g 3.83e+02
c	567	15	5.0	27991	18	CEF44A6	Caenorhabditis elegans 3.83e+02
	568	15	5.0	28230	26	HSCG1160	Human cosmid cG1160 fr 3.83e+02
c	569	15	5.0	28951	25	HSL161C2	Human DNA sequence fro 3.83e+02
	570	15	5.0	29120	18	CELC02E7	Caenorhabditis elegans 3.83e+02
c	571	15	5.0	29612	18	CEF19H6	Caenorhabditis elegans 3.83e+02
	572	15	5.0	30148	18	CEF57C7	Caenorhabditis elegans 3.83e+02
c	573	15	5.0	30281	26	HS334P19	Human DNA sequence fro 3.83e+02
	574	15	5.0	30652	18	CELF59D6	Caenorhabditis elegans 3.83e+02
c	575	15	5.0	30813	18	CET19H5	Caenorhabditis elegans 3.83e+02
	576	15	5.0	31180	18	DMC155E2	Drosophila melanogaste 3.83e+02
c	577	15	5.0	31223	18	CEZK270	Caenorhabditis elegans 3.83e+02
	578	15	5.0	31434	18	CEY68A4A	Caenorhabditis elegans 3.83e+02
c	579	15	5.0	31434	18	CEY68A4A	Caenorhabditis elegans 3.83e+02
c	580	15	5.0	31757	27	HSN53A9	Human DNA sequence fro 3.83e+02
c	581	15	5.0	32785	18	CEC53B4	Caenorhabditis elegans 3.83e+02
c	582	15	5.0	32906	26	HSIFNAR	Human IFNAR gene for i 3.83e+02
	583	15	5.0	33466	18	CEZC47	Caenorhabditis elegans 3.83e+02
	584	15	5.0	34388	18	CET28A8	Caenorhabditis elegans 3.83e+02
	585	15	5.0	34864	26	U73649	Human Chromosome 11 Co 3.83e+02
	586	15	5.0	35170	18	CELAH9	Caenorhabditis elegans 3.83e+02
c	587	15	5.0	35214	18	CELC42D4	Caenorhabditis elegans 3.83e+02
	588	15	5.0	35257	18	CELC07G3	Caenorhabditis elegans 3.83e+02
	589	15	5.0	35605	27	AC000045	Homo sapiens Chromosom 3.83e+02
	590	15	5.0	35811	23	SC9718	S.cerevisiae chromosom 3.83e+02
	591	15	5.0	36148	17	AC002321	*** SEQUENCING IN PROG 3.83e+02
c	592	15	5.0	36180	18	CET05G5	Caenorhabditis elegans 3.83e+02
c	593	15	5.0	36344	18	CELC54H2	Caenorhabditis elegans 3.83e+02
c	594	15	5.0	36490	18	CEF25D7	Caenorhabditis elegans 3.83e+02
	595	15	5.0	36631	25	HUMCOL7A1X	Homo sapiens (clones: 3.83e+02
c	596	15	5.0	36960	26	AC004742	Homo sapiens PAC clone 3.83e+02
	597	15	5.0	37056	26	AC004700	Homo sapiens chromosom 3.83e+02
c	598	15	5.0	37302	26	D87007	Human (lambda) DNA for 3.83e+02
	599	15	5.0	37666	18	CET24D5	Caenorhabditis elegans 3.83e+02
c	600	15	5.0	37675	18	CELZK370	C. elegans cosmid ZK37 3.83e+02
c	601	15	5.0	37801	18	CEAH6	Caenorhabditis elegans 3.83e+02
	602	15	5.0	38013	25	HSL17A9B	Human DNA sequence fro 3.83e+02
	603	15	5.0	38152	18	CELW03B1	Caenorhabditis elegans 3.83e+02

	604	15	5.0	38155	18	CELF11C7	Caenorhabditis elegans 3.83e+02
c	605	15	5.0	38262	26	HSN98E6	Human DNA sequence fro 3.83e+02
c	606	15	5.0	38477	23	SCPEKGA	S.cerevisiae sequence 3.83e+02
c	607	15	5.0	38503	15	MSGB1912CS	M. leprae genomic dna 3.83e+02
c	608	15	5.0	38649	26	AC004391	Homo sapiens Cosmid g5 3.83e+02
c	609	15	5.0	38675	15	MLU15180	Mycobacterium leprae c 3.83e+02
	610	15	5.0	38919	18	DMC133E12	Drosophila melanogaste 3.83e+02
	611	15	5.0	39187	18	CELC01F1	Caenorhabditis elegans 3.83e+02
	612	15	5.0	39266	18	CEF20E11	Caenorhabditis elegans 3.83e+02
c	613	15	5.0	39266	18	CEF20E11	Caenorhabditis elegans 3.83e+02
c	614	15	5.0	39305	23	SPU33009	Schizosaccharomyces po 3.83e+02
c	615	15	5.0	39337	18	CELB0286	Caenorhabditis elegans 3.83e+02
	616	15	5.0	39399	15	MSGB1970CS	Mycobacterium leprae c 3.83e+02
c	617	15	5.0	39485	26	AF015721	Homo sapiens chromosom 3.83e+02
c	618	15	5.0	39731	27	AC005760	Homo sapiens chromosom 3.83e+02
	619	15	5.0	39822	15	BSPEPSYN	B.subtilis genes for p 3.83e+02
c	620	15	5.0	40017	26	D86998	Human (lambda) DNA for 3.83e+02
c	621	15	5.0	40205	26	AP045450	Homo sapiens chromosom 3.83e+02
c	622	15	5.0	40216	18	CELC06E1	C. elegans cosmid C06E 3.83e+02
c	623	15	5.0	40222	18	CELF22B7	C.aenorhabditis elegan 3.83e+02
	624	15	5.0	40460	27	AC005492	Homo sapiens clone UWG 3.83e+02
	625	15	5.0	40822	25	HSCU177E8	Human DNA sequence fro 3.83e+02
c	626	15	5.0	40897	16	AF004835	Brevibacillus brevis t 3.83e+02
	627	15	5.0	41114	18	CELF16B4	Caenorhabditis elegans 3.83e+02
c	628	15	5.0	41172	27	AC005546	Homo sapiens chromosom 3.83e+02
	629	15	5.0	41971	18	AC003765	Drosophila melanogaste 3.83e+02
c	630	15	5.0	42045	27	L48038	Structure of the promo 3.83e+02
c	631	15	5.0	42398	26	D84401	Homo sapiens genomic D 3.83e+02
	632	15	5.0	42418	17	HSN38G6	Human DNA sequence *** 3.83e+02
	633	15	5.0	42464	27	AC005563	Homo sapiens chromosom 3.83e+02
	634	15	5.0	42648	27	AC005685	Homo sapiens clone UWG 3.83e+02
c	635	15	5.0	42936	18	CELF36A4	Caenorhabditis elegans 3.83e+02
	636	15	5.0	43808	28	AC003996	Mouse Cosmid ma66a097 3.83e+02
c	637	15	5.0	43875	17	AC005432	*** SEQUENCING IN PROG 3.83e+02
	638	15	5.0	45022	18	CEC34D1	Caenorhabditis elegans 3.83e+02
c	639	15	5.0	45110	17	AC005032	*** SEQUENCING IN PROG 3.83e+02
c	640	15	5.0	45601	27	AC004191	Homo sapiens clone UWG 3.83e+02
c	641	15	5.0	45635	18	CELT23E7	Caenorhabditis elegans 3.83e+02
	642	15	5.0	45678	25	HSV351F8	Human DNA sequence fro 3.83e+02
	643	15	5.0	45713	18	CELB0511	Caenorhabditis elegans 3.83e+02
	644	15	5.0	46251	26	AD000671	Homo sapiens DNA from 3.83e+02
c	645	15	5.0	46275	26	AC003107	Human DNA from chromos 3.83e+02
c	646	15	5.0	46864	16	AF012285	Bacillus subtilis mobA 3.83e+02
	647	15	5.0	48469	26	AC002484	Human cosmid clone LL1 3.83e+02
	648	15	5.0	49323	17	AC003067	*** SEQUENCING IN PROG 3.83e+02
	649	15	5.0	55186	18	AC001655	Drosophila melanogaste 3.83e+02
	650	15	5.0	56075	26	HS161N10	Human DNA sequence fro 3.83e+02
c	651	15	5.0	56632	26	HS511E16	Human DNA sequence fro 3.83e+02
	652	15	5.0	57688	26	AP000048	Homo sapiens genomic D 3.83e+02
c	653	15	5.0	58949	18	CELK07E12	Caenorhabditis elegans 3.83e+02
c	654	15	5.0	59105	26	AC004013	Human BAC clone GS541B 3.83e+02

	655	15	5.0	60029	26	AC003689	Human Chromosome 11 pa 3.83e+02
	656	15	5.0	60560	26	AC003678	Human Chromosome 11 pa 3.83e+02
c	657	15	5.0	61815	17	AC005453	*** SEQUENCING IN PROG 3.83e+02
	658	15	5.0	62257	17	AC005469	*** SEQUENCING IN PROG 3.83e+02
	659	15	5.0	63178	18	AC004333	Drosophila melanogaste 3.83e+02
	660	15	5.0	65691	18	PFMAL3P1	Plasmodium falciparum 3.83e+02
	661	15	5.0	68464	26	AC005248	Homo sapiens BAC clone 3.83e+02
c	662	15	5.0	71803	17	AC005504	*** SEQUENCING IN PROG 3.83e+02
	663	15	5.0	73350	18	AC002493	Drosophila melanogaste 3.83e+02
	664	15	5.0	73470	18	AC005434	Drosophila melanogaste 3.83e+02
c	665	15	5.0	74951	26	AC005219	Homo sapiens chromosom 3.83e+02
c	666	15	5.0	76848	17	PFMAL1P2	Plasmodium falciparum 3.83e+02
	667	15	5.0	78054	18	AC004296	Drosophila melanogaste 3.83e+02
	668	15	5.0	79676	24	AB013396	Arabidopsis thaliana g 3.83e+02
	669	15	5.0	80497	26	HS27K14	Human DNA sequence fro 3.83e+02
c	670	15	5.0	81370	24	AB017071	Arabidopsis thaliana g 3.83e+02
	671	15	5.0	81602	18	AC004349	Drosophila melanogaste 3.83e+02
c	672	15	5.0	81662	23	AB008265	Arabidopsis thaliana g 3.83e+02
	673	15	5.0	81826	25	HSU78045	Human collagenase and 3.83e+02
	674	15	5.0	82315	23	AB006707	Arabidopsis thaliana g 3.83e+02
	675	15	5.0	82594	17	AC005311	*** SEQUENCING IN PROG 3.83e+02
	676	15	5.0	83698	23	AB010072	Arabidopsis thaliana g 3.83e+02
	677	15	5.0	84124	17	HSAJ9614	Homo sapiens *** SEQUE 3.83e+02
	678	15	5.0	84883	18	AC004118	Drosophila melanogaste 3.83e+02
	679	15	5.0	85338	18	AC004279	Drosophila melanogaste 3.83e+02
	680	15	5.0	85372	17	AC005508	*** SEQUENCING IN PROG 3.83e+02
c	681	15	5.0	85757	18	AC001652	Drosophila melanogaste 3.83e+02
c	682	15	5.0	85837	23	SPU33010	Schizosaccharomyces po 3.83e+02
c	683	15	5.0	86080	26	HS337018	Homo sapiens DNA seque 3.83e+02
	684	15	5.0	86156	26	HUAC002550	Human Chromosome 16 BA 3.83e+02
c	685	15	5.0	86340	17	AC005441	*** SEQUENCING IN PROG 3.83e+02
	686	15	5.0	86380	23	AB010697	Arabidopsis thaliana g 3.83e+02
c	687	15	5.0	87557	17	AC005431	*** SEQUENCING IN PROG 3.83e+02
c	688	15	5.0	92802	17	AC004710	*** SEQUENCING IN PROG 3.83e+02
	689	15	5.0	93008	17	AC005120	*** SEQUENCING IN PROG 3.83e+02
	690	15	5.0	93273	26	AC004084	Homo sapiens BAC clone 3.83e+02
	691	15	5.0	93578	25	HSAC000055	Human PAC clone DJ073F 3.83e+02
c	692	15	5.0	95417	24	ATT16H5	Arabidopsis thaliana D 3.83e+02
	693	15	5.0	95655	17	AC005629	*** SEQUENCING IN PROG 3.83e+02
c	694	15	5.0	95742	26	AC005161	Homo sapiens BAC clone 3.83e+02
	695	15	5.0	95855	25	HSAC000115	Human BAC clone GS188P 3.83e+02
c	696	15	5.0	96625	25	HSAC000118	Human BAC clone RG072E 3.83e+02
	697	15	5.0	98734	23	ATAC002333	Arabidopsis thaliana c 3.83e+02
	698	15	5.0	98913	24	AF069442	Arabidopsis thaliana B 3.83e+02
	699	15	5.0	100000	26	AP000053	Homo sapiens genomic D 3.83e+02
	700	15	5.0	100000	26	AP000049	Homo sapiens genomic D 3.83e+02
c	701	15	5.0	100000	26	AP000044	Homo sapiens genomic D 3.83e+02
c	702	15	5.0	100000	26	AP000015	Homo sapiens genomic D 3.83e+02
c	703	15	5.0	101076	25	HS37M17	Human DNA sequence fro 3.83e+02
	704	15	5.0	102473	17	AC002320	*** SEQUENCING IN PROG 3.83e+02
	705	15	5.0	102857	6	AC004781	Drosophila melanogaste 3.83e+02

	706	15	5.0	103391	17	HS892F13	Human DNA sequence *** 3.83e+02
c	707	15	5.0	103682	23	SCU12980	Saccharomyces cerevisi 3.83e+02
c	708	15	5.0	104016	24	ATAC004138	Arabidopsis thaliana c 3.83e+02
	709	15	5.0	105263	17	AC005700	*** SEQUENCING IN PROG 3.83e+02
	710	15	5.0	106172	26	AC004111	Homo sapiens BAC clone 3.83e+02
	711	15	5.0	106319	26	HS747L4	Homo sapiens DNA seque 3.83e+02
	712	15	5.0	106377	26	AC004132	Homo sapiens chromosom 3.83e+02
c	713	15	5.0	107155	27	AL422092	Human DNA sequence *** 3.83e+02
	714	15	5.0	107677	26	HS440G21	Human DNA sequence fro 3.83e+02
c	715	15	5.0	108778	26	AC005194	Homo sapiens chromosom 3.83e+02
c	716	15	5.0	109279	26	HS394F12	Human DNA sequence fro 3.83e+02
	717	15	5.0	109359	27	AC004857	Homo sapiens PAC clone 3.83e+02
	718	15	5.0	109407	17	AC005694	*** SEQUENCING IN PROG 3.83e+02
c	719	15	5.0	109866	26	AF043945	Homo sapiens chromosom 3.83e+02
	720	15	5.0	110000	17	CEY39A1	Caenorhabditis elegans 3.83e+02
	721	15	5.0	110000	17	CEY37H9	Caenorhabditis elegans 3.83e+02
	722	15	5.0	110000	17	CEY38E10	Caenorhabditis elegans 3.83e+02
c	723	15	5.0	110000	17	CEY38E10	Caenorhabditis elegans 3.83e+02
c	724	15	5.0	110209	26	HUAC004533	Homo sapiens Chromosom 3.83e+02
c	725	15	5.0	110804	26	AC004748	Homo sapiens chromosom 3.83e+02
c	726	15	5.0	112400	17	HS30A23	Human DNA sequence *** 3.83e+02
c	727	15	5.0	114330	27	AC004772	Homo sapiens chromosom 3.83e+02
	728	15	5.0	114495	17	HS112K5	Human DNA sequence *** 3.83e+02
c	729	15	5.0	114621	27	AC005187	Homo sapiens chromosom 3.83e+02
	730	15	5.0	115641	24	F8K4	Arabidopsis thaliana c 3.83e+02
c	731	15	5.0	115641	24	F8K4	Arabidopsis thaliana c 3.83e+02
	732	15	5.0	115905	26	HS573E3	Homo sapiens DNA seque 3.83e+02
	733	15	5.0	117721	27	HS113J7	Human DNA sequence fro 3.83e+02
c	734	15	5.0	117994	17	HS409O10	Human DNA sequence *** 3.83e+02
	735	15	5.0	118684	26	AC004259	Human Chromosome 15q11 3.83e+02
	736	15	5.0	119707	23	PINCPTRPG	Pinus thunbergii chlor 3.83e+02
	737	15	5.0	119865	28	AC004407	Mus musculus mbac6, co 3.83e+02
	738	15	5.0	119903	27	AC004662	Homo sapiens chromosom 3.83e+02
c	739	15	5.0	120213	17	AC004150	*** SEQUENCING IN PROG 3.83e+02
	740	15	5.0	120787	23	AC002292	Genomic sequence of Ar 3.83e+02
c	741	15	5.0	121131	26	HS345B16	Human DNA sequence fro 3.83e+02
	742	15	5.0	121236	17	AC005097	*** SEQUENCING IN PROG 3.83e+02
	743	15	5.0	121780	27	HS774I24	Human DNA sequence fro 3.83e+02
c	744	15	5.0	121804	17	HS229A8	Human DNA sequence *** 3.83e+02
c	745	15	5.0	122000	17	AC004050	*** SEQUENCING IN PROG 3.83e+02
c	746	15	5.0	123193	17	AC005060	*** SEQUENCING IN PROG 3.83e+02
	747	15	5.0	124000	17	AC004060	*** SEQUENCING IN PROG 3.83e+02
c	748	15	5.0	124836	26	AC003085	Human BAC clone RG094H 3.83e+02
	749	15	5.0	125179	24	ATT4L20	Arabidopsis thaliana D 3.83e+02
c	750	15	5.0	125704	17	HS505B13	Human DNA sequence *** 3.83e+02
	751	15	5.0	126365	27	AC000026	Homo sapiens Chromosom 3.83e+02
c	752	15	5.0	126579	26	AC004692	Homo sapiens PAC clone 3.83e+02
	753	15	5.0	127039	18	AC005732	Drosophila melanogaste 3.83e+02
	754	15	5.0	127247	25	HS326L13	Human DNA sequence fro 3.83e+02
	755	15	5.0	127356	17	AC005063	*** SEQUENCING IN PROG 3.83e+02
	756	15	5.0	127816	17	AC004822	*** SEQUENCING IN PROG 3.83e+02

	757	15	5.0	128230	26	HS313L4	Human DNA sequence fro 3.83e+02
c	758	15	5.0	128499	17	AC005480	*** SEQUENCING IN PROG 3.83e+02
	759	15	5.0	128779	26	HS29K1	Human DNA sequence fro 3.83e+02
	760	15	5.0	129079	17	AC004955	*** SEQUENCING IN PROG 3.83e+02
	761	15	5.0	129252	26	HSU91327	Human chromosome 12p15 3.83e+02
c	762	15	5.0	129429	17	AC004951	*** SEQUENCING IN PROG 3.83e+02
c	763	15	5.0	130064	17	AC002509	*** SEQUENCING IN PROG 3.83e+02
	764	15	5.0	130076	17	HS510D11	Human DNA sequence *** 3.83e+02
	765	15	5.0	130381	26	AC002402	Human Chromosome 11 pa 3.83e+02
	766	15	5.0	130513	26	AC004592	Homo sapiens PAC clone 3.83e+02
c	767	15	5.0	131047	17	AC002421	*** SEQUENCING IN PROG 3.83e+02
c	768	15	5.0	132698	17	AC005051	*** SEQUENCING IN PROG 3.83e+02
c	769	15	5.0	132953	26	AC003015	Human BAC clone GS113H 3.83e+02
c	770	15	5.0	133798	17	AC004950	*** SEQUENCING IN PROG 3.83e+02
c	771	15	5.0	134488	17	AC005527	*** SEQUENCING IN PROG 3.83e+02
c	772	15	5.0	134506	27	AC005738	Homo sapiens chromosom 3.83e+02
	773	15	5.0	134534	17	HS1163J1	Human DNA sequence *** 3.83e+02
	774	15	5.0	134970	26	AC002504	Human Xp22 PAC RPCI1-2 3.83e+02
c	775	15	5.0	134970	26	AC002504	Human Xp22 PAC RPCI1-2 3.83e+02
c	776	15	5.0	135409	27	AC000025	Homo sapiens Chromosom 3.83e+02
	777	15	5.0	135705	17	HS97P20	Human DNA sequence *** 3.83e+02
	778	15	5.0	135891	17	HS221G9	Human DNA sequence *** 3.83e+02
c	779	15	5.0	136755	26	AC002529	Human BAC clone RG161N 3.83e+02
	780	15	5.0	137129	17	HSAC001236	*** SEQUENCING IN PROG 3.83e+02
	781	15	5.0	137176	26	AC005017	Homo sapiens BAC clone 3.83e+02
c	782	15	5.0	137191	17	AC005476	*** SEQUENCING IN PROG 3.83e+02
	783	15	5.0	137947	27	HS45I4	Human DNA sequence *** 3.83e+02
c	784	15	5.0	138313	17	AC004848	*** SEQUENCING IN PROG 3.83e+02
c	785	15	5.0	138343	17	AC004907	*** SEQUENCING IN PROG 3.83e+02
c	786	15	5.0	138820	17	CEY67A6	Caenorhabditis elegans 3.83e+02
	787	15	5.0	139042	17	AC004600	*** SEQUENCING IN PROG 3.83e+02
	788	15	5.0	139350	26	HS443E24	Human DNA sequence fro 3.83e+02
	789	15	5.0	139642	17	HS20N2	Human DNA sequence *** 3.83e+02
	790	15	5.0	139887	25	HUB18A8	Chromosome 22q13 BAC C 3.83e+02
	791	15	5.0	140896	17	HS163G9	Human DNA sequence *** 3.83e+02
	792	15	5.0	140977	26	AC002536	Human Chromosome 11 pa 3.83e+02
	793	15	5.0	141990	26	AC004691	Homo sapiens PAC clone 3.83e+02
c	794	15	5.0	141990	26	AC004691	Homo sapiens PAC clone 3.83e+02
c	795	15	5.0	142396	25	HSAC000120	Human BAC clone RG161K 3.83e+02
c	796	15	5.0	143319	17	HS100N22	Human DNA sequence *** 3.83e+02
	797	15	5.0	143701	27	HS29C18	Homo sapiens DNA seque 3.83e+02
c	798	15	5.0	144184	17	HS167P19	Human DNA sequence *** 3.83e+02
c	799	15	5.0	144227	26	HS119E23	Homo sapiens DNA seque 3.83e+02
c	800	15	5.0	144349	17	HS202D8	Human DNA sequence *** 3.83e+02
c	801	15	5.0	144368	26	AC004103	Homo sapiens Xp22 BAC 3.83e+02
	802	15	5.0	144545	17	AC005287	*** SEQUENCING IN PROG 3.83e+02
c	803	15	5.0	145247	17	AC004886	*** SEQUENCING IN PROG 3.83e+02
c	804	15	5.0	145831	27	HUAP001548	Human Chromosome 16 BA 3.83e+02
c	805	15	5.0	145871	26	HS302D9	Human DNA sequence fro 3.83e+02
c	806	15	5.0	145961	17	AC005805	*** SEQUENCING IN PROG 3.83e+02
c	807	15	5.0	146560	17	AC004998	*** SEQUENCING IN PROG 3.83e+02

c	808	15	5.0	146828	17	AC004968	*** SEQUENCING IN PROG 3.83e+02
c	809	15	5.0	147898	17	HS1111N9	Human DNA sequence *** 3.83e+02
	810	15	5.0	149322	17	HS1119D9	Human DNA sequence *** 3.83e+02
c	811	15	5.0	149526	26	HS41I86	Human DNA sequence fro 3.83e+02
	812	15	5.0	149752	26	AC004616	Homo sapiens Xp22 BAC 3.83e+02
	813	15	5.0	149765	26	AC004984	Homo sapiens PAC clone 3.83e+02
	814	15	5.0	150228	26	AC003071	Human BAC clone BK085E 3.83e+02
c	815	15	5.0	150620	25	HS203C2	Human DNA sequence fro 3.83e+02
c	816	15	5.0	151018	17	PFMAL1P6	Plasmodium falciparum 3.83e+02
c	817	15	5.0	151111	26	AC002308	Homo sapiens Chromosom 3.83e+02
	818	15	5.0	151166	17	AC005520	*** SEQUENCING IN PROG 3.83e+02
	819	15	5.0	151822	26	AC004478	Homo Sapiens Chromosom 3.83e+02
	820	15	5.0	151874	26	AC004594	Homo sapiens PAC clone 3.83e+02
c	821	15	5.0	152115	17	HS404K8	Human DNA sequence *** 3.83e+02
	822	15	5.0	152184	17	HS593P22	Human DNA sequence *** 3.83e+02
c	823	15	5.0	152386	17	AC005711	*** SEQUENCING IN PROG 3.83e+02
	824	15	5.0	152592	17	HS396D17	Human DNA sequence *** 3.83e+02
	825	15	5.0	153019	17	AC004969	*** SEQUENCING IN PROG 3.83e+02
c	826	15	5.0	153019	17	AC004969	*** SEQUENCING IN PROG 3.83e+02
	827	15	5.0	153735	26	AC004029	Human BAC clone RG035E 3.83e+02
	828	15	5.0	153773	26	HUAC004020	Homo sapiens Chromosom 3.83e+02
	829	15	5.0	154066	26	HSB24018	Human DNA sequence *** 3.83e+02
	830	15	5.0	154985	18	AC004641	Drosophila melanogaste 3.83e+02
	831	15	5.0	155176	26	HS453P22	Human DNA sequence fro 3.83e+02
	832	15	5.0	155407	26	AF042090	Homo sapiens chromosom 3.83e+02
	833	15	5.0	156214	26	HSAC002064	Human BAC clone RG016J 3.83e+02
c	834	15	5.0	156601	26	HS212G6	Homo sapiens DNA seque 3.83e+02
	835	15	5.0	158829	17	CEY68A4	Caenorhabditis elegans 3.83e+02
c	836	15	5.0	158829	17	CEY68A4	Caenorhabditis elegans 3.83e+02
	837	15	5.0	159500	28	AC005742	Mus musculus chromosom 3.83e+02
	838	15	5.0	159863	26	AC002349	Homo sapiens Xp22 PAC 3.83e+02
c	839	15	5.0	160218	17	HS500L14	Human DNA sequence *** 3.83e+02
c	840	15	5.0	160986	17	AC005505	*** SEQUENCING IN PROG 3.83e+02
	841	15	5.0	161742	17	AC005519	*** SEQUENCING IN PROG 3.83e+02
c	842	15	5.0	162671	26	AC002352	Homo sapiens 12q24 PAC 3.83e+02
c	843	15	5.0	163908	27	AC005411	Homo sapiens chromosom 3.83e+02
c	844	15	5.0	163974	27	AB011399	Homo sapiens gene for 3.83e+02
c	845	15	5.0	164074	17	AC005803	*** SEQUENCING IN PROG 3.83e+02
	846	15	5.0	164536	17	HS134N8	Human DNA sequence *** 3.83e+02
c	847	15	5.0	165015	17	HS1158E12	Human DNA sequence *** 3.83e+02
	848	15	5.0	165228	26	AC005325	Homo sapiens chromosom 3.83e+02
c	849	15	5.0	165382	26	AF064860	Homo sapiens chromosom 3.83e+02
	850	15	5.0	165507	17	AC004903	*** SEQUENCING IN PROG 3.83e+02
	851	15	5.0	165608	26	AC004492	Homo sapiens BAC clone 3.83e+02
c	852	15	5.0	166001	17	AC005225	*** SEQUENCING IN PROG 3.83e+02
c	853	15	5.0	166219	26	AC003084	Human BAC clone RG084D 3.83e+02
c	854	15	5.0	167372	27	AC004945	Homo sapiens PAC clone 3.83e+02
	855	15	5.0	169283	17	AC004935	*** SEQUENCING IN PROG 3.83e+02
	856	15	5.0	169465	26	HSU96409	Human Xp22 contig of 3 3.83e+02
c	857	15	5.0	171101	17	AC005666	*** SEQUENCING IN PROG 3.83e+02
	858	15	5.0	171147	17	HS1059H15	Human DNA sequence *** 3.83e+02

c	859	15	5.0	171258	17	HS873P14	Human DNA sequence *** 3.83e+02
c	860	15	5.0	172000	16	AP000005	Pyrococcus horikoshii 3.83e+02
	861	15	5.0	172995	17	AC004926	*** SEQUENCING IN PROG 3.83e+02
	862	15	5.0	173034	26	AC002059	Homo sapiens Chromosom 3.83e+02
c	863	15	5.0	173494	17	HS662M3	Human DNA sequence *** 3.83e+02
	864	15	5.0	173717	27	AC005695	Homo sapiens chromosom 3.83e+02
	865	15	5.0	175270	17	HS652L8	Human DNA sequence *** 3.83e+02
	866	15	5.0	175405	17	AC005589	*** SEQUENCING IN PROG 3.83e+02
c	867	15	5.0	176154	17	AC005317	*** SEQUENCING IN PROG 3.83e+02
	868	15	5.0	176511	27	AC003035	Homo sapiens Xp22 bins 3.83e+02
	869	15	5.0	176714	27	AC005618	Homo sapiens chromosom 3.83e+02
c	870	15	5.0	177311	26	AC004385	Human Chromosome X clo 3.83e+02
c	871	15	5.0	177328	17	HS462D8	Human DNA sequence *** 3.83e+02
c	872	15	5.0	177821	17	AC005030	*** SEQUENCING IN PROG 3.83e+02
c	873	15	5.0	181795	26	AC003685	Homo sapiens Xp22 BAC 3.83e+02
	874	15	5.0	182329	17	AC005022	*** SEQUENCING IN PROG 3.83e+02
c	875	15	5.0	183315	17	HS90G24	Human DNA sequence *** 3.83e+02
	876	15	5.0	183338	17	HS329A5	Human DNA sequence *** 3.83e+02
	877	15	5.0	185516	26	AC004485	Homo sapiens BAC clone 3.83e+02
	878	15	5.0	186454	17	AC005678	*** SEQUENCING IN PROG 3.83e+02
	879	15	5.0	188251	17	AC005041	*** SEQUENCING IN PROG 3.83e+02
c	880	15	5.0	189357	17	HS596H12	Human DNA sequence *** 3.83e+02
	881	15	5.0	189476	17	AC004976	*** SEQUENCING IN PROG 3.83e+02
c	882	15	5.0	189959	17	AC005023	*** SEQUENCING IN PROG 3.83e+02
	883	15	5.0	190000	17	AC005598	*** SEQUENCING IN PROG 3.83e+02
c	884	15	5.0	190000	17	AC005150	*** SEQUENCING IN PROG 3.83e+02
	885	15	5.0	191191	25	HSAC000119	Human BAC clone RG104I 3.83e+02
c	886	15	5.0	191908	17	CEY51B9	Caenorhabditis elegans 3.83e+02
	887	15	5.0	194985	28	AC002406	Mouse chromosome X BAC 3.83e+02
	888	15	5.0	196559	17	HS399M14	Human DNA sequence *** 3.83e+02
c	889	15	5.0	196672	18	AC004767	Drosophila melanogaste 3.83e+02
c	890	15	5.0	196836	17	AC004836	*** SEQUENCING IN PROG 3.83e+02
c	891	15	5.0	197273	26	AC002461	Human BAC clone RG204I 3.83e+02
c	892	15	5.0	197620	26	AF001905	Homo sapiens cosmids E 3.83e+02
c	893	15	5.0	199082	17	AC005632	*** SEQUENCING IN PROG 3.83e+02
c	894	15	5.0	199927	17	HS101A4	Human DNA sequence *** 3.83e+02
	895	15	5.0	200000	17	AC004661	*** SEQUENCING IN PROG 3.83e+02
c	896	15	5.0	200000	17	AC004664	*** SEQUENCING IN PROG 3.83e+02
c	897	15	5.0	202096	17	AC005036	*** SEQUENCING IN PROG 3.83e+02
c	898	15	5.0	204596	17	AC005069	*** SEQUENCING IN PROG 3.83e+02
c	899	15	5.0	206214	17	AC005013	*** SEQUENCING IN PROG 3.83e+02
c	900	15	5.0	208230	15	BSUB0008	Bacillus subtilis comp 3.83e+02
	901	15	5.0	209617	17	HS54B20	Human DNA sequence *** 3.83e+02
	902	15	5.0	212610	15	BSUB0019	Bacillus subtilis comp 3.83e+02
c	903	15	5.0	212610	15	BSUB0019	Bacillus subtilis comp 3.83e+02
c	904	15	5.0	213420	15	BSUB0014	Bacillus subtilis comp 3.83e+02
c	905	15	5.0	213633	26	HUAC002288	Homo sapiens Chromosom 3.83e+02
	906	15	5.0	213863	27	AC003037	Homo sapiens Xp22 BAC 3.83e+02
c	907	15	5.0	220731	17	AC004671	*** SEQUENCING IN PROG 3.83e+02
c	908	15	5.0	222922	17	AC005015	*** SEQUENCING IN PROG 3.83e+02
c	909	15	5.0	224645	17	AC004153	*** SEQUENCING IN PROG 3.83e+02

c	910	15	5.0	227538	28	AC002397	Mouse chromosome 6 BAC 3.83e+02
c	911	15	5.0	228500	17	HS212A2	Human DNA sequence *** 3.83e+02
c	912	15	5.0	233780	15	BSUB0010	Bacillus subtilis comp 3.83e+02
	913	15	5.0	239566	26	HSAJ03147	Homo sapiens complete 3.83e+02
c	914	15	5.0	243041	26	AC004388	Homo Sapiens Chromosom 3.83e+02
c	915	15	5.0	250599	26	AF015262	Homo sapiens Down Synd 3.83e+02
	916	15	5.0	251199	26	HUAE000659	Homo sapiens T-cell re 3.83e+02
	917	15	5.0	253516	17	CEY6B3	Caenorhabditis elegans 3.83e+02
	918	15	5.0	257000	16	AP000002	Pyrococcus horikoshii 3.83e+02
c	919	15	5.0	259894	26	HUAC002302	Homo sapiens Chromosom 3.83e+02
c	920	15	5.0	268399	17	AC004469	*** SEQUENCING IN PROG 3.83e+02
	921	15	5.0	269078	17	CEY76A2	Caenorhabditis elegans 3.83e+02
c	922	15	5.0	275138	17	CEY54B5	Caenorhabditis elegans 3.83e+02
c	923	15	5.0	288002	26	HS229043	Homo sapiens 959 kb co 3.83e+02
	924	15	5.0	302988	17	CEY46G5	Caenorhabditis elegans 3.83e+02
c	925	15	5.0	302988	17	CEY46G5	Caenorhabditis elegans 3.83e+02
c	926	15	5.0	307360	17	HS796E4	Human DNA sequence *** 3.83e+02
c	927	15	5.0	310126	17	CEY48B6	Caenorhabditis elegans 3.83e+02
	928	15	5.0	312429	17	AC005529	*** SEQUENCING IN PROG 3.83e+02
c	929	15	5.0	335867	17	CEY87G2	Caenorhabditis elegans 3.83e+02
	930	14	4.7	160	25	HS21B11R	H.sapiens CpG island D 1.63e+03
c	931	14	4.7	225	26	HSA230087	Homo sapiens rearrange 1.63e+03
c	932	14	4.7	259	26	HSA230096	Homo sapiens rearrange 1.63e+03
c	933	14	4.7	374	33	SVU27795	Stealth virus clone C1 1.63e+03
c	934	14	4.7	420	25	HSTCRB13	Human mRNA for T-cell 1.63e+03
c	935	14	4.7	510	16	AF082303	Azospirillum brasilens 1.63e+03
c	936	14	4.7	600	23	CHSOTRNI	Spinach chloroplast tr 1.63e+03
c	937	14	4.7	743	20	RTCRYB23	Rana temporaria mRNA f 1.63e+03
c	938	14	4.7	1124	25	HSIL1BR	Human mRNA for interle 1.63e+03
	939	14	4.7	1250	28	MMCD40	M.musculus mRNA for CD 1.63e+03
c	940	14	4.7	1367	33	PSTVCP	Peanut stripe virus ge 1.63e+03
c	941	14	4.7	1657	33	VSVGLYPY	Vesicular stomatitis v 1.63e+03
c	942	14	4.7	1657	33	VSVGLYPA	Vesicular stomatitis v 1.63e+03
c	943	14	4.7	2066	23	NRPHOSLPC	N.rustica mRNA for pho 1.63e+03
	944	14	4.7	2583	24	TSPGPH	Schizosaccharomyces po 1.63e+03
c	945	14	4.7	2682	25	HUMGTRAN	Human protein-glutamin 1.63e+03
	946	14	4.7	2937	20	BRSVPA	B.rerio svp 44 mRNA fo 1.63e+03
c	947	14	4.7	3168	33	HVSX99518	Herpesvirus saimiri vi 1.63e+03
	948	14	4.7	3187	28	MUSOSF2	Mouse mRNA for osteobl 1.63e+03
c	949	14	4.7	3283	33	HPU22461	Human papillomavirus t 1.63e+03
c	950	14	4.7	5089	33	PPMCG	Monkey B-lymphotropic 1.63e+03
	951	14	4.7	5926	26	HSU48959	Homo sapiens myosin li 1.63e+03
c	952	14	4.7	6030	28	AF020345	Cavia porcellus cytoch 1.63e+03
c	953	14	4.7	7180	33	HPENSSP	Hepatitis E virus (mex 1.63e+03
c	954	14	4.7	7824	25	HUMIL1B	Human interleukin 1-be 1.63e+03
c	955	14	4.7	7905	33	HPU21941	Human papillomavirus t 1.63e+03
	956	14	4.7	9610	20	AF027728	Xenopus laevis kinesin 1.63e+03
	957	14	4.7	9645	26	HSU27266	Human myosin binding p 1.63e+03
c	958	14	4.7	16519	28	MMPAFALIS4	Mus musculus Pafaha/Li 1.63e+03
	959	14	4.7	17013	15	EAAMASL	E.amylovara (Ea7/74) a 1.63e+03
	960	14	4.7	26962	18	CEZK836	Caenorhabditis elegans 1.63e+03

c	961	14	4.7	27717	18	CEP34H10	Caenorhabditis elegans 1.63e+03
	962	14	4.7	35397	18	CELF37B4	Caenorhabditis elegans 1.63e+03
c	963	14	4.7	39470	26	AC005390	Homo sapiens chromosom 1.63e+03
c	964	14	4.7	39595	26	AC004509	Homo sapiens chromosom 1.63e+03
	965	14	4.7	39874	18	CEZK287	Caenorhabditis elegans 1.63e+03
c	966	14	4.7	41511	18	CELC26B2	Caenorhabditis elegans 1.63e+03
c	967	14	4.7	42759	24	AB004535	Schizosaccharomyces po 1.63e+03
c	968	14	4.7	47812	27	HS127L4	Human DNA sequence fro 1.63e+03
	969	14	4.7	55529	17	HS390E6	Human DNA sequence *** 1.63e+03
c	970	14	4.7	56500	26	AB013139	Homo sapiens gene for 1.63e+03
c	971	14	4.7	65589	27	AF049895	Homo sapiens nibrin (N 1.63e+03
c	972	14	4.7	65691	18	PFMAL3P1	Plasmodium falciparum 1.63e+03
	973	14	4.7	77612	17	AC004726	*** SEQUENCING IN PROG 1.63e+03
	974	14	4.7	87557	17	AC005431	*** SEQUENCING IN PROG 1.63e+03
	975	14	4.7	92128	24	F17O7	Arabidopsis thaliana c 1.63e+03
	976	14	4.7	93904	26	AC004128	Homo sapiens PAC clone 1.63e+03
c	977	14	4.7	100267	25	HSMHC3W36A	Human HLA class III re 1.63e+03
c	978	14	4.7	108754	26	AC004226	Homo sapiens chromosom 1.63e+03
	979	14	4.7	122638	26	AB001523	Homo sapiens gene for 1.63e+03
	980	14	4.7	123203	26	AC003036	Homo sapiens Xp22 PAC 1.63e+03
	981	14	4.7	125930	17	HS124K22	Human DNA sequence *** 1.63e+03
	982	14	4.7	134450	26	HUAC003049	Homo sapiens Chromosom 1.63e+03
	983	14	4.7	139666	18	AC004545	Drosophila melanogaste 1.63e+03
c	984	14	4.7	140823	17	HS1104E15	Human DNA sequence *** 1.63e+03
c	985	14	4.7	150000	17	AC004701	*** SEQUENCING IN PROG 1.63e+03
c	986	14	4.7	156641	26	HUAC002544	Homo sapiens Chromosom 1.63e+03
	987	14	4.7	162485	17	AC004840	*** SEQUENCING IN PROG 1.63e+03
	988	14	4.7	163023	17	AC004815	*** SEQUENCING IN PROG 1.63e+03
	989	14	4.7	164296	17	HSAC000380	*** SEQUENCING IN PROG 1.63e+03
c	990	14	4.7	172241	26	AC004014	Human BAC clone RG022K 1.63e+03
c	991	14	4.7	174582	26	HUAC003010	Homo sapiens Chromosom 1.63e+03
c	992	14	4.7	175405	17	AC005589	*** SEQUENCING IN PROG 1.63e+03
c	993	14	4.7	180000	17	AC004578	*** SEQUENCING IN PROG 1.63e+03
	994	14	4.7	189370	15	AF010496	Rhodobacter capsulatus 1.63e+03
	995	14	4.7	200000	17	AC004795	*** SEQUENCING IN PROG 1.63e+03
c	996	14	4.7	200000	23	MIATGENA	A.thaliana mitochondri 1.63e+03
c	997	14	4.7	200000	17	AC005289	*** SEQUENCING IN PROG 1.63e+03
c	998	14	4.7	222930	26	HSU47924	Human chromosome 12p13 1.63e+03
	999	14	4.7	250529	26	HUAE000658	Homo sapiens T-cell re 1.63e+03
	1000	14	4.7	250611	28	MMAE000663	Mus musculus TCR beta 1.63e+03

ALIGNMENTS

```
RESULT    1
LOCUS       AF051782     5635 bp    mRNA            PRI       11-MAR-1998
DEFINITION  Homo sapiens diaphanous 1 (HDIA1) mRNA, complete cds.
ACCESSION   AF051782
NID         g2947237
KEYWORDS    .
SOURCE      human.
```



```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5635)
  AUTHORS   Leon,P.E., Raventos,H., Lynch,E., Morrow,J. and King,M.C.
  TITLE     The gene for an inherited form of deafness maps to chromosome 5q31
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 89 (11), 5181-5184 (1992)
  MEDLINE   92279290
REFERENCE   2  (bases 1 to 5635)
  AUTHORS   Lynch,E.D., Lee,M.K., Morrow,J.E., Welcsh,P.L., Leon,P.E. and
            King,M.C.
  TITLE     Nonsyndromic deafness DFNA1 associated with mutation of a human
            homolog of the Drosophila gene diaphanous
  JOURNAL   Science 278 (5341), 1315-1318 (1997)
  MEDLINE   98028756
REFERENCE   3  (bases 1 to 5635)
  AUTHORS   Lynch,E.D., Lee,M.K. and King,M.-C.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-FEB-1998) Medical Genetics, University of Washington,
            Box 357720, Seattle, WA 98195-7720, USA
FEATURES              Location/Qualifiers
     source           1. .5635
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="5"
                      /map="5q31"
     gene             1. .5635
                      /note="mutation associated with nonsyndromic deafness
                      DFNA1"
                      /gene="HDIA1"
     CDS              1. .3747
                      /gene="HDIA1"
                      /note="similar to Drosophila diaphanous"
                      /codon_start=1
                      /product="diaphanous 1"
                      /db_xref="PID:g2947238"
                      /translation="MEPPGGSLGPGRETRDKKKGRSPDELPSAGGDGGKSKKFLERFT
                      SMRIKKEKEKPNSAHRNSSASYGDDPTAQSLQDVSDEQVLVLFEQMLLDMNLNEEKQQ
                      PLREKDIIIKREMVSQYLYTSKAGMSQKESSKSAMMYIQELRSGLRDMPLLSCLESLR
                      VSLNNNPVSWVQTFGAEGLASLLDILKRLHDEKEETAGSYDSRNKHEIIRCLKAFMNN
                      KFGIKTHLETEEGILLLVRAMDPAVPNMMIDAAKLLSALCILPQPEDMNERVLEAMTE
                      RAEMDEVERPQPLLDGLKSGTTIALKVGCLQLINALITPAEELDFRVHIRSELMRLGL
                      HQVLQDLREIENEDMRVQLNVFDEQGEEDSYDLKGRLDDIRMEMDDFNEVFQILLNTV
                      KDSKAEPHFLSILQHLLLVRNDYEARPQYYKLIEECISQIVLHKNGADPDFKCRHLQI
                      EIEGLIDQMIDKTKVEKSEAKAAELEKKLDSELTARHELQVEMKKMESDFEQKLQDLQ
                      GEKDALHSEKQQIATEKQDLEAEVSQLTGEVAKLTEELEDAKKEMASLGAAAITVPPS
                      VPSRAPVPPAPPLPGDSGTIIPPPPAPGDSTTPPPPPPPPPPPPPLPGGTAISPPPPL
                      SGDATIPPPPPLPEGVGIPSPSSLPGGTAIPPPPPLPGSARIPPPPPPLPGSAGIPPP
                      PPPLPGEAGMPPPPPPLPGGPGIPPPPPPFPGGPGIPPPPPPGMGMPPPPPFGFGVPAAP
                      VLPPGLTPKKLYKPEVQLRRPNWSKLVAEDLSQDCFWTKVKEDRFENNELFAKLTLTF
                      SAQTKTKKDQEGGEEKKSVQKKKVKELKVLDSKTAQNLSIFLGSFRMPYQEIKNVILE
```



```
                     VNEAVLTESMIQNLIKQMPEPEQLKMLSELKDEYDDLAESEQFGVVMGTVPRLRPRLN
                     AILFKLQFSEQVENIKPEIVSVTAACEELRKSESFSNLLEITLLVGNYMNAGSRNAGA
                     FGFNISFLCKLRDTKSTDQKMTLLHFLAELCENDYPDVLKFPDELAHVEKASRVSAEN
                     LQKNLDQMKKQISDVERDVQNFPAATDEKDKFVEKMTSFVKDAQEQYNKLRMMHSNME
                     TLYKELGEYFLFDPKKLSVEEFFMDLHNFRNMFLQAVKEMQKRRKTEEKMRRAKLAKE
                     KAEKERLEKQQKREQLIDMNAEGDETGVMDSLLELQSGAAPRKRGPRQANRKAGCA
                     VTSLLASELTKDDAMAAVPAKVSKNSETFPTILEEAKELVGRAS"
BASE COUNT     1432 a   1406 c   1421 g   1376 t
ORIGIN

  Query Match          74.8%;  Score 223;  DB 26;  Length 5635;
  Best Local Similarity  97.4%;  Pred. No. 1.52e-243;
  Matches   223;  Conservative    0;  Mismatches   6;  Indels   0;  Gaps   0;

Db    4885 GATCCTCTCTCCCAGGGAATCCGACACAGGAGGAACCCCTTCTCTGGTTGAGCTGGGCCA 4944
           ||||||||||||||||||||||||||||||||||||||||||| ||||||| ||||||||
Qy       1 GATCCTCTCTCCCAGGGAATCCGACACAGGAGGAACCCCTTCTNTGGTTGANCTGGGCCA 60

Db    4945 GGCCTAAGAGTAGCAGGAACTCTAAGACCACAGAGTTTTTTATAAATGTATAAATGTATC 5004
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      61 GGCCTAAGAGTAGCAGGAACTCTAAGACCACAGAGTTTTTTATAAATGTATAAATGTATC 120

Db    5005 AAGCCAAATGTGCAGATGCTAACTGGACATTCTGGGGAACTGGGCACCAGGAGTGCCTTC 5064
           ||||||||| |||||||||||||||||||||||||||||||||||||||||| ||||||||
Qy     121 AAGCCAAATNTGCAGATGCTAACTGGACATTCTGGGGAACTGGGCACCAGGNGTGCCTTC 180

Db    5065 ATACACTGTACCCCAGCTCTCTTCTAAAAGAGAAGTGGGTGGGCACACT 5113
           |||||||||||||||||| |||||||||||||| ||||||||||||
Qy     181 ATACACTGTACCCCAGCTCTNTTCTAAAAGAGAAGTNGGTGGGCACACT 229


RESULT    2
LOCUS       AC005366    68067 bp    DNA             PRI       01-AUG-1998
DEFINITION  Homo sapiens chromosome 5, BAC clone 45m2 (LBNL H145), complete
            sequence.
ACCESSION   AC005366
NID         g3367508
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 68067)
  AUTHORS   Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
            Kadner,K., Miguel,T., Miller,C., Pitluck,S., Pollard,M.,
            Rojeski,H., Subramanian,S. and Martin,C.H.
  TITLE     Sequencing of human chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 68067)
  AUTHORS   Ricke,D.O.
  TITLE     Large Scale Sequence Analysis and Annotation with the Sequence
```



```
            Comparison Analysis (SCAN) System
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 68067)
  AUTHORS   Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
            Davis,C.A., Kadner,K., Miguel,T., Pitluck,S., Pollard,M.,
            Rojeski,H., Subramanian,S. and Martin,C.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-AUG-1998) Human Genome Center, DOE Joint Genome
            Institute, Lawrence Berkeley National Laboratory, MS 74-157,
            Berkeley, CA 94720, U.S.A.
COMMENT     Sequence submitted by:
            DOE Joint Genome Institute.
FEATURES              Location/Qualifiers
     source           1. .68067
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /map="5q"
                      /clone="45m2"
                      /chromosome="5"
                      /note="LBNL H145"
     repeat_region    complement(860. .1135)
                      /rpt_family="Alu"
     misc_feature     1675. .1795
                      /note="GRAIL 2 excellent exon, frame 0"
     misc_feature     2007. .2246
                      /note="GRAIL 2 excellent exon, frame 2"
     misc_feature     2366. .2495
                      /note="GRAIL 2 excellent exon, frame 1"
     repeat_region    3226. .3248
                      /note="(T)23"
                      /rpt_type=tandem
                      /rpt_unit=T
     misc_feature     3251. .3433
                      /note="GRAIL 2 excellent exon, frame 0"
     repeat_region    complement(3529. .3738)
                      /rpt_family="Alu"
     repeat_region    4030. .4249
                      /rpt_family="Alu"
     misc_feature     4495. .4657
                      /note="GRAIL 2 excellent exon, frame 1"
     misc_feature     4788. .4916
                      /note="GRAIL 2 excellent exon, frame 1"
     misc_feature     6725. .6811
                      /note="GRAIL 2 excellent exon, frame 0"
     repeat_region    8356. .8377
                      /note="(T)22"
                      /rpt_type=tandem
                      /rpt_unit=T
     repeat_region    9855. .10057
                      /rpt_family="Alu"
     misc_feature     10148. .10275
```

```
                     /note="GRAIL 2 excellent exon, frame 0"
     repeat_region   complement(10288. .10358)
                     /rpt_family="THE1"
     repeat_region   complement(10549. .10646)
                     /rpt_family="THE1"
     repeat_region   12158. .12425
                     /rpt_family="Alu"
     misc_feature    13946. .14103
                     /note="GRAIL 2 excellent exon, frame 0"
     misc_feature    complement(19785. .20004)
                     /note="GRAIL 2 excellent exon, frame 2"
     repeat_region   20183. .20390
                     /rpt_family="MIR"
     repeat_region   complement(20664. .20875)
                     /rpt_family="Alu"
     repeat_region   complement(20866. .20968)
                     /rpt_family="MIR"
     repeat_region   24489. .24509
                     /note="(T)21"
                     /rpt_type=tandem
                     /rpt_unit=T
     repeat_region   complement(24492. .24781)
                     /rpt_family="Alu"
     repeat_region   complement(25029. .25230)
                     /rpt_family="MIR"
     misc_feature    complement(25473. .25561)
                     /note="GRAIL 2 excellent exon, frame 0"
     repeat_region   complement(25807. .25992)
                     /rpt_family="Alu"
     repeat_region   complement(27340. .27635)
                     /rpt_family="Alu"
     repeat_region   27852. .28087
                     /rpt_family="Alu"
     repeat_region   28491. .28780
                     /rpt_family="MER42"
     repeat_region   complement(29213. .29352)
                     /rpt_family="MIR"
     repeat_region   29620. .29657
                     /note="(GT)19"
                     /rpt_type=tandem
                     /rpt_unit=GT
     repeat_region   29791. .30070
                     /rpt_family="Alu"
     repeat_region   30445. .30708
                     /rpt_family="Alu"
     repeat_region   30722. .30761
                     /note="(AATA)10"
                     /rpt_type=tandem
                     /rpt_unit=AATA
     repeat_region   30961. .31253
                     /rpt_family="Alu"
```



```
     repeat_region   31237. .31258
                     /note="(A)22"
                     /rpt_type=tandem
                     /rpt_unit=A
     repeat_region   31920. .31971
                     /note="(CTTC)13"
                     /rpt_type=tandem
                     /rpt_unit=CTTC
     repeat_region   complement(32017. .32286)
                     /rpt_family="Alu"
     repeat_region   33109. .33403
                     /rpt_family="Alu"
     repeat_region   complement(33811. .34024)
                     /rpt_family="Alu"
     misc_feature    complement(39067. .41652)
                     /note="GRAIL 2 excellent exon, frame 0"
     misc_feature    complement(43155. .45770)
                     /note="GRAIL 2 excellent exon, frame 1"
     repeat_region   complement(47851. .48154)
                     /rpt_family="Alu"
     repeat_region   complement(48839. .48998)
                     /rpt_family="Alu"
     repeat_region   49009. .49030
                     /note="(T)22"
                     /rpt_type=tandem
                     /rpt_unit=T
     repeat_region   complement(49013. .49425)
                     /rpt_family="Alu"
     repeat_region   complement(49876. .50143)
                     /rpt_family="Alu"
     repeat_region   complement(50519. .50839)
                     /rpt_family="Alu"
     misc_feature    complement(52236. .54827)
                     /note="GRAIL 2 excellent exon, frame 1"
     misc_feature    complement(55687. .55781)
                     /note="GRAIL 2 excellent exon, frame 0"
     repeat_region   56506. .56645
                     /rpt_family="MER5"
     repeat_region   complement(56856. .57002)
                     /rpt_family="Alu"
     repeat_region   complement(57947. .58229)
                     /rpt_family="Alu"
     repeat_region   complement(58392. .58683)
                     /rpt_family="Alu"
     repeat_region   59594. .59895
                     /rpt_family="Alu"
     repeat_region   59878. .59897
                     /note="(A)20"
                     /rpt_type=tandem
                     /rpt_unit=A
     repeat_region   59939. .60240
```



```
                     /rpt_family="Alu"
     repeat_region   60242. .60355
                     /rpt_family="MER45"
     repeat_region   complement(60807. .61083)
                     /rpt_family="Alu"
     repeat_region   complement(61377. .61941)
                     /rpt_family="Alu"
     repeat_region   complement(62074. .62355)
                     /rpt_family="Alu"
     repeat_region   complement(62596. .62924)
                     /rpt_family="Alu"
     repeat_region   62986. .63277
                     /rpt_family="Alu"
     repeat_region   63592. .63888
                     /rpt_family="Alu"
     repeat_region   64602. .64769
                     /rpt_family="Alu"
     repeat_region   complement(64813. .65092)
                     /rpt_family="Alu"
     repeat_region   65891. .66202
                     /rpt_family="Alu"
     misc_feature    66304. .66378
                     /note="GRAIL 2 excellent exon, frame 1"
     repeat_region   complement(66660. .66949)
                     /rpt_family="Alu"
     repeat_region   67199. .67495
                     /rpt_family="Alu"
     repeat_region   67478. .67505
...
Note: remainder of annotations omitted.

  Query Match           74.8%;  Score 223;  DB 26;  Length 68067;
  Best Local Similarity  97.4%;  Pred. No. 1.52e-243;
  Matches  223;  Conservative    0;  Mismatches   6;  Indels   0;  Gaps   0;

Db  15181 GATCCTCTCTCCCAGGGAATCCGACACAGGAGGAACCCCTTCTCTGGTTGAGCTGGGCCA 15240
          |||||||||||||||||||||||||||||||||||||||||||| ||||||| ||||||||
Qy      1 GATCCTCTCTCCCAGGGAATCCGACACAGGAGGAACCCCTTCTNTGGTTGANCTGGGCCA 60

Db  15241 GGCCTAAGAGTAGCAGGAACTCTAAGACCACAGAGTTTTTTATAAATGTATAAATGTATC 15300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     61 GGCCTAAGAGTAGCAGGAACTCTAAGACCACAGAGTTTTTTATAAATGTATAAATGTATC 120

Db  15301 AAGCCAAATGTGCAGATGCTAACTGGACATTCTGGGGAACTGGGCACCAGGAGTGCCTTC 15360
          ||||||||| |||||||||||||||||||||||||||||||||||||||||| ||||||||
Qy    121 AAGCCAAATNTGCAGATGCTAACTGGACATTCTGGGGAACTGGGCACCAGGNGTGCCTTC 180

Db  15361 ATACACTGTACCCCAGCTCTCTTCTAAAAGAGAAGTGGGTGGGCACACT 15409
          |||||||||||||||||||| |||||||||||||||| ||||||||||||
Qy    181 ATACACTGTACCCCAGCTCTNTTCTAAAAGAGAAGTNGGTGGGCACACT 229
```



```
RESULT    3
LOCUS       PFSC03017   27207 bp    DNA             INV       20-MAR-1998
DEFINITION  Plasmodium falciparum DNA *** SEQUENCING IN PROGRESS *** from
            contig 3-17, complete sequence.
ACCESSION   AL008969
NID         g2982533
KEYWORDS    HTG.
SOURCE      malaria parasite.
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 27207)
  AUTHORS   Lawson,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAR-1998) P.falciparum Genome Sequencing Consortium,
            The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA, UK
COMMENT     On Mar 25, 1998 this sequence version replaced gi:2894448.
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and the release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc.
            For more information about this sequence or the Malaria Project,
            see http://www.sanger.ac.uk/Projects/P_falciparum.
FEATURES              Location/Qualifiers
     source           1. .27207
                      /organism="Plasmodium falciparum"
                      /db_xref="taxon:5833"
                      /chromosome="3"
                      /clone="3-17"
BASE COUNT    10775 a   2658 c   2663 g  11111 t
ORIGIN

  Query Match           6.4%;  Score 19;  DB 18;  Length 27207;
  Best Local Similarity 100.0%;  Pred. No. 4.97e-01;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  13660 ATACATTTATACATTTATA 13678
          |||||||||||||||||||
Cp    119 ATACATTTATACATTTATA 101


RESULT    4
LOCUS       CEF53B7     35955 bp    DNA             INV       21-SEP-1998
DEFINITION  Caenorhabditis elegans cosmid F53B7, complete sequence.
ACCESSION   Z72510
NID         g1301685
KEYWORDS    HTG; 156G surface protein; G-protein coupled receptor; UTR3.
SOURCE      Caenorhabditis elegans.
```

```
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 35955)
  AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
  TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
  JOURNAL   Nature 368 (6466), 32-38 (1994)
  MEDLINE   94150718
REFERENCE   2  (bases 1 to 35955)
  AUTHORS   McMurray,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-NOV-1996) Nematode Sequencing Project, Sanger Centre,
            Hinxton, Cambridge CB10 1RQ, England and Department of Genetics,
            Washington University, St. Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     Coding sequences below are predicted from computer analysis, using
            predictions from Genefinder (P. Green, U. Washington), and other
            available information.
            For a graphical representation of this sequence and its analysis
            see:-
            http://webace.sanger.ac.uk/cgi-
            bin/display?db=wormace&class=Sequence &object=F53B7
            Current sequence finishing criteria for the C. elegans genome
            sequencing consortium are that all bases are either sequenced
            unambiguously on both strands, or on a single strand with both a
            dye primer and dye terminator reaction, from distinct subclones.
            Exceptions are indicated by an explicit note.
            IMPORTANT:  This sequence is NOT necessarily the entire insert of
            the specified clone.  It may be shorter because we only sequence
            overlapping sections once, or longer because we arrange for a small
            overlap between neighbouring submissions.
            This sequence is the entire insert of clone F53B7. The true left
            end of clone C13G3 is at 35852 in this sequence. The true right end
            of clone F17C11 is at 8546 in this sequence. The start of this
            sequence (1. .104) overlaps with the end of sequence Z72507.
            The end of this sequence (35852. .35955) overlaps with the start of
            sequence Z74027.
FEATURES                Location/Qualifiers
     source             1. .35955
                        /organism="Caenorhabditis elegans"
                        /db_xref="taxon:6239"
                        /chromosome="V"
                        /clone="F53B7"
     gene               20722. .23602
                        /gene="F53B7.2"
     CDS                join(20722. .20983,21032. .21132,21185. .21449,21491. .21773,
                        22140. .22220,22385. .22589,22680. .22871,22916. .22994,
                        23043. .23234,23544. .23602)
                        /gene="F53B7.2"
                        /note="similar to G-protein coupled receptor"
                        /codon_start=1
                        /db_xref="PID:e239505"
                        /db_xref="PID:g1301686"
                        /db_xref="SPTREMBL:Q20715"
                        /translation="MHRSLSTFSFLINRFDSFFVSFSVCISNVFYLHKLSTSFSILKN
                        DGFQVRAFRLMTSTVIIVPDIQELPECEPSVFNDIQATIRLFGGMPIAIPGFITNVIN
                        VIVFCDPEMRCSLVNHFLLVLSISDLVLLVCNFFMLIFPVIASMSNSYLLHDYYPVFL
                        WFAYPVGLSTQTCGVYLTVLVSVHRYLGVCHPFRAKHWVSGKFVKWAIIGSIIFSIVI
                        NLHTWLELDIRPCYSINFNAPISSIILTSLRQKSSYNLITKCIMTTLIMFIIPFTILI
                        IVNCRIVVALKESTRMRNGQSMKKSTQSRIMNNFRMLKGAKYSELPGRFGRLNFNPLK
                        TPSLLKTNGNSLRDRSVTLMLLAIVAIFLCCNCLAPCNNIYENVQHVKKHASDSQSLN
                        QTSPVPEIEEQSQDYDGEWSIFNEWNFDLSVEISNLLISLNSSSSKFVYLIFSSKYRS
                        IIKHWLGLEKRKRTNGVALTTVMAAQKALELSILPDEVEARRHRKEKSHFVKNKKQMN
                        KSAQLFLTTSEIDLRKARTIRKEQEDEEEEEEEIREIQSDEPSAANDQNRSSSKRKLL
                        RFATLA"
     gene               complement(28506. .29465)
                        /gene="F53B7.3"
     CDS                complement(join(28506. .29007,29055. .29209,29253. .29327,
                        29394. .29465))
                        /gene="F53B7.3"
                        /note="similarity to yeast UTR3 protein (Swiss Prot
                        accession number P21374); cDNA EST EMBL:D72822 comes from
                        this gene; cDNA EST EMBL:D75763 comes from this gene; cDNA
                        EST yk274e3.3 comes from this gene; cDNA EST yk274e3.5
                        comes from this gene"
                        /codon_start=1
                        /db_xref="PID:e239379"
                        /db_xref="PID:g1301687"
                        /db_xref="SPTREMBL:Q20716"
                        /translation="MARNAEKAMTALARWRRMKEEEERGPIARRPHDVKDCRNLSDAE
                        RFRREIVRDASKKITAIQNPGLGEFKLRDLNDEVNRLIKLKHAWEQRIRELGGTDYRK
                        YAQKELDAIGRETGNSRGYKYFGAAKDLPGVRELFEKSTEGEEQQRRHRADLLRNIDAH
                        YFGYLDDEDGRLIPLEKLIEEKNIERINKEFAEKQAQKQQTASDAAPENIYKVEEDDD
                        DDLETQESTVIGEDGRPMTIRHVLLPTQQDIEEMLLEQKKQELMAKYLD"
     gene               complement(29774. .31319)
                        /gene="F53B7.4"
     CDS                complement(join(29774. .29964,30007. .30185,30341. .30446,
                        30492. .30591,30639. .30727,30777. .30915,30959. .31184,
                        31249. .31319))
                        /gene="F53B7.4"
                        /codon_start=1
```



```
                        /db_xref="PID:e239380"
                        /db_xref="PID:g1301688"
                        /db_xref="SPTREMBL:Q20717"
                        /translation="MFLNLTFILFFITIEANAYCPFKKAYVDDILMITTDDYLSVGQD
                        ETSEVCEIEIQSLNGNLILVSITNFLMFSHRRPRNPNEIILRNSDMQFIVDSKPKGYQ
                        LMKTPVWLTIPMNLNFRSILLFSNPKECSSYESIRCPEYSLCYNCSKYCFSSSLECDG
                        LDTCGEDESTCAFITPRTPHTSYRYGRDALSSSQIMLIGFFALNGSPIVFPLLILSLCS
                        AGLFYRCPFYAKNSNASRHTFIRYLRQAIPPALAPPPSVQGDQPKSDEGEKTLYTPSTP
                        TGHTRHRNSIIRINFQPAPALSEIEPEVEDREQSPQESICEPPAHQPEVFLRRQSLP
                        FGLTAVPVSHDLNSKTDDKRNRFASRAMSPQH"
     gene               join(complement(7332. .8823),complement(7153. .7289),
                        complement(6754. .7092),complement(5960. .6701),
                        complement(5574. .5915),complement(5139. .5528),
                        complement(4770. .5060),complement(4160. .4725),
                        complement(3561. .4107),complement(3371. .3507),
                        complement(2832. .3318),complement(1149. .2740),
                        complement(916. .1105),complement(672. .868),
                        complement(39. .602),complement(Z72507:36517. .36668),
                        complement(Z72507:36224. .36344))
                        /gene="F53B7.5"
     CDS                join(complement(7332. .8823),complement(7153. .7289),
                        complement(6754. .7092),complement(5960. .6701),
                        complement(5574. .5915),complement(5139. .5528),
                        complement(4770. .5060),complement(4160. .4725),
                        complement(3561. .4107),complement(3371. .3507),
                        complement(2832. .3318),complement(1149. .2740),
                        complement(916. .1105),complement(672. .868),
                        complement(39. .602),complement(Z72507:36517. .36668),
                        complement(Z72507:36224. .36344))
                        /gene="F53B7.5"
                        /note="similar to Thrombospondin type 1 domain; cDNA EST
                        EMBL:D34389 comes from this gene; cDNA EST EMBL:D37437
                        comes from this gene; cDNA EST EMBL:D64645 comes from this
                        gene; cDNA EST EMBL:D65908 comes from this gene; cDNA EST
                        EMBL:D67737 comes from this gene; cDNA EST EMBL:D69515
                        comes from this gene; cDNA EST EMBL:C13585 comes from this
                        gene; cDNA EST EMBL:C11441 comes from this gene; cDNA EST
                        yk356f12.3 comes from this gene; cDNA EST yk356f12.5 comes
                        from this gene; cDNA EST yk270f1.3 comes from this gene;
                        cDNA EST yk270f1.5 comes from this gene; cDNA EST
                        yk226b1.3 comes from this gene; cDNA EST yk226b1.5 comes
                        from this gene"
                        /codon_start=1
                        /db_xref="PID:e1322504"
                        /db_xref="PID:g3642634"
                        /translation="MLDETTTTEMTTTDATTADTTTTESETTTEASATTAASTTKDSE
                        TTTEASATTAASTTKDSETTTEASATTAATTTQESITTTESDSVTTTEGSATTGAATT
                        KDSVTTTEGSVTDKETTTEASATTTEGSATQATTAEETTAGEVTDGVSNPSVQTTLDF
                        TPPSNAPGVDTTLPIEGSPAPYDDCAPPVTTVTTMSPGTLGTLFPGGGGGGGGLDPF
                        TFPPTPTAMPGVTRSPACKTTPVLTTTIPFKTCTMIAVSTTTPMISLTTESMQSTTSS
                        SEISTTLDSSSTTGSVSSTESVTSDSSDSTTPLDVSSSPSESTESQSSTPSDASSSES
```



```
                        ASSSSPAATTEESGSSTDAVSSSPSGSSSESPNSSTGGSDSSSSESVTTEPSGSSTEP
                        SGSSTEPSGSSTEPSGSSTEPSGSSDSSSSELPNSSTDSGSTSEPSMSSEAPGASSSD
                        FPSSTPDSSVSSEEPNASSNSVSTETSPGSSDETTIGVIEETTTEYTWSTCNDCSQIN
                        VQLFNGNPFDDASSVLNEYSASGCTYVQPFCRPIEIGDTTPVSTILSGDTSYLSAPLS
                        ANLACRDKEWVLLSSTSPLAASVTSLGCVYANPTTTTRRPTTIPGIDPNNPCMNCTNL
                        VVSQSIHPEGVTVLDHTLDKCLSVDAHCLPINGGTLVELMYNNKGILDSGSTVNRTPT
                        CTKNSKWLDTTTNSIVNNISCIMPEVATTTPLVASTLHELPSTDGGPPTKDSSFKPVT
                        EPKITPEAEPPEEVTTPSGIMSTTQNDMTTVDSMPSTPGGGTLTRGPATGSPIDETTV
                        EELESTTTTLPSVCNNCVALEPLAKNAAGTYNGMLVLWHYKQNGCNRVSISCQATMSE
                        NDVAYLYYNPNGEVMSGVGELAMDLVCTNGEYVYEGKQISTVSCLASTTAVVPITTLP
                        PITTPAPGDDCSGCNRLPVAILSSSSEYTNGLLKATTSYSNCMSVQLTCYGVAATQAA
                        YLIYADATLATSRANANLTLTCNKAGWVTGSLQTVESISCGYTSQTTTTTALPSACG
                        SCANNGSAKALLQTNTYDGMIVLYNYVDPVTKCKKVDISCQGTTGTEVASLIPEPMGT
                        VSNGPCILEISLNCSSDGTWRYTGSVINSVSCIISTGVTVTPPTTVVVPTTATTLNPA
                        IPMCAQCSRLNAEAVSSIDNSLQNGVTVMTLSTVNQCSVIQVTCSGITANERVTLLAA
                        TTSLVSGVGSVSFNFTCNSAASWMTAAGTIVDSVACGRYNLTTSTPTTTTITSTSPRT
                        TTTTIPTQSQCSKCANMLAKALNPSDGSYDGMIVLNHHTSAGCRRVDISCGPTLNTEN
                        ATIFPGSSVVATGPAVHSPSIVCDSNANWIYDITSVSTVSCLISTASNGVTAPPTTST
                        TTVKDDTAPACSQCGRAPIVSVGTGSTYLNGFTTVVQSTDAQGCSLLPVKCQGMTSAD
                        AVALLSSGMPYSSGFGSTTANLTCNADSQWIGRTSTAITSLACGHKNAITTPSIITTL
                        APTGQCGSCPNMVAVAMSSGVSNSYDGMVIMDHFVNPATNCRALTVTCGPTVSTENAT
                        LYPNSVPISTGMATTLELACDSSAAWIKITTVNNVGCLISTGSGAVATTVSTLAPV
                        TNPSTNLCSQCARLPIQSVNDTEFLNGFTTLTTSINAPGCMTVSMMCEGVLPSDPVSI
                        VINGNYAETANGNVNYTDFACNSQGQWTSSTFGVVTSVICQRRSDTVIVTTALATTTT
                        TSNGLCNTCPNIVPRLMGTGASYDGMLVINHFKDRATACRKAVIDCMGTMDYETAELY
                        LDSVKINSSSQVEVSLQCTNNAVWMYENTQYSTASCLISNTTSSPSTGKKRKKRQAGT
                        GETACQSCGVLPMTSLHAFYPYWNGISTISIEEESTCMTAVAACEGVTSSDTVAWIAS
                        GTILSSGTDMINMTFTCNSMSNWQTSAGIVFSSLSCGRQYATTTVATTTVTSLAQNCS
                        NLIAAPLTSTELGENESNGQLILDHYINITTMERIVTITCYGGSATDNTTISFNGGSR
                        TMSGTGQIVLTMTCSSSGQWMRNGIGVEYVACIVVTPPGSTIPLTTTWATIPTTSAGS
                        SPNTNARNVMLTIPSGYNAGYLGMNTISSETSLQVVMSCEAPTGISESLLLSSTGAT
                        SSTCTPTTNMTLNCNSASQWVVSSASGTGIPVGTVITSIACVVAISTTTLVPATVAAG
                        CKLCTNLPAIGLEPNQLDANGRNGQFILDHVTTATACRVIYLSCAGSSASENSTILFN
                        GGAQTMSNQGITSTVVTCTAASTWDWYGSAVGYASCKVADRALTTTAAPVSVTTTAYV
                        VPGDTGLACSRSSLVNPLIEGDAGAYLVLDTKIVNGGLSTTLSCAAPDDSSTAVLTNG
                        GGSVLGTSDNGLINMTLTCNSNAQWVITAISASTSGTSAALGTVVDDLKCGEIPAATT
                        TLATTTASTACQLCPNMQAIALTSSQLVINGATNGQLLLSHSVDQSSKCRIVVIKCR
                        GENTNQNATIFFNSNGQSLSAMYGQVATSMACSSSNTWQRNGIEIDGVACMITSVVGA
                        TSTSPAPITTPAPSSTTQSSSGPSASCNTGFVTAVSTPGYTSGFMTVDTTTSGSIMSA
                        DITCASPVTSQTAALLGSNNNQVSSGVGSTFMTLTCNSNSQWVTPDGTVITSLSCGTV
                        APATTAATTTTLAPAATTTSTAAASDCSTSTWNAWQEWSTCTETCGSCGTQQRFRSCN
```

...

Note: remainder of annotations omitted.

```
 Query Match          6.4%;  Score 19;  DB 18;  Length 35955;
 Best Local Similarity 100.0%;  Pred. No. 4.97e-01;
 Matches    19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Db  26787 ATACATTTATACATTTATA 26805
          |||||||||||||||||||
Cp    119 ATACATTTATACATTTATA 101
```

```
RESULT    5
LOCUS     AC004157   147476 bp    DNA             HTG       20-AUG-1998
DEFINITION  *** SEQUENCING IN PROGRESS *** Plasmodium falciparum 3D7 chromosome
          12 PFTAC293 genomic sequence; HTGS phase 1, 32 unordered pieces.
ACCESSION   AC004157
NID         g3445191
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      malaria parasite P. falciparum.
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 147476)
  AUTHORS   Hyman,R.W., Qin,F., Fung,E.L., Conway,A.B. and Davis,R.W.
  TITLE     Plasmodium falciparum 3D7 chromosome 12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 147476)
  AUTHORS   Hyman,R.W., Qin,F., Fung,E.L., Conway,A.B. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-FEB-1998) Stanford DNA Sequencing and Technology
            Center, Stanford University, 855 California Avenue, Palo Alto, CA
            94304, USA
COMMENT     On Aug 20, 1998 this sequence version replaced gi:3335329.
            ***                                                          ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                          ***
            * This sequence is unfinished. It consists of 32 contigs for
            * which the order is not known; their order in this record is
            * arbitrary. In some cases, the exact lengths of the gaps
            * between the contigs are also unknown; these gaps are  presented
            * as runs of N as a convenience only. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
            *        1     4009: contig of 4009 bp in length
            *                     gap of unknown length
            *     4010     9108: contig of 5099 bp in length
            *                     gap of unknown length
            *     9109    15781: contig of 6673 bp in length
            *                     gap of unknown length
            *    15782    23960: contig of 8179 bp in length
            *                     gap of unknown length
            *    23961    28258: contig of 4298 bp in length
            *                     gap of unknown length
            *    28259    32224: contig of 3966 bp in length
            *                     gap of unknown length
            *    32225    34642: contig of 2418 bp in length
            *                     gap of unknown length
            *    34643    42603: contig of 7961 bp in length
            *                     gap of unknown length
            *    42604    47668: contig of 5065 bp in length
            *                     gap of unknown length
```



```
            *    47669    60968: contig of 13300 bp in length
            *                     gap of unknown length
            *    60969    69481: contig of 8513 bp in length
            *                     gap of unknown length
            *    69482    77569: contig of 8088 bp in length
            *                     gap of unknown length
            *    77570    85052: contig of 7483 bp in length
            *                     gap of unknown length
            *    85053    91429: contig of 6377 bp in length
            *                     gap of unknown length
            *    91430    96323: contig of 4894 bp in length
            *                     gap of unknown length
            *    96324   100874: contig of 4551 bp in length
            *                     gap of unknown length
            *   100875   105380: contig of 4506 bp in length
            *                     gap of unknown length
            *   105381   109230: contig of 3850 bp in length
            *                     gap of unknown length
            *   109231   113061: contig of 3831 bp in length
            *                     gap of unknown length
            *   113062   116412: contig of 3351 bp in length
            *                     gap of unknown length
            *   116413   119574: contig of 3162 bp in length
            *                     gap of unknown length
            *   119575   122674: contig of 3100 bp in length
            *                     gap of unknown length
            *   122675   125699: contig of 3025 bp in length
            *                     gap of unknown length
            *   125700   128520: contig of 2821 bp in length
            *                     gap of unknown length
            *   128521   131315: contig of 2795 bp in length
            *                     gap of unknown length
            *   131316   134048: contig of 2733 bp in length
            *                     gap of unknown length
            *   134049   136616: contig of 2568 bp in length
            *                     gap of unknown length
            *   136617   139161: contig of 2545 bp in length
            *                     gap of unknown length
            *   139162   141468: contig of 2307 bp in length
            *                     gap of unknown length
            *   141469   143602: contig of 2134 bp in length
            *                     gap of unknown length
            *   143603   145712: contig of 2110 bp in length
            *                     gap of unknown length
            *   145713   147476: contig of 1764 bp in length.
FEATURES             Location/Qualifiers
     source          1. .147476
                     /organism="Plasmodium falciparum"
                     /db_xref="taxon:5833"
                     /chromosome="12"
BASE COUNT    56806 a  16366 c  17047 g  57154 t    103 others
```



```
ORIGIN

  Query Match           6.4%;  Score 19;  DB 17;  Length 147476;
  Best Local Similarity 100.0%;  Pred. No. 4.97e-01;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  52547 ATACATTTATACATTTATA 52565
          ||||||||||||||||||
Cp    119 ATACATTTATACATTTATA 101


RESULT    6
LOCUS     HS742C19   178615 bp    DNA             HTG       08-OCT-1998
DEFINITION  Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone
          742C19; HTGS phase 1.
ACCESSION   AL031846
NID         g3724248
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 178615)
  AUTHORS   Griffiths,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-OCT-1998) Wellcome Trust Genome Campus, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquires:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and the release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc. Order of segments is not known; 800 n's separate
            segments. Unfinished: dJ742C19  Contig_ID: 01948  acc=  Length:
            176771 bp Unfinished: dJ742C19  Contig_ID: 02029  acc=  Length:
            1044 bp.
            ***                                                          ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                          ***
            * This sequence is unfinished. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
FEATURES             Location/Qualifiers
     source          1. .178615
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="22"
                     /clone="742C19"
BASE COUNT    46053 a  45222 c  45012 g  41476 t    852 others
ORIGIN
```



```
  Query Match           6.4%;  Score 19;  DB 17;  Length 178615;
  Best Local Similarity 95.0%;  Pred. No. 4.97e-01;
  Matches   19;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db  97131 ACGCCTGGTGCCCAGTTCCC 97150
          || ||||||||||||||||||
Cp    174 ACNCCTGGTGCCCAGTTCCC 155


RESULT    7
LOCUS     MTCMXAFAA     600 bp    DNA             BCT       02-MAR-1995
DEFINITION  Methylobacter albus methanol dehydrogenase alpha subunit (mxaF)
          gene, partial cds.
ACCESSION   L33682
NID         g497609
KEYWORDS    .
SOURCE      Methylobacter albus (strain BG8) DNA.
  ORGANISM  Methylobacter albus
            Eubacteria; Proteobacteria; gamma subdivision; Methylococcaceae;
            Methylobacter.
REFERENCE   1  (bases 1 to 600)
  AUTHORS   Chistoserdova,L., Kuhn,M. and Lidstrom,M.E.
  TITLE     Identification of a promoter region for mxaF (moxF) from the type I
            methanotroph, Methylobacter albus BG8
  JOURNAL   FEMS Microbiol. Lett. 121 (3), 343-348 (1994)
  MEDLINE   95011500
FEATURES             Location/Qualifiers
     source          1. .600
                     /organism="Methylobacter albus"
                     /strain="BG8"
                     /db_xref="taxon:34057"
     gene            421. .600
                     /gene="mxaF"
     CDS             421. .>600
                     /gene="mxaF"
                     /note="formerly called moxF"
                     /codon_start=1
                     /transl_table=11
                     /product="methanol dehydrogenase alpha subunit"
                     /db_xref="PID:g688433"
                     /translation="MRISKITRTAVSLACGGAFLAAALPAEANKELDRLSKEDTNWVM
                     QTKDYGATHFSKMTQI"
BASE COUNT      132 a    183 c    147 g    138 t
ORIGIN

  Query Match           6.0%;  Score 18;  DB 15;  Length 600;
  Best Local Similarity 90.0%;  Pred. No. 2.92e+00;
  Matches   18;  Conservative    0;  Mismatches   2;  Indels   0;  Gaps   0;

Db     57 GATACTTTTGGCCCTGCTTG 76
```

```
              ||||||||||||||| | ||
Qy     283 GATACTTTTGGCCCTNCNTG 302


RESULT    8
LOCUS       AF030635      1514 bp     mRNA             ROD       16-NOV-1997
DEFINITION  Mus musculus FK-506 binding protein homolog (SAM11) mRNA, complete
            cds.
ACCESSION   AF030635
NID         g2623223
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1514)
  AUTHORS   Pedersen,K.M., Finsen,B., Celis,J.E. and Jensen,N.A.
  TITLE     muFKBP38: A novel murine 38kD immunophillin homolog differentially
            expressed in Schwannoma cells and the forebrain
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1514)
  AUTHORS   Pedersen,K.M. and Jensen,N.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-OCT-1997) Department of Medical Biochemistry,
            University of Aarhus, Ole Worms Alle, Bldg.170, Aarhus C. 8000,
            Denmark
FEATURES             Location/Qualifiers
     source          1. .1514
                     /organism="Mus musculus"
                     /strain="C57BL6/DBA2"
                     /db_xref="taxon:10090"
                     /cell_type="Schwannoma"
     gene            1. .1514
                     /gene="SAM11"
     CDS             29. .1096
                     /gene="SAM11"
                     /note="muFKBP38"
                     /codon_start=1
                     /product="FK-506 binding protein homolog"
                     /db_xref="PID:g2623224"
                     /translation="MGQPTVEEAEQPGALAREFLAATEPEPAPAPAPEEWLDILGNGL
                     LRKETLVPGPPGSSRPLKGQVVTVHLQMSLENGTRVQEEPELAFTLGDCDVIQALDLS
                     VPLMDVGETAMVTADSKYCYGPQGRSPYIPPHAALCLEVTLKTAEDGPDLEMLSQQER
                     VALANRKRECGNAHYQRADFVLAANSYDLAIKAITSNTKVDMTCEEEELLQLKVKCL
                     NNLAASQLKLDHYRAALRSCSQVLEHQPDNIKALPRKGKVLAQQGEYSEAIPILRAAL
                     KLEPSNKTIHAELSKLVKKRAAQRSTETALYRKMLGNPSRLPAKCPGKGAWSIPWKWL
                     FGATAVALGGVALSVVIAGRN"
BASE COUNT      360 a    457 c    434 g    263 t
ORIGIN

  Query Match             6.0%;  Score 18;  DB 28;  Length 1514;
```



```
  Best Local Similarity 100.0%;  Pred. No. 2.92e+00;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    137 CTGGACATTCTGGGGAAC 154
          ||||||||||||||||||
Qy    143 CTGGACATTCTGGGGAAC 160


RESULT    9
LOCUS       HUMFKBPA      1641 bp     mRNA             PRI       16-MAY-1996
DEFINITION  Human FK-506 binding protein homologue (FKBP38) mRNA, complete cds.
ACCESSION   L37033
NID         g965469
KEYWORDS    binding protein; leucine zipper; tetratricopeptide family.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1641)
  AUTHORS   Lam,E., Martin,M. and Wiederrecht,G.
  TITLE     Isolation of a cDNA encoding a novel human FK506-binding protein
            homolog containing leucine zipper and tetratricopeptide repeat
            motifs
  JOURNAL   Gene 160 (2), 297-302 (1995)
  MEDLINE   95369708
COMMENT     Related entries: acc# T09208 T08452 M79056 M62106.
FEATURES             Location/Qualifiers
     source          1. .1641
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="Jurkat library (Stratagene)"
     gene            141. .1208
                     /gene="FKBP38"
     CDS             141. .1208
                     /gene="FKBP38"
                     /codon_start=1
                     /product="FK-506 binding protein homologue"
                     /db_xref="PID:g965470"
                     /translation="MGQPPAEEAEQPGALAREFLAAMEPEPAPAPAPEEWLDILGNGL
                     LRKKTLVPGPPGSSRPVKGQVVTVHLQTSLENGTRVQEEPELVFTLGDCDVIQALDLS
                     VPLMDVGETAMVTADSKYCYGPQGRSPYIPPHAALCLEVTLKTAVDGPDLEMLTGQER
                     VALANRKRECGNAHYQRADFVLAANSYDLAIKAITSSAKVDMTFEEEAQLLQLKVKCL
                     NNLAASQLKLDHYRAALRSCSLVLEHQPDNIKALFRKGKVLAQQGEYSEAIPILRAAL
                     KLEPSNKTIHAELSKLVKKHAAQRSTETALYRKMLGNPSRLPAKCPGKGAWSIPWKWL
                     FGATAVALGGVALSVVIAARN"
BASE COUNT      324 a    550 c    507 g    260 t
ORIGIN

  Query Match             6.0%;  Score 18;  DB 25;  Length 1641;
  Best Local Similarity 100.0%;  Pred. No. 2.92e+00;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```



```
Db    249 CTGGACATTCTGGGGAAC 266
          ||||||||||||||||||
Qy    143 CTGGACATTCTGGGGAAC 160


RESULT   10
LOCUS       HSL226F1      8717 bp     DNA              PRI       19-DEC-1995
DEFINITION  Human DNA sequence from cosmid L226F1, Huntington's Disease Region,
            chromosome 4p16.3 *.
ACCESSION   Z68278
NID         g1130690
KEYWORDS    4p16.3.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 8717)
  AUTHORS   Williamson,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-DEC-1995) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1RQ, England. E-mail enquires regarding this sequence:
            humpub@sanger.ac.uk
COMMENT     IMPORTANT:
            This sequence is not the entire insert of clone L226F1.  It may be
            shorter because we only sequence overlapping sections once, or
            longer because we arrange for a small overlap between neighbouring
            submissions.
            The true left end of clone L226F1 is at 1 in this sequence. The
            true left end of clone L17A9 is at 8618 in this sequence. This
            sequence has been finished according to sequence map criteria as
            follows.  An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu).  Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            L226F1 is from cosmid library LA04NC01 constructed at the Human
            Genome Center, Los Alamos National Laboratory, NM 87545 under the
            auspices of the U.S. Department of Energy. The library was
            constructed using flow-sorted human chromosome 4 from a
            Hamster-Human hybrid cell line ( UV20HL21-27 ) containing human
            chromosomes 4, 8 and 21.
            VECTOR: sCos1
            L226F1 is contained in a clone contig spanning
            2Mb which is described in Baxendale et al, Nature Genetics 4 ( 1993
            ) 181- 186.  See also Myers et al, Cytogenet Cell Genet. 66 ( 1994
            ) 218-230.
FEATURES             Location/Qualifiers
     source          1. .8717
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
```



```
                     /chromosome="4"
                     /map="4p16.3"
                     /clone="L226F1"
                     /clone_lib="LA04NC01"
                     /cell_line="UV20HL21-27"
     misc_feature    complement(1024. .1389)
                     /note="matches STS L09876"
     repeat_region   2847. .2906
                     /note="MSTC element fragment"
     repeat_region   3032. .3085
                     /note="MSTC element fragment"
     repeat_region   4856. .5147
                     /note="Alu repeat: matches 308. .1 of consensus"
BASE COUNT     1971 a   2461 c   2270 g   2015 t
ORIGIN

  Query Match             6.0%;  Score 18;  DB 25;  Length 8717;
  Best Local Similarity 100.0%;  Pred. No. 2.92e+00;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   5528 ATAAATGTATAAATGTAT 5545
          ||||||||||||||||||
Qy    102 ATAAATGTATAAATGTAT 119


RESULT   11
LOCUS       AC005758    41414 bp    DNA              PRI       02-OCT-1998
DEFINITION  Homo sapiens chromosome 19, cosmid R27946, complete sequence.
ACCESSION   AC005758
NID         g3688087
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 41414)
  AUTHORS   Lamerdin,J.E., McCready,P.M., Skowronski,E., Viswanathan,V.,
            Burkhart-Schultz,K., Gordon,L., Dias,J., Ramirez,M., Stilwagen,S.,
            Phan,H., Velasco,N., Do,L., Regala,W., Terry,A., Garnes,J.,
            Danganan,L., Poundstone,P., Christensen,M., Georgescu,A., Avila,J.,
            Liu,S., Attix,C., Andreise,T., Trankheim,M., Amico-Keller,G.,
            Coefield,J., Duarte,S., Lucas,S., Bruce,R., Thomas,P., Quan,G.,
            Kronmiller,B., Arellano,A., Montgomery,M., Ow,D., Nolan,M.,
            Trong,S., Kobayashi,A., Olsen,A.S. and Carrano,A.V.
  TITLE     Sequence analysis of a 500 kb region in 19q12 between D19S766 and
            D19S222
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 41414)
  AUTHORS   Lamerdin,J.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-OCT-1998) Joint Genome Institute, Lawrence Livermore
```

```
            National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
  COMMENT   Map and sequence oriented from q centromere to
            telomere. Cosmid R27946 is expected to overlaps BAC 326584
            (CIT-978SK-459f4) to the left by approximately 10 kb, and is
            separated from cosmid F16593 to the right by approximately 3 kb.
            Additional map and sequence information may be obtained at:
            http://www-bio.llnl.gov/bbrp/genome/genome.html.
  FEATURES             Location/Qualifiers
       source          1. .41414
                       /organism="Homo sapiens"
                       /db_xref="taxon:9606"
                       /clone="R27946"
                       /chromosome="19"
                       /map="19q12 between D19S766 and D19S222"
                       /cell_line="5HL2-B"
                       /clone_lib="LL19NC03 R chromosome 19-specific cosmid"
                       /note="Cosmid library constructed at LLNL from flow-sorted
                       chromosomes from hybrid 5HL2-B, which carries chromosome
                       19 as its only human chromosome."
       misc_feature    complement(453. .578)
                       /note="predicted exon, program: grail2exons_human_1.3,
                       frame: 0, quality: good, score: 61.000"
       repeat_region   1332. .1370
                       /rpt_family="MIR"
       misc_feature    complement(1436. .1499)
                       /note="predicted exon, program: grail2exons_human_1.3,
                       frame: 1, quality: excellent, score: 99.000"
       repeat_region   complement(1486. .1564)
                       /rpt_family="MIR"
       repeat_region   1578. .1879
                       /rpt_family="AluSx"
       repeat_region   4171. .4243
                       /rpt_family="MIR"
       repeat_region   complement(4320. .4616)
                       /rpt_family="AluJb"
       repeat_region   complement(4632. .4690)
                       /rpt_family="AT_rich"
       repeat_region   complement(4830. .4875)
                       /rpt_family="(CAT)n"
       misc_feature    complement(5046. .5148)
                       /note="predicted exon, program: grail2exons_human_1.3,
                       frame: 0, quality: excellent, score: 75.000"
       repeat_region   5355. .5410
                       /rpt_family="(TA)n"
       repeat_region   complement(5429. .5556)
                       /rpt_family="FLAM_C"
       repeat_region   complement(6009. .6052)
                       /rpt_family="AT_rich"
       misc_feature    complement(6141. .6293)
                       /note="predicted exon, program: grail2exons_human_1.3,
                       frame: 0, quality: excellent, score: 89.000"
```



```
       repeat_region   6717. .6832
                       /rpt_family="AluJo"
       repeat_region   complement(7703. .7809)
                       /rpt_family="(CAGG)n"
       misc_feature    10295. .10491
                       /note="predicted exon, program: grail2exons_human_1.3,
                       frame: 0, quality: good, score: 62.000"
       misc_feature    complement(10340. .10759)
                       /note="BLASTN similarity to AQ141027 (1. .420); match:
                       0.98, score: 7.3e-162; database searched: month.na;
                       HS_3141_B2_C06_MR CIT Approved Human Genomic Sperm Library
                       D Homo sapiens genomic clone Plate=3141 Col=12 Row=F,"
       repeat_region   complement(10684. .10963)
                       /rpt_family="MER58B"
       repeat_region   complement(11383. .12146)
                       /rpt_family="L1MB8"
       repeat_region   complement(12290. .12403)
                       /rpt_family="L1"
       repeat_region   complement(12443. .12717)
                       /rpt_family="AluJb"
       repeat_region   complement(12829. .13017)
                       /rpt_family="L1PA7"
       repeat_region   complement(13037. .13311)
                       /rpt_family="L1"
       repeat_region   complement(13312. .13431)
                       /rpt_family="L1MA10"
       repeat_region   complement(13581. .13853)
                       /rpt_family="L1"
       repeat_region   complement(13851. .14141)
                       /rpt_family="L1PA16"
       repeat_region   complement(14745. .14772)
                       /rpt_family="AT_rich"
       repeat_region   14897. .15158
                       /rpt_family="L1MB4"
       repeat_region   15177. .15244
                       /rpt_family="(CATA)n"
       misc_feature    16083. .16119
                       /note="predicted exon, program: grail2exons_human_1.3,
                       frame: 2, quality: good, score: 72.000"
       misc_feature    16698. .16782
                       /note="predicted exon, program: grail2exons_human_1.3,
                       frame: 2, quality: excellent, score: 92.000"
       misc_feature    19059. .19168
                       /note="predicted exon, program: grail2exons_human_1.3,
                       frame: 2, quality: excellent, score: 83.000"
       misc_feature    19251. .19306
                       /note="predicted exon, program: grail2exons_human_1.3,
                       frame: 2, quality: good, score: 55.000"
       repeat_region   19454. .19732
                       /rpt_family="AluJo"
       repeat_region   20334. .20414
```



```
                       /rpt_family="(GA)n"
       repeat_region   complement(21685. .21976)
                       /rpt_family="AluSq"
       repeat_region   complement(22087. .22386)
                       /rpt_family="AluSg"
       repeat_region   complement(22465. .22765)
                       /rpt_family="AluSq"
       misc_feature    24809. .24896
                       /note="predicted exon, program: grail2exons_human_1.3,
                       frame: 1, quality: good, score: 55.000"
       repeat_region   complement(26305. .26539)
                       /rpt_family="AluJo"
       misc_feature    26865. .27104
                       /note="predicted exon, program: grail2exons_human_1.3,
                       frame: 2, quality: good, score: 65.000"
       repeat_region   complement(27110. .27181)
                       /rpt_family="(GAAAA)n"
       repeat_region   complement(27632. .27723)
                       /rpt_family="POLY_A"
       repeat_region   complement(27839. .28140)
                       /rpt_family="THE1B"
       repeat_region   29453. .29701
                       /rpt_family="LTR4"
       repeat_region   30191. .30335
                       /rpt_family="LINE2"
       repeat_region   complement(30735. .30774)
                       /rpt_family="AT_rich"
       repeat_region   complement(30895. .30919)
                       /rpt_family="AT_rich"
       misc_feature    complement(31251. .31408)
                       /note="predicted exon, program: grail2exons_human_1.3,
                       frame: 1, quality: excellent, score: 87.000"
       repeat_region   complement(31672. .31741)
                       /rpt_family="LINE2"
       repeat_region   32622. .32731
                       /rpt_family="MIR"
       repeat_region   32768. .32804
                       /rpt_family="LINE2"
       repeat_region   33222. .33380
                       /rpt_family="MIR"
       repeat_region   35511. .35819
                       /rpt_family="AluSx"
       repeat_region   complement(36685. .36983)
                       /rpt_family="AluSc"
       repeat_region   complement(37232. .37271)
                       /rpt_family="AT_rich"
       repeat_region   complement(37315. .37453)
                       /rpt_family="FLAM_A"
       repeat_region   complement(39244. .39891)
                       /rpt_family="LINE2"
       repeat_region   39981. .40203
```



```
                       /rpt_family="L1MC/D"
       repeat_region   complement(40230. .41414)
                       /rpt_family="L1PA8"
  BASE COUNT    11377 a   8473 c   8956 g  12608 t
  ORIGIN

    Query Match           6.0%;  Score 18;  DB 27;  Length 41414;
    Best Local Similarity  90.0%;  Pred. No. 2.92e+00;
    Matches   18;  Conservative    0;  Mismatches   2;  Indels   0;  Gaps   0;

  Db  35893 TGTTCTAAAAGAGAAGTGGG 35912
            | ||||||||||||||| ||
  Qy    200 TNTTCTAAAAGAGAAGTNGG 219


  RESULT   12
  LOCUS       AC005512    44219 bp    DNA             PRI       22-AUG-1998
  DEFINITION  Homo sapiens chromosome 19, cosmid F23259, complete sequence.
  ACCESSION   AC005512
  NID         g3450745
  KEYWORDS    HTG.
  SOURCE      human.
    ORGANISM  Homo sapiens
              Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
              Primates; Catarrhini; Hominidae; Homo.
  REFERENCE   1  (bases 1 to 44219)
    AUTHORS   Lamerdin,J.E., McCready,P.M., Skowronski,E., Adamson,A.W.,
              Burkhart-Schultz,K., Gordon,L., Kyle,A., Ramirez,M., Stilwagen,S.,
              Phan,H., Velasco,N., Do,L., Regala,W., Terry,A., Garnes,J.,
              Danganan,L., Poundstone,P., Christensen,M., Georgescu,A., Avila,J.,
              Liu,S., Attix,C., Andreise,T., Trankheim,M., Amico-Keller,G.,
              Coefield,J., Duarte,S., Lucas,S., Bruce,R., Thomas,P., Quan,G.,
              Kronmiller,B., Arellano,A., Montgomery,M., Ow,D., Nolan,M.,
              Trong,S., Kobayashi,A., Olsen,A.S. and Carrano,A.V.
    TITLE     Sequence analysis of a 3.5 Mb contig in human 19p13.3 containing a
              serine protease gene cluster
    JOURNAL   Unpublished
  REFERENCE   2  (bases 1 to 44219)
    AUTHORS   Lamerdin,J.E.
    TITLE     Direct Submission
    JOURNAL   Submitted (22-AUG-1998) Joint Genome Institute, Lawrence Livermore
              National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
  COMMENT     Map and sequence are oriented from p telomere to centromere. Cosmid
              F23259 overlaps cosmid F30017 to the left from bases 1 to 2,155 of
              this accession. There is currently a sequence gap to the left
              between F23259 and fosmid 39347 of approximately 5 kb. Additional
              map and sequence information may be obtained at:
              http://www-bio.llnl.gov/bbrp/genome/genome.html.
  FEATURES             Location/Qualifiers
       source          1. .44219
                       /organism="Homo sapiens"
```

```
                     /db_xref="taxon:9606"
                     /clone="F23259"
                     /chromosome="19"
                     /map="19p13.3 between CDC34 and D19S342"
                     /cell_line="UV5HL9-5B"
                     /clone_lib="LL19NC02 F chromosome 19-specific cosmid
                     library"
                     /note="Cosmid library constructed at LLNL from flow-sorted
                     chromosomes from hybrid UV5HL9-5B, which carries
                     chromosome 19 as its only human chromosome."
     repeat_region   1. .238
                     /rpt_family="AluSc"
     repeat_region   complement(520. .552)
                     /rpt_family="(TAAAA)n"
     repeat_region   complement(553. .854)
                     /rpt_family="AluY"
     repeat_region   complement(857. .1130)
                     /rpt_family="AluSq"
     repeat_region   complement(1590. .1717)
                     /rpt_family="FLAM_C"
     misc_feature    complement(2587. .2981)
                     /note="BLASTN similarity to cpg254e2.rt1a (1. .395); match:
                     1, score: 6.9e-158; database searched: Sanger CpG; bases
                     78 to 472 (SL to QR)"
     repeat_region   3224. .3331
                     /rpt_family="GC_rich"
     repeat_region   complement(3426. .3461)
                     /rpt_family="GC_rich"
     misc_feature    3655. .4093
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 0, quality: good, score: 54.000"
     repeat_region   complement(4149. .4463)
                     /rpt_family="AluSx"
     repeat_region   complement(4467. .4769)
                     /rpt_family="AluSq"
     repeat_region   5622. .5713
                     /rpt_family="LINE2"
     repeat_region   complement(5984. .6285)
                     /rpt_family="AluY"
     repeat_region   complement(6306. .6606)
                     /rpt_family="AluSx"
     misc_feature    7008. .7039
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 2, quality: good, score: 56.000"
     repeat_region   complement(7351. .7393)
                     /rpt_family="(TAAAA)n"
     repeat_region   complement(7396. .7698)
                     /rpt_family="AluSx"
     repeat_region   complement(7722. .7928)
                     /rpt_family="L1MC2"
     repeat_region   complement(7934. .7983)
```



```
                     /rpt_family="MIR"
     repeat_region   complement(8375. .8410)
                     /rpt_family="(CA)n"
     repeat_region   complement(8942. .8964)
                     /rpt_family="GC_rich"
     repeat_region   complement(9166. .9221)
                     /rpt_family="MIR"
     repeat_region   complement(9918. .9964)
                     /rpt_family="(GGA)n"
     repeat_region   complement(10410. .10719)
                     /rpt_family="AluSx"
     repeat_region   complement(10720. .10968)
                     /rpt_family="AluJo"
     repeat_region   10975. .11114
                     /rpt_family="MIR"
     repeat_region   complement(11435. .11684)
                     /rpt_family="MLT1D"
     repeat_region   complement(11748. .12045)
                     /rpt_family="AluSp"
     repeat_region   complement(12091. .12392)
                     /rpt_family="AluJo"
     repeat_region   complement(12632. .12696)
                     /rpt_family="MIR"
     repeat_region   12726. .13025
                     /rpt_family="AluY"
     repeat_region   complement(13637. .13938)
                     /rpt_family="AluY"
     repeat_region   complement(14520. .14821)
                     /rpt_family="AluSp"
     repeat_region   15281. .15582
                     /rpt_family="AluSx"
     repeat_region   15755. .15875
                     /rpt_family="HERVL"
     repeat_region   complement(15994. .16175)
                     /rpt_family="L1MC2"
     repeat_region   complement(16177. .16469)
                     /rpt_family="AluSx"
     repeat_region   complement(16470. .16587)
                     /rpt_family="L1MC2"
     repeat_region   complement(16635. .16733)
                     /rpt_family="MLT1D"
     repeat_region   complement(16747. .17045)
                     /rpt_family="AluY"
     repeat_region   complement(17046. .17073)
                     /rpt_family="AT_rich"
     repeat_region   complement(17074. .17167)
                     /rpt_family="U6"
     repeat_region   complement(17193. .17482)
                     /rpt_family="AluSp"
     repeat_region   complement(17598. .17788)
                     /rpt_family="MER45"
```



```
     misc_feature    17916. .17940
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 1, quality: good, score: 63.000"
     repeat_region   18207. .18490
                     /rpt_family="MLT1D"
     repeat_region   complement(18491. .18526)
                     /rpt_family="AT_rich"
     repeat_region   complement(18609. .18688)
                     /rpt_family="L1PA10"
     repeat_region   18692. .18985
                     /rpt_family="AluSx"
     repeat_region   complement(18986. .19260)
                     /rpt_family="L1PA9"
     repeat_region   complement(19261. .19554)
                     /rpt_family="AluY"
     repeat_region   19794. .20098
                     /rpt_family="AluJo"
     repeat_region   20243. .20556
                     /rpt_family="AluSp"
     repeat_region   complement(20644. .20944)
                     /rpt_family="AluSx"
     repeat_region   complement(20989. .21022)
                     /rpt_family="AT_rich"
     repeat_region   21245. .21381
                     /rpt_family="MIR"
     repeat_region   complement(21568. .21685)
                     /rpt_family="(GA)n"
     repeat_region   complement(21686. .21984)
                     /rpt_family="AluY"
     repeat_region   complement(22033. .22058)
                     /rpt_family="(CA)n"
     repeat_region   complement(22059. .22359)
                     /rpt_family="AluSx"
     repeat_region   complement(22401. .22705)
                     /rpt_family="AluSp"
     repeat_region   22835. .23135
                     /rpt_family="AluY"
     repeat_region   23199. .23273
                     /rpt_family="MADE1"
     repeat_region   23311. .23631
                     /rpt_family="L1MB7"
     repeat_region   complement(23666. .23961)
                     /rpt_family="AluSq"
     repeat_region   complement(24077. .24379)
                     /rpt_family="AluSq"
     repeat_region   complement(24396. .24700)
                     /rpt_family="AluSg"
     repeat_region   complement(24701. .24996)
                     /rpt_family="AluSx"
     misc_feature    25247. .26140
                     /note="DDS similarity to overlapping ESTs:-(25247. .25624)
```



```
                     N49963 yz11b12.s1 Homo sapiens cDNA clone 282719 3'.
                     Score: 690  Identity: 372/384 (96%).--(25247. .25714)
                     N47607 yy59e06.s1 Homo sapiens cDNA clone 277858 3'.
                     Score: 891  Identity: 462/471 (98%).--(25909. .25621)
                     H46880 yo19f02.r1 Homo sapiens cDNA clone 178395 5'.
                     Score: 550  Identity: 282/289 (97%).--(26140. .25687)
                     W79897 zd79a07.r1 Soares fetal heart NbHH19W Homo sapiens
                     cDNA clone 346836 5'; Score: 866  Identity: 447/456
                     (98%).--Additional EST matches:-H46023, and many others"
     repeat_region   27946. .28244
                     /rpt_family="AluSq"
     mRNA            complement(28477. .34746)
...
Note: remainder of annotations omitted.

  Query Match             6.0%;  Score 18;  DB 27;  Length 44219;
  Best Local Similarity  90.0%;  Pred. No. 2.92e+00;
  Matches    18;  Conservative    0;  Mismatches   2;  Indels   0;  Gaps   0;

Db  26962 CCCAACCACAGGAGGCTGCA 26981
          || |||||||||| ||||||
Qy    246 CCNAACCACAGGANGCTGCA 265


RESULT   13
LOCUS       AC004461    44897 bp    DNA                 PRI       20-JUN-1998
DEFINITION  Homo sapiens Chromosome 22q11.2 Cosmid Clone 119f4 In DGCR Region,
            complete sequence.
ACCESSION   AC004461 U30597
NID         g3242742
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 44897)
  AUTHORS   Roe,B.A.
  TITLE     HTGS Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 44897)
  AUTHORS   Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-MAR-1998) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   3  (bases 1 to 44897)
  AUTHORS   Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-APR-1998) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
```

```
REFERENCE   4  (bases 1 to 44897)
  AUTHORS   Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAY-1998) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   5  (bases 1 to 44897)
  AUTHORS   Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-1998) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   6  (bases 1 to 44897)
  AUTHORS   Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUN-1998) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On Jun 20, 1998 this sequence version replaced gi:3132494.
FEATURES             Location/Qualifiers
     source          1. .44897
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT     9298 a  11738 c  12086 g  11775 t
ORIGIN

  Query Match           6.0%;  Score 18;  DB 26;  Length 44897;
  Best Local Similarity  94.7%;  Pred. No. 2.92e+00;
  Matches   18;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db  16164 GCACCAGGTGTGCCTTCAT 16182
          ||||||||| ||||||||||
Qy    164 GCACCAGGNGTGCCTTCAT 182


RESULT   14
LOCUS       HS313F4     49496 bp    DNA             HTG       23-SEP-1998
DEFINITION  Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone 313F4;
            HTGS phase 1.
ACCESSION   Z97633
NID         g3421033
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 49496)
  AUTHORS   Lloyd,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-SEP-1998) Wellcome Trust Genome Campus, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquires:
```



```
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 18, 1998 this sequence version replaced gi:2578083.
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and the release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc. Order of segments is not known; 800 n's separate
            segments. Unfinished: c313F4  Contig_ID: 00059  acc=Z97633  Length:
            49496 bp.
            ***                                                          ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                          ***
            * This sequence is unfinished. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
FEATURES             Location/Qualifiers
     source          1. .49496
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="313F4"
BASE COUNT    12260 a  13047 c  12724 g  11465 t
ORIGIN

  Query Match           6.0%;  Score 18;  DB 17;  Length 49496;
  Best Local Similarity  94.7%;  Pred. No. 2.92e+00;
  Matches   18;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db   8109 CTGGGCACCAGGTGTGCCT 8127
          ||||||||||| ||||||
Qy    160 CTGGGCACCAGGNGTGCCT 178


RESULT   15
LOCUS       HS496H19    73864 bp    DNA             HTG       29-SEP-1998
DEFINITION  Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone
            496H19; HTGS phase 1.
ACCESSION   AL023582
NID         g3550061
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 73864)
  AUTHORS   Smalley,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-SEP-1998) Wellcome Trust Genome Campus, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquires:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
```



```
COMMENT     On Sep 8, 1998 this sequence version replaced gi:3452364.
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and the release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc. Order of segments is not known; 800 n's separate
            segments. Unfinished: dJ496H19  Contig_ID: 00837  acc=AL023582
            Length: 66473 bp Unfinished: dJ496H19  Contig_ID: 00720
            acc=AL023582  Length: 1015 bp Unfinished: dJ496H19  Contig_ID:
            00263  acc=AL023582  Length: 1108 bp Unfinished: dJ496H19
            Contig_ID: 01380  acc=AL023582  Length: 1017 bp Unfinished:
            dJ496H19  Contig_ID: 00404  acc=AL023582  Length: 1051 bp.
            ***                                                          ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                          ***
            * This sequence is unfinished. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
FEATURES             Location/Qualifiers
     source          1. .73864
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="496H19"
BASE COUNT    22353 a  14478 c  13790 g  20042 t   3201 others
ORIGIN

  Query Match           6.0%;  Score 18;  DB 17;  Length 73864;
  Best Local Similarity  94.7%;  Pred. No. 2.92e+00;
  Matches   18;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db  23774 AGGCACTCCTGGTGCCCAG 23792
          |||||| ||||||||||||
Cp    178 AGGCACNCCTGGTGCCCAG 160


RESULT   16
LOCUS       ATT12H17    89350 bp    DNA             PLN       03-FEB-1998
DEFINITION  Arabidopsis thaliana DNA chromosome 4, BAC clone T12H17 (ESSAII
            project).
ACCESSION   AL021635
NID         g2827538
KEYWORDS    .
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryotae; Viridiplantae; Charophyta/Embryophyta group;
            Embryophyta; Tracheophyta; seed plants; Magnoliophyta;
            eudicotyledons; Rosidae; Capparales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 89350)
  AUTHORS   Bevan,M., Hilbert,H., Braun,M., Holzer,E., Brandt,A.,
```



```
            Duesterhoeft,A., Bancroft,I., Mewes,H.W., Mayer,K. and Schueller,C.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 89350)
  AUTHORS   EU Arabidopsis sequencing,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-FEB-1998) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, Project
            Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
            Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
            E-mail: michael.bevan@bbsrc.ac.uk
FEATURES             Location/Qualifiers
     source          1. .89350
                     /organism="Arabidopsis thaliana"
                     /variety="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="4"
     gene            669. .1151
                     /gene="T12H17.10"
     exon            669. .1151
                     /gene="T12H17.10"
                     /number=1
     CDS             669. .1151
                     /gene="T12H17.10"
                     /note="similarity to auxin-induced protein 10A, Glycine
                     max., PIR2:JQ1099"
                     /codon_start=1
                     /product="predicted protein"
                     /db_xref="PID:e1248670"
                     /db_xref="PID:g2827539"
                     /translation="MKRLRGFKIGHRFVKIFKWIIRSRRNQTGKRQCLTGILNPVTKI
                     YSLARRCLRRGANRLCGGMKPGQTRLGNEPKTPTVPRGHLVVHVGESGEDTRRVVVPV
                     IYFNHPLPGELLEQAERVYGFEQPGRIMIPCRVSDPEKVQMRIAAWDHCRRRKSTFKIL
                     *
     gene            2406. .2983
                     /gene="T12H17.20"
     exon            2406. .2754
                     /gene="T12H17.20"
                     /number=1
     CDS             join(2406. .2754,2844. .2983)
                     /gene="T12H17.20"
                     /note="similarity to predicted proteins, Arabidopsis
                     thaliana"
                     /codon_start=1
                     /product="predicted protein"
                     /db_xref="PID:e1248671"
                     /db_xref="PID:g2827540"
                     /translation="MAYTMKVAVAVGAAVVPLAVVMNPRWTEAQTYPKLDRLCVMMIP
                     DILEECFTHDRLKPTEDCCNDLKNATMTQVDCLCDNFLESLSFSDLSRTFSAGVLKKC
                     DVSHKYMCQAAKNRGEAKGGRNSTTTCDNSITNTSVGGKNKVATSMSAFGLVAILLFV
                     MF"
     intron          2755. .2843
```

```
                     /gene="T12H17.20"
                     /number=1
     exon            2844. .2983
                     /gene="T12H17.20"
                     /number=2
     tRNA            3327. .3400
                     /note="Asn-tRNA_(GTT)"
     gene            4696. .5489
                     /gene="T12H17.30"
     CDS             join(4696. .5035,5407. .5489)
                     /gene="T12H17.30"
                     /note="contains EST gb:N38564"
                     /codon_start=1
                     /product="putative protein"
                     /db_xref="PID:e1248672"
                     /db_xref="PID:g2827541"
                     /translation="MAYTNKVTISAAVATMMLFLAVTIVDAQSMPPMPKFNPVCALAD
                     LPNIVQLCYFNLDLTPSEECCNDLKSSSTIQVNCLCDNFIAHPSNGNISQARYDLVNS
                     ACGVADKFACKGGDASGGSTNKIAASMVLLGLVASLFF"
     exon            4696. .5035
                     /gene="T12H17.30"
                     /number=1
     intron          5036. .5406
                     /gene="T12H17.30"
                     /number=1
     exon            5407. .5489
                     /gene="T12H17.30"
                     /number=2
     exon            6595. .6940
                     /gene="T12H17.40"
                     /number=1
     gene            6595. .7105
                     /gene="T12H17.40"
     CDS             join(6595. .6940,7035. .7105)
                     /gene="T12H17.40"
                     /codon_start=1
                     /product="hypothetical protein"
                     /db_xref="PID:e1248673"
                     /db_xref="PID:g2827542"
                     /translation="MAVAVTAVLFLAVVIAPQWTETKKPPRPSDTSDTSGTSGRDRRT
                     MCPLSIPGIVQNCYATLNAFPSKECCKDLKTASKREVTCLCNNVIAHPDDPLYTNTNQV
                     NKACGVLDKYACDVAPTFTIVYTNYDMIIIAVIQDP"
     intron          6941. .7034
                     /gene="T12H17.40"
                     /number=1
     exon            7035. .7105
                     /gene="T12H17.40"
                     /number=2
     gene            8019. .9209
                     /gene="T12H17.50"
     CDS             complement(8019. .9209)
```



```
                     /gene="T12H17.50"
                     /note="similarity to predicted protein, Arabidopsis
                     thaliana"
                     /codon_start=1
                     /product="predicted protein"
                     /db_xref="PID:e1248674"
                     /db_xref="PID:g2827543"
                     /translation="MEKMQMPNTWSDLPLDLLNLVPKRLSPANFRQAKSVCSGWYSAS
                     KQSVPKNQIPWLMLFPKDKNNNKNSSCTIFFNPEDKDQLYQTQDLGVEFAKSVCLATY
                     GSWLLMQDSKYNLYILNPFTYEKIGLPAIESQQVGMVKVDQTIDDDFLTFDDHNHVKL
                     FKGNNTVRTPVFWIDEKTKEYIALWGLGYWCVVYAKNGDKLWNQIPEIILDSLDMVYK
                     DHKLYSPSYRNLFTILDFSGEIPRKAPQCFMHVYRSEWLSPRSRQLSNSWCVAETKLV
                     VTVTGDVLLVERMLRHWSRIQSFNVYKFYSSGTFLDKYELADSLGDEAMLLDLGITVL
                     ANEVEGLHRNTIYFSDSHDTTTKDLFLFNYETREMEPLHKFDPDCSLLELSARWFLPS
                     FSHT"
     exon            complement(8019. .9209)
                     /gene="T12H17.50"
                     /number=1
     exon            9990. .10023
                     /gene="T12H17.60"
                     /number=1
     gene            9990. .13709
                     /gene="T12H17.60"
     CDS             join(9990. .10023,10239. .10535,10564. .10664,10910. .11292,
                     11755. .11809,11898. .12053,12143. .12311,12391. .12455,
                     12534. .12674,12769. .12885,13024. .13096,13244. .13380,
                     13468. .13525,13612. .13709)
                     /gene="T12H17.60"
                     /note="similarity to p48, Homo sapiens, PATCHX:G904032;
                     contains EST gb:Aa605571, H76919, H76638, R30271, Z25650"
                     /codon_start=1
                     /product="HSP associated protein like"
                     /db_xref="PID:e1248675"
                     /db_xref="PID:g2827544"
                     /translation="MRRTWVLAEESSVAIAPLLAEPQSTMFPEMTPECATVMPDLLEK
                     CFATGSVTPTEDCCTDLKSATSTQVTCLCDNYIANPAVSNITGPYSKAITTKCGVPDK
                     YSCDCTSKVISNKNGAPMVINGFWFCVVKEEKKRKEEAAVAMGVLTLLGLINGLIVAH
                     GKTRQQNKPVCKLILFQFQHFGFCSIRFSANKKARSRSKTGDAYPFRNEKVLEQKPYS
                     LPFVPQILEIETRLKTKKTRLLTLRPKPTLPLFLSLSRRSEADFSQKDGFNEAYLGAK
                     IPTGVHEEDKDTKPRSFVVEESDDDMDETEEVKPKVEEEEEEDEIVESDVELEGDTVE
                     PDNDPPQKMGDSSVEVTDENREAAQEAKGKAMEALSEGNPDEAIEHLTRAITLNPTSA
                     IMYGNRASVYIKLKKPNAAIRDANAALEINPDSAKGYKSRGMARAMLGEWAEAAKDLH
                     LASTIDYDEEISAVLKKVEPNAHKLEEHRRKYDRLRKEREDKKAERDRLRRRAEAQAA
                     YDKAKKEEQSSSSRPSGGGFPGGMGGGMPGMGGGMPAGMGGGGMPGAGGGMPGGGGMP
                     GGMDFSKILNDPELMTAFSDPEVMAALQDVMKNPANLAKHQANPKVAPVIAKMMGKFA
                     GPQ"
     intron          10024. .10238
                     /gene="T12H17.60"
                     /number=1
     exon            10239. .10535
                     /gene="T12H17.60"
```



```
                     /number=2
     intron          10536. .10563
                     /gene="T12H17.60"
                     /number=2
     exon            10564. .10664
                     /gene="T12H17.60"
                     /number=3
     intron          10665. .10909
                     /gene="T12H17.60"
                     /number=3
     exon            10910. .11292
                     /gene="T12H17.60"
                     /number=4
     intron          11293. .11754
                     /gene="T12H17.60"
                     /number=4
     exon            11755. .11809
                     /gene="T12H17.60"
                     /number=5
     intron          11810. .11897
                     /gene="T12H17.60"
                     /number=5
     exon            11898. .12053
                     /gene="T12H17.60"
                     /number=6
     intron          12054. .12142
                     /gene="T12H17.60"
                     /number=6
     exon            12143. .12311
                     /gene="T12H17.60"
                     /number=7
     intron          12312. .12390
                     /gene="T12H17.60"
                     /number=7
...
Note: remainder of annotations omitted.

  Query Match           6.0%;  Score 18;  DB 23;  Length 89350;
  Best Local Similarity 100.0%;  Pred. No. 2.92e+00;
  Matches    18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  80013 TTTTTATAAATGTATAAA 80030
          ||||||||||||||||||
Qy     97 TTTTTATAAATGTATAAA 114


RESULT   17
LOCUS       AC003065    90832 bp    DNA             PRI       10-SEP-1998
DEFINITION  Homo sapiens Chromosome 2p13 BAC Clone h173, complete sequence.
ACCESSION   AC003065
NID         g3581728
```



```
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 90832)
  AUTHORS   Hua,A. and Roe,B.A.
  TITLE     Homo sapiens Chromosome 2p13 BAC Clone h173
  JOURNAL   Unpublished (1997)
REFERENCE   2  (bases 1 to 90832)
  AUTHORS   Meisler,M.
  JOURNAL   Unpublished (1997)
REFERENCE   3  (bases 1 to 90832)
  AUTHORS   Hua,A. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-NOV-1997) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   4  (bases 1 to 90832)
  AUTHORS   Hua,A. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-SEP-1998) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On Sep 10, 1998 this sequence version replaced gi:3548823.
FEATURES             Location/Qualifiers
     source          1. .90832
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2p13"
                     /clone="h173"
BASE COUNT    27135 a  19291 c  18194 g  26212 t
ORIGIN

  Query Match           6.0%;  Score 18;  DB 27;  Length 90832;
  Best Local Similarity 100.0%;  Pred. No. 2.92e+00;
  Matches    18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  35999 AATGTATAAATGTATCAA 36016
          ||||||||||||||||||
Qy    105 AATGTATAAATGTATCAA 122


RESULT   18
LOCUS       HUMDGCRCEN 108400 bp    DNA             PRI       14-JUN-1996
DEFINITION  Homo sapiens DiGeorge syndrome critical region, centromeric end.
ACCESSION   L77570
NID         g1377755
KEYWORDS    .
SOURCE      Homo sapiens DNA.
```

```
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 108400)
  AUTHORS   Gong,W., Emanuel,B.S., Collins,J., Kim,D.H., Wang,Z., Chen,F.,
            Zhang,G., Roe,B. and Budarf,M.L.
  TITLE     A transcription map of the DiGeorge and velo-cardio-facial syndrome
            minimal critical region on 22q11
  JOURNAL   Hum. Mol. Genet. 5 (6), 789-800 (1996)
  MEDLINE   96372815
FEATURES             Location/Qualifiers
     source          1. .108400
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT    29924 a  28123 c  26278 g  24037 t     38 others
ORIGIN

  Query Match             6.0%;  Score 18;  DB 27;  Length 108400;
  Best Local Similarity  94.7%;  Pred. No. 2.92e+00;
  Matches    18;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db  21784 ATGAAGGCACACCTGGTGC 21802
          ||||||||||  ||||||||
Cp    182 ATGAAGGCACNCCTGGTGC 164


RESULT   19
LOCUS       ATF7H19    139316 bp    DNA             PLN       17-SEP-1998
DEFINITION  Arabidopsis thaliana DNA chromosome 4, BAC clone F7H19 (ESSAII
            project).
ACCESSION   AL031018
NID         g3292807
KEYWORDS    .
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Charophyta/Embryophyta group;
            Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta;
            Magnoliophyta; eudicotyledons; Rosidae; Capparales; Brassicaceae;
            Arabidopsis.
REFERENCE   1  (bases 1 to 139316)
  AUTHORS   Peters,S.A., van Staveren,M., Dirkse,W., Stiekema,W., Bancroft,I.,
            Mewes,H.W., Mayer,K.F.X., Schueller,C. and Bevan,M.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 139316)
  AUTHORS   EU Arabidopsis sequencing,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUL-1998) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
            schuelle@mips.biochem.mpg.de Project Coordinator: Mike Bevan,
            Molecular Genetics Department, Cambridge Laboratory, John Innes
            Centre, Colney Lane, NR4 7UJ Norwich, UK, E-mail:
```



```
            michael.bevan@bbsrc.ac.uk
FEATURES             Location/Qualifiers
     source          1. .139316
                     /organism="Arabidopsis thaliana"
                     /variety="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="4"
     exon            complement(915. .1445)
                     /gene="F7H19.10"
                     /number=1
     gene            915. .1445
                     /gene="F7H19.10"
     CDS             complement(915. .1445)
                     /gene="F7H19.10"
                     /note="similarity to PVPR3 protein, Phaseolus vulgaris,
                     PID:g169363; contains EST gb:AA597458, H36610"
                     /codon_start=1
                     /product="putative protein"
                     /db_xref="PID:e1310051"
                     /db_xref="PID:g3292808"
                     /translation="MGSEQNDSTSFTQSQASEPKLCVKGCGFFGSPSNMDLCSKCYRG
                     ICAEEAQTAVAKAAVEKSFKPSPPRSLFIAEPPAVVVEPKPEKAAVVVVSAEPSSSAV
                     PEANEPSRPARTNRCLCCNKKVGIMGFKCKCGSTFCGEHRYPETHDCSFDFKEVGRGE
                     IAKANPVVKADKIQRF"
     gene            4680. .6720
                     /gene="F7H19.20"
     CDS             complement(join(4680. .4751,4828. .4910,5004. .5257,
                     5355. .5607,5685. .5858,5956. .6100,6319. .6720))
                     /gene="F7H19.20"
                     /note="similarity to Na(+) dependent transporter, Aquifex
                     aeolicus, PIR2:E70482"
                     /codon_start=1
                     /product="putative protein"
                     /db_xref="PID:e1310052"
                     /db_xref="PID:g3292809"
                     /translation="MSVITTPIETLHLKSTLRLLPRAVYRSQRIQVFPPNIFSNTSLS
                     SPLRIDPISQGLFILLSSILRRLCLLGTCKIYTFFSHGFLQPTIYTVGGSRNLWRRYA
                     SDNFSEMGLDPGADPFKVCAMGSPLFHSMLCWVIEKPSIVDRMKKANSILPHVVLAST
                     ILALIYPPSFTWFTSRFFPLEQGFLMFAVGINSNEKDFLEAFKRPKAILLGYVGQYLV
                     KPVLGFIPGLAAVSLFQLPTPIGAGIMLVSCVSGAQLSNYATFLTDPALAPLSIVMTS
                     LSTATAVLVTPMLSLLLIGKKLPVDVKGMISSILQVVIAPIAAGLLLNKLFPKVSNAI
                     RPFLPILSVLDTACCVGAPLALNINSVMSPFGATILLLVTMFHLSAFLAGYFLTGSVF
                     RNAPDAKAMQRTLSYETGMQSSLLALALATKFPQDPLVGIPPAISTVVMSLMGFTLVL
                     IWSKEKSNTF"
     exon            complement(4680. .4751)
                     /gene="F7H19.20"
                     /number=1
     intron          complement(4752. .4827)
                     /gene="F7H19.20"
                     /number=1
     exon            complement(4828. .4910)
```



```
                     /gene="F7H19.20"
                     /number=2
     intron          complement(4911. .5003)
                     /gene="F7H19.20"
                     /number=2
     exon            complement(5004. .5257)
                     /gene="F7H19.20"
                     /number=3
     intron          complement(5258. .5354)
                     /gene="F7H19.20"
                     /number=3
     exon            complement(5355. .5607)
                     /gene="F7H19.20"
                     /number=4
     intron          complement(5608. .5684)
                     /gene="F7H19.20"
                     /number=4
     exon            complement(5685. .5858)
                     /gene="F7H19.20"
                     /number=5
     intron          complement(5859. .5955)
                     /gene="F7H19.20"
                     /number=5
     exon            complement(5956. .6100)
                     /gene="F7H19.20"
                     /number=6
     intron          complement(6101. .6318)
                     /gene="F7H19.20"
                     /number=6
     exon            complement(6319. .6720)
                     /gene="F7H19.20"
                     /number=7
     CDS             join(7238. .7489,8330. .8547,8802. .8985)
                     /gene="F7H19.30"
                     /note="similarity to hypothetical protein, Synechocystis
                     sp., PIR2:S76980; contains EST gb:F14154, F14162"
                     /codon_start=1
                     /product="putative protein"
                     /db_xref="PID:e1310053"
                     /db_xref="PID:g3292810"
                     /translation="MTVTEDEKPVLELLVRVEDDDEKGPYVKLSEVLEKGQESEQEKE
                     EEKTDSSRFWFWVKLSLLFAFLAALAVVGYIWIGPLIMDKGWLERYPDQAAVLRAAGE
                     GNWLHQFLLVTLIRISPFPYILYNYCSVATRVKYGPYITGSLLGMVPEVPVAIYTGIL
                     VRTLAEASSAEEQGLSVTQVILNILGFLATVATTILITKYAKRQLETMKKEDEALLQ"
     exon            7238. .7489
                     /gene="F7H19.30"
                     /number=1
     gene            7238. .8985
                     /gene="F7H19.30"
     intron          7490. .8329
                     /gene="F7H19.30"
```



```
                     /number=1
     exon            8330. .8547
                     /gene="F7H19.30"
                     /number=2
     intron          8548. .8801
                     /gene="F7H19.30"
                     /number=2
     exon            8802. .8985
                     /gene="F7H19.30"
                     /number=3
     exon            10779. .10953
                     /gene="F7H19.40"
                     /number=1
     gene            10779. .14325
                     /gene="F7H19.40"
     CDS             join(10779. .10953,11160. .11318,11459. .11545,11780. .11912,
                     12005. .12116,12278. .12328,12422. .12519,12603. .12678,
                     12773. .12835,12931. .13005,13132. .13190,13307. .13385,
                     13465. .13572,13791. .13853,14079. .14185,14307. .14325)
                     /gene="F7H19.40"
                     /codon_start=1
                     /product="hypothetical protein"
                     /db_xref="PID:e1310054"
                     /db_xref="PID:g3292811"
                     /translation="MEMDEDMEIEPMVPEVKEIDLEYEFDAARWYDPTREELPEESQA
                     AEFWFHSAPSYAPSPPVTKLLREEVYDDKTEASTRSEDEEVVADVCDKYKGTPQQPN
                     HLATDFNKTGSRMRSGIFSSQQQQQGCNLKKVPWQPIWKGPTVSNHKHTDKPKFRARSS
                     IRPTPRSSTLMRPTASQLAKQNNASKFHMQVDQSHDKGSCGTSGTEVQASKRQKLDGG
                     LLRKVADTKQEMSFVHKIPKKDTTLDKNAQHARTKITIPQEPDFATSQRAHRIRHKND
                     AKLEQEQPAVYRPKARPFNRKIFDAPSLPIRKKSTPKLPEPQEPHLKTSERAMQHSSA
                     VTTRSHAYKESDKSIITDMLDGDNRESRRPSGMDISKHDVSEGSHVFKARPLNKKILS
                     SRGDMGIFKNSKRETTVPLVWQSYQLPFALDLLLSIKSEIQPNDGSRTRLFQPKVTRM
                     AAAKPVSSAGQQMQPGNSGITPETNQQWTARRNLGIR"
     intron          10954. .11159
                     /gene="F7H19.40"
                     /number=1
     exon            11160. .11318
                     /gene="F7H19.40"
                     /number=2
     intron          11319. .11458
                     /gene="F7H19.40"
                     /number=2
     exon            11459. .11545
                     /gene="F7H19.40"
                     /number=3
     intron          11546. .11779
                     /gene="F7H19.40"
                     /number=3
     exon            11780. .11912
                     /gene="F7H19.40"
                     /number=4
```

```
     intron          11913. .12004
                     /gene="F7H19.40"
                     /number=4
     exon            12005. .12116
                     /gene="F7H19.40"
                     /number=5
     intron          12117. .12277
                     /gene="F7H19.40"
                     /number=5
     exon            12278. .12328
                     /gene="F7H19.40"
                     /number=6
     intron          12329. .12421
                     /gene="F7H19.40"
                     /number=6
     exon            12422. .12519
                     /gene="F7H19.40"
                     /number=7
     intron          12520. .12602
                     /gene="F7H19.40"
                     /number=7
     exon            12603. .12678
                     /gene="F7H19.40"
...
Note: remainder of annotations omitted.

  Query Match           6.0%;  Score 18;  DB 24;  Length 139316;
  Best Local Similarity 100.0%;  Pred. No. 2.92e+00;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      19 TTTTTATAAATGTATAAA 36
           ||||||||||||||||||
Qy      97 TTTTTATAAATGTATAAA 114


RESULT   20
LOCUS       HSU52111   153460 bp    DNA             PRI       11-SEP-1997
DEFINITION  Homo sapiens Xq28 genomic DNA in the region of the ALD locus
            containing the genes for creatine transporter (SLC6A8), CDM,
            adrenoleukodystrophy (ALD), Na+-isocitrate dehydrogenase gamma
            subunit (IDH), and translocon-associated protein delta (TRAP)
            genes, complete cds, plexin related protein (PLEXR) and serine
            kinase (SK) genes, partial cds, Xq28lul gene and cytochrome C (CCp)
            pseudogene.
ACCESSION   U52111
NID         g1302649
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
```



```
REFERENCE   1  (bases 1 to 153460)
  AUTHORS   Brenner,V., Nyakatura,G., Rosenthal,A. and Platzer,M.
  TITLE     Genomic organization of two novel genes on human Xq28: compact head
            to head arrangement of IDH gamma and TRAP delta is conserved in rat
            and mouse
  JOURNAL   Genomics 44 (1), 8-14 (1997)
  MEDLINE   97432815
REFERENCE   2  (bases 1 to 153460)
  AUTHORS   Platzer,M., Bauer,D. and Drescher,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAR-1996) Matthias Platzer, Institute of Molecular
            Biotechnology, Department of Genome Analysis, Beutenbergstr. 11,
            Jena 07745 Germany
FEATURES             Location/Qualifiers
     source          1. .153460
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="Xq28"
                     /chromosome="X"
     mRNA            join(1042. .1876,3415. .3546,4344. .4593,5015. .5147,
                     6081. .6215,6303. .6406,6502. .6626,6945. .7057,7170. .7307,
                     7384. .7486,7573. .7673,7759. .7929,8114. .9625)
                     /gene="SLC6A8"
     exon            1042. .1876
                     /gene="SLC6A8"
                     /number=1
     5'UTR           1042. .1614
                     /gene="SLC6A8"
     gene            1042. .9625
                     /gene="SLC6A8"
     repeat_region   1122. .1607
                     /rpt_family="TAR1"
     CDS             join(1615. .1876,3415. .3546,4344. .4593,5015. .5147,
                     6081. .6215,6303. .6406,6502. .6626,6945. .7057,7170. .7307,
                     7384. .7486,7573. .7673,7759. .7929,8114. .8254)
                     /gene="SLC6A8"
                     /note="CRTR"
                     /codon_start=1
                     /product="creatine transporter"
                     /db_xref="PID:g1302650"
                     /translation="MAKKSAENGIYSVSGDEKKGPLIAPGPDGAPAKGDGPVGLGTPG
                     GRLAVPPRETWTRQMDFIMSCVGFAVGLGNVWRFPYLCYKNGGGVFLIPYVLIALVGG
                     IPIFFLEISLGQFMKAGSINVWNICPLFKGLGYASMVIVFFCNTYYIMVLAWGFYYLV
                     KSFTTGLPWATCGHTWNTPDCVEIFRHEDCANASLANLTCDQLADRRSPVIEFWENKV
                     LRLSGGLEVPGALNWEVTLCLLACWVLVYFCVWKGVKSTGKIVTFTATFPYVVLVVLL
                     VRGVLLPGALDGIIYYLKPDWSKLGSPQVWIDAGTQIFFSYAIGLGALTALGSYNRFN
                     NNCYKDAIILALINSGTSFFAGFVVFSILGFMAAEQGVHISKVAESGPGLAFIAYPRA
                     VTLMPVAPLWAALFFFMLLLLGLDSQFVGVEGFITGLLDLLPASYYFRFQREISVALC
                     CALCFVIDLSMVTDGGMYVFQLFDYYSASGTTLLWQAFWECVVVAWVYGADRFMDDIA
                     CMIGYRPCPWMKWCWSFFTPLVCMGIFIFNVVYYEPLVTNNTYVYPWWGEAMGWAFAL
                     SSMLCVPLHLLGCLLRAKGTMAERWQHLTQPIWGLHHLEYRAQDADVRGLTTLTPVSE
```



```
                     SSKVVVVESVM"
     exon            3415. .3546
                     /gene="SLC6A8"
                     /number=2
     exon            4344. .4593
                     /gene="SLC6A8"
                     /number=3
     exon            5015. .5147
                     /gene="SLC6A8"
                     /number=4
     exon            6081. .6215
                     /gene="SLC6A8"
                     /number=5
     exon            6303. .6406
                     /gene="SLC6A8"
                     /number=6
     exon            6502. .6626
                     /gene="SLC6A8"
                     /number=7
     exon            6945. .7057
                     /gene="SLC6A8"
                     /number=8
     exon            7170. .7307
                     /gene="SLC6A8"
                     /number=9
     exon            7384. .7486
                     /gene="SLC6A8"
                     /number=10
     exon            7573. .7673
                     /gene="SLC6A8"
                     /number=11
     exon            7759. .7929
                     /gene="SLC6A8"
                     /number=12
     exon            8114. .9625
                     /gene="SLC6A8"
                     /number=13
     3'UTR           8255. .9625
                     /gene="SLC6A8"
     repeat_region   10650. .11006
                     /rpt_family="L1MB7"
     mRNA            complement(join(13537. .14013,15045. .15145,15973. .16096,
                     16997. .17132,28581. .28728,33911. .34011,36192. .36327,
                     37373. .37464))
                     /gene="CDM"
     exon            complement(13537. .14013)
                     /gene="CDM"
                     /number=8
     gene            complement(13537. .37464)
                     /gene="CDM"
     3'UTR           complement(13537. .13974)
```



```
     CDS             /gene="CDM"
                     complement(join(13975. .14013,15045. .15145,15973. .16096,
                     16997. .17132,28581. .28728,33911. .34011,36192. .36283))
                     /gene="CDM"
                     /note="DXS1375E"
                     /codon_start=1
                     /product="CDM protein"
                     /db_xref="PID:g1508820"
                     /translation="MSLQWTAVATFLYAEVFVVLLLCIPPISPKRWQKIFKSRLVELL
                     VSYGNTFFVVLIVILVLLVIDAVREIRKYDDVTEKVNLQNNPGAMEHFHMKLFRAQRN
                     LYIAGFSLLLSFLLRRLVTLISQQATLLASNEAFKKQAESASEAAKKYMEENDQLKKG
                     AAVDGGKLDVGNAEVKLEEENRSLKADLQKLKDELASTKQKLEKAENQVLAMRKQSEG
                     LTKETDRLLEERAKLQAAVDGPMDKKEE"
     exon            complement(15045. .15145)
                     /gene="CDM"
                     /number=7
     exon            complement(15973. .16096)
                     /gene="CDM"
                     /number=6
     exon            complement(16997. .17132)
                     /gene="CDM"
                     /number=5
     repeat_region   complement(17561. .17736)
                     /rpt_family="MIR"
     repeat_region   24665. .24951
                     /rpt_family="Alu-Sz"
     repeat_region   25179. .25350
                     /rpt_family="Alu-Jo"
     repeat_region   25351. .25449
                     /rpt_family="Alu-J"
     repeat_region   25963. .26253
                     /rpt_family="Alu-Ya"
     repeat_region   complement(26937. .27215)
                     /rpt_family="Alu-Sx"
     exon            complement(28581. .28728)
                     /gene="CDM"
                     /number=4
     repeat_region   28915. .28949
                     /rpt_family="Alu"
     repeat_region   28950. .29213
                     /rpt_family="Alu-Sq"
     repeat_region   29498. .29700
                     /rpt_family="Alu-Jo"
     repeat_region   29701. .29804
                     /rpt_family="Alu-J"
     repeat_region   complement(30498. .30753)
                     /rpt_family="Alu-Jb"
     repeat_region   31073. .31364
                     /rpt_family="Alu-Sz"
     repeat_region   complement(33413. .33702)
                     /rpt_family="Alu-Sq"
```

```
     exon            complement(33911. .34011)
                     /gene="CDM"
                     /number=3
     repeat_region   34205. .34482
                     /rpt_family="Alu-Sz"
     repeat_region   complement(34545. .34615)
                     /rpt_family="Alu-J"
     repeat_region   34645. .34945
                     /rpt_family="Alu-Sx"
     exon            complement(36192. .36327)
                     /gene="CDM"
                     /number=2
     exon            complement(37373. .37464)
                     /gene="CDM"
                     /number=1
     5'UTR           complement(join(37373. .37464,36284. .36327))
                     /gene="CDM"
     mRNA            join(37920. .39205,42271. .42451,49150. .49292,49383. .49551,
                     50195. .50289,53130. .53275,53612. .53757,56025. .56109,
                     56259. .56384,56527. .56899)
                     /gene="ALD"
     gene            37920. .56899
                     /gene="ALD"
     exon            37920. .39205
                     /gene="ALD"
                     /number=1
     5'UTR           37920. .38305
                     /gene="ALD"
     CDS             join(38306. .39205,42271. .42451,49150. .49292,49383. .49551,
                     50195. .50289,53130. .53275,53612. .53757,56025. .56109,
                     56259. .56384,56527. .56773)
                     /gene="ALD"
...
Note: remainder of annotations omitted.

  Query Match           6.0%;  Score 18;  DB 27;  Length 153460;
  Best Local Similarity  90.0%;  Pred. No. 2.92e+00;
  Matches   18;  Conservative    0;  Mismatches   2;  Indels   0;  Gaps   0;

Db  62686 TTTGGTTGATCTGGGCCAGG 62705
          | ||||||| ||||||||||
Qy     43 TNTGGTTGANCTGGGCCAGG 62


RESULT   21
LOCUS       HSAC002069 158537 bp    DNA             PRI       09-MAY-1997
DEFINITION  Human BAC clone RG326K09 from 7q21, complete sequence.
ACCESSION   AC002069
NID         g2076718
KEYWORDS    HTG.
SOURCE      human.
```



```
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 158537)
  AUTHORS   Geisel,C, Beck,C, Smith,A and Twyman,B.
  TITLE     The sequence of H. sapiens BAC clone RG326K09
  JOURNAL   Unpublished (1997)
REFERENCE   2  (bases 1 to 158537)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-1997)
COMMENT     Genome Sequencing Center
            Department of Genetics, Washington University
            St. Louis, MO 63108, USA
            http://genome.wustl.edu/gsc
            e-mail: sapiens@watson.wustl.edu

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            sections once, or longer because we provide a small overlap between
            neighboring submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded or sequenced with an alternate
            chemistry; an attempt was made to resolve all sequencing problems,
            such as compressions and repeats; all regions were covered by
            sequence from more than one subclone; and the assembly was
            confirmed by restriction digest.

            MAPPING INFORMATION:
            The sequence of this clone was established as part of a mapping and
            sequencing collaboration between the NHGRI Chromosome 7 Mapping
            Project and the Washington University Genome Sequencing Center.
            For additional information about the map position of this sequence,
            see http://www.nhgri.nih.gov/DIR/GTB/CHR7 or send an E-mail to
            egreen@nhgri.nih.gov

            SOURCE INFORMATION:
            This clone is from a release of the human BAC library. The library
            contains cloned DNA from a human sperm.  For references see:
            Shizuya et al., Proc. Natl. Acad. Sci. 89:8794-8797 (1992); Kim et
            al., Genomics 34:213-218 (1996).
            VECTOR:  pBELO
            Selection:  chloramphenicol

            NEIGHBORING SEQUENCE INFORMATION:
            Actual start of this clone is at base position 1 of H_RG326K09;
            actual end is at 158537 of H_RG326K09.  This clone lies in an
            unanchored cluster of unknown orientation.
```



```
            This clone contains sWSS3233 (NID:g1113712).
FEATURES            Location/Qualifiers
     source         1. .158537
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="7"
                    /clone="RG326K09"
                    /clone_lib="CITB-HS-A"
                    /map="7q21"
     repeat_region  276. .1999
                    /rpt_family="L1"
     repeat_region  complement(580. .998)
                    /rpt_family="L1"
     repeat_region  3260. .3282
                    /rpt_family="L1"
     repeat_region  4062. .4178
                    /rpt_family="L1"
     repeat_region  4193. .4376
                    /rpt_family="L1"
     repeat_region  4553. .4725
                    /rpt_family="L1"
     repeat_region  4767. .4808
                    /rpt_family="L1"
     repeat_region  5097. .6621
                    /rpt_family="L1"
     repeat_region  complement(5236. .5642)
                    /rpt_family="L1"
     repeat_region  6764. .6910
                    /rpt_family="L1"
     repeat_region  6922. .6974
                    /rpt_family="L1"
     repeat_region  complement(6997. .7223)
                    /rpt_family="ALU"
     repeat_region  7606. .7712
                    /rpt_family="L1"
     repeat_region  7866. .7936
                    /rpt_family="L1"
     repeat_region  11439. .11461
                    /rpt_family="L1"
     repeat_region  11641. .11715
                    /rpt_family="L1"
     repeat_region  12310. .12397
                    /rpt_family="L1"
     repeat_region  12589. .12629
                    /rpt_family="L1"
     repeat_region  13203. .13236
                    /rpt_family="L1"
     repeat_region  13854. .14031
                    /rpt_family="L1"
     repeat_region  16674. .17204
                    /rpt_family="MER"
```



```
     repeat_region  complement(17532. .17751)
                    /rpt_family="L1"
     repeat_region  complement(19040. .19066)
                    /rpt_family="L1"
     misc_feature   20125. .21945
                    /note="Probable pseudogene. Similar to hSRPlalpha U28386
                    (NID:g899538)."
     misc_feature   20132. .20288
                    /note="similar to human EST W56840 (NID:g1358716)
                    zc97d02.r1"
     misc_feature   20278. .20388
                    /note="similar to human EST W56840 (NID:g1358716)
                    zc97d02.r1"
     repeat_region  complement(26748. .28317)
                    /rpt_family="THR"
     repeat_region  29378. .35406
                    /rpt_family="L1"
     repeat_region  complement(33995. .34413)
                    /rpt_family="L1"
     repeat_region  complement(35884. .35905)
                    /rpt_family="L1"
     repeat_region  complement(35910. .36194)
                    /rpt_family="ALU"
     misc_feature   40798. .45213
                    /note="Probable retroviral pseudogene. Similar to POL_GALV
                    POL POLYPROTEIN P21414 (NID:g130602)."
     repeat_region  complement(46549. .47052)
                    /rpt_family="MER"
     repeat_region  complement(47102. .47157)
                    /rpt_family="MER"
     repeat_region  47521. .50154
                    /rpt_family="L1"
     repeat_region  51874. .52163
                    /rpt_family="ALU"
     repeat_region  53038. .53063
                    /rpt_family="L1"
     repeat_region  complement(56184. .56476)
                    /rpt_family="ALU"
     repeat_region  56478. .56736
                    /rpt_family="ALU"
     repeat_region  59292. .60851
                    /rpt_family="THR"
     repeat_region  61080. .61206
                    /rpt_family="L1"
     repeat_region  complement(65139. .65325)
                    /rpt_family="L1"
     repeat_region  75780. .76108
                    /rpt_family="L1"
     repeat_region  78275. .78293
                    /rpt_family="L1"
     repeat_region  79752. .79781
```

```
                     /rpt_family="L1"
     repeat_region   80350. .80411
                     /rpt_family="ALU"
     repeat_region   80412. .80438
                     /rpt_family="L1"
     repeat_region   80761. .83307
                     /rpt_family="L1"
     repeat_region   83308. .83601
                     /rpt_family="ALU"
     repeat_region   83602. .83652
                     /rpt_family="L1"
     repeat_region   83744. .83767
                     /rpt_family="L1"
     repeat_region   83784. .83813
                     /rpt_family="L1"
     repeat_region   84400. .84452
                     /rpt_family="L1"
     repeat_region   84599. .84655
                     /rpt_family="L1"
     repeat_region   complement(85751. .85835)
                     /rpt_family="L1"
     repeat_region   85770. .85847
                     /rpt_family="L1"
     repeat_region   89152. .89193
                     /rpt_family="L1"
     repeat_region   93896. .93929
                     /rpt_family="L1"
     repeat_region   complement(95512. .95708)
                     /rpt_family="ALU"
     repeat_region   97302. .97346
                     /rpt_family="L1"
     repeat_region   98231. .98289
                     /rpt_family="L1"
     repeat_region   98500. .98555
                     /rpt_family="L1"
     repeat_region   complement(99131. .99423)
                     /rpt_family="ALU"
     repeat_region   107834. .107888
                     /rpt_family="L1"
     repeat_region   108592. .108880
                     /rpt_family="ALU"
...
Note: remainder of annotations omitted.

  Query Match             6.0%;  Score 18;  DB 26;  Length 158537;
  Best Local Similarity 100.0%;  Pred. No. 2.92e+00;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db 141627 TAAGAGTAGCAGGAACTC 141644
          ||||||||||||||||||
Qy     65 TAAGAGTAGCAGGAACTC 82
```



```
RESULT   22
LOCUS       AC005033   167468 bp    DNA             HTG       12-JUN-1998
DEFINITION  *** SEQUENCING IN PROGRESS *** Homo sapiens clone NH0140K04; HTGS
            phase 1, 4 unordered pieces.
ACCESSION   AC005033
NID         g3212958
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 167468)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 167468)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-1998) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     ***                                                          ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                          ***
            * This sequence is unfinished. It consists of 4 contigs for
            * which the order is not known; their order in this record is
            * arbitrary. In some cases, the exact lengths of the gaps
            * between the contigs are also unknown; these gaps are  presented
            * as runs of N as a convenience only. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
            *       1     8201: contig of 8201 bp in length
            *    8202     8219: gap of unknown length
            *    8220    35367: contig of 27148 bp in length
            *   35368    35385: gap of unknown length
            *   35386    89438: contig of 54053 bp in length
            *   89439    89456: gap of unknown length
            *   89457   167468: contig of 78012 bp in length.
FEATURES             Location/Qualifiers
     source          1. .167468
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="NH0140K04"
BASE COUNT    45231 a  38247 c  39065 g  44871 t     54 others
ORIGIN

  Query Match             6.0%;  Score 18;  DB 17;  Length 167468;
  Best Local Similarity 100.0%;  Pred. No. 2.92e+00;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```



```
Db  75391 TTGATACATTTATACATT 75408
          ||||||||||||||||||
Cp    122 TTGATACATTTATACATT 105


RESULT   23
LOCUS       AC005182   175162 bp    DNA             HTG       18-SEP-1998
DEFINITION  *** SEQUENCING IN PROGRESS *** Homo sapiens chromosome 12p13.3;
            HTGS phase 1, 48 unordered pieces.
ACCESSION   AC005182
NID         g3628574
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 175162)
  AUTHORS   Muzny,D., Arenson,A.D., Brundage,E., Carvelli,K., Chen,E., Di,W.,
            Ding,Y., Dugan,S., Durbin,J., Forcum,J., Ganesh,R., Garcia,C.,
            Goodman,M., Gorrell,J.H., Haywood,M., Jackson,L., Kampal,R.,
            Karpathy,S., Leal,B., Li,Y., Liu,W., Logan,O., Lu,J., Ly,T.,
            Martinez,C., Oswal,G., Perez,L., Rashid,N.D., Rowland,K.,
            Savage,L., Scherer,S.S., Shen,H., Timms,K.M., Todd,J., Vo,Q.,
            Worley,K.C., Yu,W., Chinault,C., Nelson,D. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 175162)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JUN-1998) Molecular and Human Genetics, Baylor
            College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Sep 18, 1998 this sequence version replaced gi:3366552.
            ***                                                          ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                          ***
            * This sequence is unfinished. It consists of 48 contigs for
            * which the order is not known; their order in this record is
            * arbitrary. In some cases, the exact lengths of the gaps
            * between the contigs are also unknown; these gaps are  presented
            * as runs of N as a convenience only. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
            *       1   131304: contig of 131304 bp in length
            *  131305   131331: gap of unknown length
            *  131332   133356: contig of 2025 bp in length
            *  133357   133383: gap of unknown length
            *  133384   134849: contig of 1466 bp in length
            *  134850   134876: gap of unknown length
            *  134877   136341: contig of 1465 bp in length
            *  136342   136368: gap of unknown length
```



```
            *  136369   137808: contig of 1440 bp in length
            *  137809   137835: gap of unknown length
            *  137836   139222: contig of 1387 bp in length
            *  139223   139249: gap of unknown length
            *  139250   140478: contig of 1229 bp in length
            *  140479   140505: gap of unknown length
            *  140506   141659: contig of 1154 bp in length
            *  141660   141686: gap of unknown length
            *  141687   142766: contig of 1080 bp in length
            *  142767   142793: gap of unknown length
            *  142794   143822: contig of 1029 bp in length
            *  143823   143849: gap of unknown length
            *  143850   144859: contig of 1010 bp in length
            *  144860   144886: gap of unknown length
            *  144887   145895: contig of 1009 bp in length
            *  145896   145922: gap of unknown length
            *  145923   146929: contig of 1007 bp in length
            *  146930   146956: gap of unknown length
            *  146957   147943: contig of 987 bp in length
            *  147944   147970: gap of unknown length
            *  147971   148945: contig of 975 bp in length
            *  148946   148971: gap of unknown length
            *  148972   149934: contig of 963 bp in length
            *  149935   149960: gap of unknown length
            *  149961   150883: contig of 923 bp in length
            *  150884   150909: gap of unknown length
            *  150910   151818: contig of 909 bp in length
            *  151819   151844: gap of unknown length
            *  151845   152752: contig of 908 bp in length
            *  152753   152778: gap of unknown length
            *  152779   153666: contig of 888 bp in length
            *  153667   153692: gap of unknown length
            *  153693   154572: contig of 880 bp in length
            *  154573   154598: gap of unknown length
            *  154599   155442: contig of 844 bp in length
            *  155443   155468: gap of unknown length
            *  155469   156308: contig of 840 bp in length
            *  156309   156334: gap of unknown length
            *  156335   157173: contig of 839 bp in length
            *  157174   157199: gap of unknown length
            *  157200   157982: contig of 783 bp in length
            *  157983   158008: gap of unknown length
            *  158009   158821: contig of 813 bp in length
            *  158822   158847: gap of unknown length
            *  158848   159657: contig of 810 bp in length
            *  159658   159683: gap of unknown length
            *  159684   160487: contig of 804 bp in length
            *  160488   160513: gap of unknown length
            *  160514   161312: contig of 799 bp in length
            *  161313   161338: gap of unknown length
            *  161339   162131: contig of 793 bp in length
```

```
              *   162132   162157: gap of unknown length
              *   162158   162947: contig of 790 bp in length
              *   162948   162973: gap of unknown length
              *   162974   163754: contig of 781 bp in length
              *   163755   163780: gap of unknown length
              *   163781   164559: contig of 779 bp in length
              *   164560   164585: gap of unknown length
              *   164586   165359: contig of 774 bp in length
              *   165360   165385: gap of unknown length
              *   165386   166157: contig of 772 bp in length
              *   166158   166183: gap of unknown length
              *   166184   166945: contig of 762 bp in length
              *   166946   166971: gap of unknown length
              *   166972   167553: contig of 582 bp in length
              *   167554   167579: gap of unknown length
              *   167580   168336: contig of 757 bp in length
              *   168337   168362: gap of unknown length
              *   168363   169118: contig of 756 bp in length
              *   169119   169144: gap of unknown length
              *   169145   169891: contig of 747 bp in length
              *   169892   169917: gap of unknown length
              *   169918   170656: contig of 739 bp in length
              *   170657   170682: gap of unknown length
              *   170683   171415: contig of 733 bp in length
              *   171416   171441: gap of unknown length
              *   171442   172128: contig of 687 bp in length
              *   172129   172154: gap of unknown length
              *   172155   172866: contig of 712 bp in length
              *   172867   172892: gap of unknown length
              *   172893   173552: contig of 660 bp in length
              *   173553   173578: gap of unknown length
              *   173579   174193: contig of 615 bp in length
              *   174194   174219: gap of unknown length
              *   174220   174832: contig of 613 bp in length
              *   174833   174858: gap of unknown length
              *   174859   175162: contig of 304 bp in length.
  FEATURES             Location/Qualifiers
       source          1. .175162
                       /organism="Homo sapiens"
                       /db_xref="taxon:9606"
                       /chromosome="12p13.3"
  BASE COUNT    46973 a  41273 c  39564 g  45222 t   2130 others
  ORIGIN

    Query Match           6.0%;  Score 18;  DB 17;  Length 175162;
    Best Local Similarity  90.0%;  Pred. No. 2.92e+00;
    Matches    18;  Conservative    0;  Mismatches   2;  Indels   0;  Gaps   0;

  Db  29845 TTGCAGCGTCCTGTGGTTTG 29864
            ||||||| |||||||||| |
  Cp    266 TTGCAGCNTCCTGTGGTTNG 247
```



```
  RESULT   24
  LOCUS       HS191F10R      130 bp    DNA             PRI       18-OCT-1995
  DEFINITION  H.sapiens CpG island DNA genomic MseI fragment, clone 191f10,
              reverse read cpg191f10.rt1b.
  ACCESSION   Z57722
  NID         g1028953
  KEYWORDS    CpG island; genomic MseI fragment.
  SOURCE      human.
    ORGANISM  Homo sapiens
              Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
              Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE   1  (bases 1 to 130)
    AUTHORS   Dodsworth,S.J., Huckle,E., Wilkinson,P. and Micklem,G.
    TITLE     Direct Submission
    JOURNAL   Submitted (16-OCT-1995) The Sanger Centre, Hinxton, Cambridgeshire,
              CB10 1RQ, England. E-mail contact: humquery@sanger.ac.uk
  REFERENCE   2  (bases 1 to 130)
    AUTHORS   Cross,S.H., Charlton,J.A., Nan,X. and Bird,A.P.
    TITLE     Purification of CpG islands using a methylated DNA binding column
    JOURNAL   Nature Genet. 6 (3), 236-244 (1994)
    MEDLINE   94282070
  COMMENT     Vector: pGEM-5Zf(-)
              Clones are available from the UK MRC Human Genome Mapping Project
              Resource Centre, Hinxton, Cambridgeshire CB10 1RQ, UK. See URL:
              http://www.hgmp.mrc.ac.uk/ for details
              or contact: biohelp@hgmp.mrc.ac.uk.
  FEATURES             Location/Qualifiers
       source          1. .130
                       /organism="Homo sapiens"
                       /db_xref="taxon:9606"
                       /dev_stage="adult"
                       /sex="male"
                       /tissue_type="blood"
                       /clone_lib="CGI-1"
                       /clone="191f10"
  BASE COUNT       57 a     14 c     22 g     36 t      1 others
  ORIGIN

    Query Match           5.7%;  Score 17;  DB 25;  Length 130;
    Best Local Similarity  94.4%;  Pred. No. 1.61e+01;
    Matches    17;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

  Db      60 CTGTTCTAAAAGAGAAGT 77
             || ||||||||||||||||
  Qy     199 CTNTTCTAAAAGAGAAGT 216


  RESULT   25
  LOCUS       HS191F9F      130 bp    DNA             PRI       23-OCT-1995
```



```
  DEFINITION  H.sapiens CpG island DNA genomic MseI fragment, clone 191f9,
              forward read cpg191f9.ft1b.
  ACCESSION   Z65038
  NID         g1037860
  KEYWORDS    CpG island; genomic MseI fragment.
  SOURCE      human.
    ORGANISM  Homo sapiens
              Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
              Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE   1  (bases 1 to 130)
    AUTHORS   Dodsworth,S.J., Huckle,E., Wilkinson,P. and Micklem,G.
    TITLE     Direct Submission
    JOURNAL   Submitted (16-OCT-1995) The Sanger Centre, Hinxton, Cambridgeshire,
              CB10 1RQ, England. E-mail contact: humquery@sanger.ac.uk
  REFERENCE   2  (bases 1 to 130)
    AUTHORS   Cross,S.H., Charlton,J.A., Nan,X. and Bird,A.P.
    TITLE     Purification of CpG islands using a methylated DNA binding column
    JOURNAL   Nature Genet. 6 (3), 236-244 (1994)
    MEDLINE   94282070
  COMMENT     Vector: pGEM-5Zf(-)
              Clones are available from the UK MRC Human Genome Mapping Project
              Resource Centre, Hinxton, Cambridgeshire CB10 1RQ, UK. See URL:
              http://www.hgmp.mrc.ac.uk/ for details
              or contact: biohelp@hgmp.mrc.ac.uk.
  FEATURES             Location/Qualifiers
       source          1. .130
                       /organism="Homo sapiens"
                       /db_xref="taxon:9606"
                       /dev_stage="adult"
                       /sex="male"
                       /tissue_type="blood"
                       /clone_lib="CGI-1"
                       /clone="191f9"
  BASE COUNT       37 a     22 c     15 g     55 t      1 others
  ORIGIN

    Query Match           5.7%;  Score 17;  DB 25;  Length 130;
    Best Local Similarity  94.4%;  Pred. No. 1.61e+01;
    Matches    17;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

  Db      55 ACTTCTCTTTTAGAACAG 72
             |||||||||||||||| ||
  Cp    216 ACTTCTCTTTTAGAANAG 199


  RESULT   26
  LOCUS       AF081430      263 bp    DNA             PLN       07-AUG-1998
  DEFINITION  Oidiodendron chlamydosporicum 5.8S ribosomal RNA gene, partial
              sequence; internal transcribed spacer 2, complete sequence; and 28S
              ribosomal RNA gene, partial sequence.
  ACCESSION   AF081430
```



```
  NID         g3420859
  KEYWORDS    .
  SOURCE      Oidiodendron chlamydosporicum.
    ORGANISM  Oidiodendron chlamydosporicum
              Eukaryota; Fungi; Ascomycota; Euascomycetes; Plectomycetes;
              Onygenales; Myxotrichaceae; mitosporic Myxotrichaceae;
              Oidiodendron.
  REFERENCE   1  (bases 1 to 263)
    AUTHORS   Monreal,M.A., Berch,S. and Berbee,M.L.
    TITLE     Molecular diversity of ericoid mycorrhizal fungi
    JOURNAL   Unpublished
  REFERENCE   2  (bases 1 to 263)
    AUTHORS   Monreal,M.A., Berch,S. and Berbee,M.L.
    TITLE     Direct Submission
    JOURNAL   Submitted (30-JUL-1998) Botany, University of British Columbia,
              6270 University Blvd., Vancouver, B.C. V6T 1Z4, Canada
  FEATURES             Location/Qualifiers
       source          1. .263
                       /organism="Oidiodendron chlamydosporicum"
                       /strain="UAMH6520"
                       /db_xref="taxon:78143"
       rRNA            <1. .107
                       /product="5.8S ribosomal RNA"
       misc_RNA        108. .249
                       /note="ITS2"
                       /product="internal transcribed spacer 2"
       rRNA            250. .>263
                       /product="28S ribosomal RNA"
  BASE COUNT       51 a     78 c     74 g     60 t
  ORIGIN

    Query Match           5.7%;  Score 17;  DB 24;  Length 263;
    Best Local Similarity  89.5%;  Pred. No. 1.61e+01;
    Matches    17;  Conservative    0;  Mismatches   2;  Indels   0;  Gaps   0;

  Db     238 CCCCTTCTATGGTTGACCT 256
             |||||||| ||||||| ||
  Qy      36 CCCCTTCTNTGGTTGANCT 54


  RESULT   27
  LOCUS       AF081429      264 bp    DNA             PLN       07-AUG-1998
  DEFINITION  Oidiodendron scytaloides 5.8S ribosomal RNA gene, partial sequence;
              internal transcribed spacer 2, complete sequence; and 28S ribosomal
              RNA gene, partial sequence.
  ACCESSION   AF081429
  NID         g3420858
  KEYWORDS    .
  SOURCE      Oidiodendron scytaloides.
    ORGANISM  Oidiodendron scytaloides
              Eukaryota; Fungi; Ascomycota; Euascomycetes; Plectomycetes;
```

```
            Onygenales; Myxotrichaceae; mitosporic Myxotrichaceae;
            Oidiodendron.
REFERENCE   1  (bases 1 to 264)
  AUTHORS   Monreal,M.A., Berch,S. and Berbee,M.L.
  TITLE     Molecular diversity of ericoid mycorrhizal fungi
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 264)
  AUTHORS   Monreal,M.A., Berch,S. and Berbee,M.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JUL-1998) Botany, University of British Columbia,
            6270 University Blvd., Vancouver, B.C. V6T 1Z4, Canada
FEATURES             Location/Qualifiers
     source          1. .264
                     /organism="Oidiodendron scytaloides"
                     /strain="UAMH6521"
                     /db_xref="taxon:78152"
     rRNA            <1. .107
                     /product="5.8S ribosomal RNA"
     misc_RNA        108. .250
                     /note="ITS2"
                     /product="internal transcribed spacer 2"
     rRNA            251. .>264
                     /product="28S ribosomal RNA"
BASE COUNT       51 a     79 c     74 g     60 t
ORIGIN

  Query Match           5.7%;  Score 17;  DB 24;  Length 264;
  Best Local Similarity  89.5%;  Pred. No. 1.61e+01;
  Matches   17;  Conservative    0;  Mismatches   2;  Indels   0;  Gaps   0;

Db    239 CCCCTTCTATGGTTGACCT 257
          ||||||||| ||||||| ||
Qy     36 CCCCTTCTNTGGTTGANCT 54


RESULT   28
LOCUS       AF062804      513 bp    DNA             PLN       02-OCT-1998
DEFINITION  Oidiodendron scytaloides internal transcribed spacer 1, partial
            sequence; 5.8S ribosomal RNA gene and internal transcribed spacer
            2, complete sequence; and 28S ribosomal RNA gene, partial sequence.
ACCESSION   AF062804
NID         g3687762
KEYWORDS    .
SOURCE      Oidiodendron scytaloides.
  ORGANISM  Oidiodendron scytaloides
            Eukaryota; Fungi; Ascomycota; Euascomycetes; Plectomycetes;
            Onygenales; Myxotrichaceae; mitosporic Myxotrichaceae;
            Oidiodendron.
REFERENCE   1  (bases 1 to 513)
  AUTHORS   Hambleton,S., Egger,K.N. and Currah,R.S.
  TITLE     The genus Oidiodendron: species delimitation and phylogenetic
```



```
            relationships based on nuclear ribosomal DNA analysis
  JOURNAL   Mycologia (1998) In press
REFERENCE   2  (bases 1 to 513)
  AUTHORS   Hambleton,S., Egger,K.N. and Currah,R.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAY-1998) Biological Sciences, University of Alberta,
            CW-405 Biological Sciences Centre, Edmonton, Alberta T6G 2E9,
            Canada
FEATURES             Location/Qualifiers
     source          1. .513
                     /organism="Oidiodendron scytaloides"
                     /strain="UAMH 6521"
                     /db_xref="taxon:78152"
                     /note="authentic strain; isolated from forest soil,
                     Sweden"
     misc_RNA        <1. .160
                     /note="ITS1"
                     /product="internal transcribed spacer 1"
     rRNA            161. .318
                     /product="5.8S ribosomal RNA"
     misc_RNA        319. .458
                     /note="ITS2"
                     /product="internal transcribed spacer 2"
     rRNA            459. .>513
                     /product="28S ribosomal RNA"
BASE COUNT      102 a    152 c    140 g    117 t      2 others
ORIGIN

  Query Match           5.7%;  Score 17;  DB 24;  Length 513;
  Best Local Similarity  89.5%;  Pred. No. 1.61e+01;
  Matches   17;  Conservative    0;  Mismatches   2;  Indels   0;  Gaps   0;

Db    449 CCCCTTCTATGGTTGACCT 467
          ||||||||| ||||||| ||
Qy     36 CCCCTTCTNTGGTTGANCT 54


RESULT   29
LOCUS       AF062789      522 bp    DNA             PLN       02-OCT-1998
DEFINITION  Oidiodendron chlamydosporicum 18S ribosomal RNA gene, partial
            sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene
            and internal transcribed spacer 2, complete sequence; and 28S
            ribosomal RNA gene, partial sequence.
ACCESSION   AF062789
NID         g3687747
KEYWORDS    .
SOURCE      Oidiodendron chlamydosporicum.
  ORGANISM  Oidiodendron chlamydosporicum
            Eukaryota; Fungi; Ascomycota; Euascomycetes; Plectomycetes;
            Onygenales; Myxotrichaceae; mitosporic Myxotrichaceae;
            Oidiodendron.
```



```
REFERENCE   1  (bases 1 to 522)
  AUTHORS   Hambleton,S., Egger,K.N. and Currah,R.S.
  TITLE     The genus Oidiodendron: species delimitation and phylogenetic
            relationships based on nuclear ribosomal DNA analysis
  JOURNAL   Mycologia (1998) In press
REFERENCE   2  (bases 1 to 522)
  AUTHORS   Hambleton,S., Egger,K.N. and Currah,R.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAY-1998) Biological Sciences, University of Alberta,
            CW-405 Biological Sciences Centre, Edmonton, Alberta T6G 2E9,
            Canada
FEATURES             Location/Qualifiers
     source          1. .522
                     /organism="Oidiodendron chlamydosporicum"
                     /strain="UAMH 6520"
                     /db_xref="taxon:78143"
                     /note="ex-type culture; isolated from soil, ex boreal
                     forest, Saskatchewan, Canada"
     rRNA            <1. .5
                     /product="18S ribosomal RNA"
     misc_RNA        6. .170
                     /note="ITS1"
                     /product="internal transcribed spacer 1"
     rRNA            171. .327
                     /product="5.8S ribosomal RNA"
     misc_RNA        328. .467
                     /note="ITS2"
                     /product="internal transcribed spacer 2"
     rRNA            468. .>522
                     /product="28S ribosomal RNA"
BASE COUNT      106 a    151 c    142 g    123 t
ORIGIN

  Query Match           5.7%;  Score 17;  DB 24;  Length 522;
  Best Local Similarity  89.5%;  Pred. No. 1.61e+01;
  Matches   17;  Conservative    0;  Mismatches   2;  Indels   0;  Gaps   0;

Db    458 CCCCTTCTATGGTTGACCT 476
          ||||||||| ||||||| ||
Qy     36 CCCCTTCTNTGGTTGANCT 54


RESULT   30
LOCUS       CAHIS1GEN     1462 bp    DNA             PLN       14-MAR-1996
DEFINITION  C.albicans HIS1 gene for ATP phosphoribosyltransferase.
ACCESSION   X83871
NID         g860899
KEYWORDS    ATP phosphoribosyltransferase; his1 gene.
SOURCE      Candida albicans.
  ORGANISM  Candida albicans
            Eukaryotae; mitochondrial eukaryotes; Fungi; Ascomycota;
```



```
            Hemiascomycetes; Saccharomycetales; Candidaceae; Candida.
REFERENCE   1  (bases 1 to 1462)
  AUTHORS   Pla,J., Perez-Diaz,R.M., Navarro-Garcia,F., Sanchez,M. and
            Nombela,C.
  TITLE     Cloning of the Candida albicans HIS1 gene by direct complementation
            of a C. albicans histidine auxotroph using an improved double-ARS
            shuttle vector
  JOURNAL   Gene 165 (1), 115-120 (1995)
  MEDLINE   96084963
REFERENCE   2  (bases 1 to 1462)
  AUTHORS   Nombela,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JAN-1995) C. Nombela, Dept. de Microbiologica II,
            Fac. de Farmacia, Univ. Complutense de Madrid, E- 28040 Madrid,
            SPAIN
FEATURES             Location/Qualifiers
     source          1. .1462
                     /organism="Candida albicans"
                     /strain="1001 (ATCC 64385)"
                     /db_xref="taxon:5476"
                     /chromosome="5"
                     /clone="pRM21"
     gene            539. .1435
                     /gene="HIS1"
     CDS             539. .1435
                     /gene="HIS1"
                     /EC_number="2.4.2.17"
                     /codon_start=1
                     /transl_table=12
                     /product="ATP phosphoribosyltransferase"
                     /db_xref="PID:g860900"
                     /db_xref="SWISS-PROT:P46586"
                     /translation="MDLVNHLPDRLLFAVPKKGRLYEKCCNLLSGADIQFRRSNRLDI
                     ALSTNLPIALIPLPAADIPVFVGEGNCDLGITGLDQIKEAEQFDNIEDLLDLKFGSCK
                     LQIQVPADGEYEKPEQLVGKKIVSSFTKLSTDYFKQLSDKPTNIRYVGGSVEASCALG
                     VADAIVDLVESGETMKAAGLKAIETILETSAHLISSKKSKFPEMVNIIVQRLQGVLAA
                     QEYVLCNYNAPKSIQAKCLTITPGRRAATVSTLDKHSDDEEDWVAISSMVNRKEIGNV
                     MDELKKAGATDILVLEISNCRV"
BASE COUNT      469 a    246 c    303 g    444 t
ORIGIN

  Query Match           5.7%;  Score 17;  DB 23;  Length 1462;
  Best Local Similarity 100.0%;  Pred. No. 1.61e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    138 AGTTTTTTATAAATGTA 154
          |||||||||||||||||
Qy     94 AGTTTTTTATAAATGTA 110


RESULT   31
```

```
LOCUS       AC002182     2881 bp    DNA             PRI       03-JUN-1997
DEFINITION  Homo sapiens (subclone 2_f6 from BAC H111) DNA sequence, complete
            sequence.
ACCESSION   AC002182
NID         g2168264
KEYWORDS    HTG.
SOURCE      Homo sapiens (Subclones in pOT2 from BAC clone H111) DNA.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 2881)
  AUTHORS   Martin,C.H., Arcaina,T.T., Bondoc,M.M., Chiang,A., Critz,P.A.,
            Davis,C.A., Doyle,C.M., Ericsson,C.L., Farfan,D.E., Gunning,K.M.,
            Houston,K.A., Jaklevic,M.A., Kadner,K.E., Kim,K., Kim,S.F.,
            Ko,C.L., Lewis,K.D., Li,M., Lindquist,K.J., Lomotan,M.A.,
            Lustre,V.M., Machrus,M.U., Mayeda,C.A., Miguel,T.M., Miller,C.A.,
            Mok,M.S., Pacleb,J.M., Patel,S.G., Santos,R.F., Subramanian,S.,
            Wan,K.H., Whitelaw,K.R., Yee,A., Yeh,R.T., Yu,C. and Palazzolo,M.J.
  TITLE     Sequencing of
  JOURNAL   Unpublished (1997)
REFERENCE   2  (bases 1 to 2881)
  AUTHORS   Martin,C.H., Alcivare,D.A., Arcaina,T.T., Bondoc,M.M., Chiang,A.,
            Critz,P.A., Davis,C.A., Doyle,C.M., Ericsson,C.L., Farfan,D.E.,
            Gunning,K.M., Houston,K.A., Hummasti,S.R., Jaklevic,M.A.,
            Kadner,K.E., Karra,K., Kim,K., Kim,S.F., Ko,C.L., Lewis,K.D.,
            Li,M., Lindquist,K.J., Lomotan,M.A., Lustre,V.M., Machrus,M.U.,
            Mayeda,C.A., Mazda,P., Miguel,T.M., Miller,C.A., Mok,M.S.,
            Pacleb,J.M., Patel,S.G., Salva,A., Santos,R.F., Snir,E.,
            Subramanian,S., Wan,K.H., Whitelaw,K.R., Yee,A., Yeh,R.T., Yu,C.,
            Zieran,L.L. and Palazzolo,M.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUN-1997)
COMMENT     Sequence submitted by:
            Human Genome Center
            Lawrence Berkeley National Laboratory, MS 74-157
            Berkeley, CA 94720
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www-hgc.lbl.gov/sequence-archive.html) or
            send email to human@genome.lbl.gov.
FEATURES              Location/Qualifiers
     source           1. .2881
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome=""
                      /clone="BAC H111 (bac187F3)"
                      /sub_clone="2_f6"
BASE COUNT     1072 a    502 c    444 g    863 t
ORIGIN
```



```
  Query Match           5.7%;  Score 17;  DB 26;  Length 2881;
  Best Local Similarity 100.0%;  Pred. No. 1.61e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2677 TTTATAAATGTATAAAT 2693
           |||||||||||||||||
Qy     99 TTTATAAATGTATAAAT 115


RESULT   32
LOCUS       XLU75996     2900 bp    mRNA            VRT       20-DEC-1996
DEFINITION  Xenopus laevis eomesodermin mRNA, complete cds.
ACCESSION   U75996
NID         g1743868
KEYWORDS    .
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea;
            Pipidae; Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 2900)
  AUTHORS   Ryan,K., Garrett,N., Mitchell,A. and Gurdon,J.B.
  TITLE     Eomesodermin, a key early gene in Xenopus mesoderm differentiation
  JOURNAL   Cell 87 (6), 989-1000 (1996)
  MEDLINE   97133207
REFERENCE   2  (bases 1 to 2900)
  AUTHORS   Ryan,K., Garrett,N. and Gurdon,J.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-OCT-1996) Wellcome/CRC Institute and Dept. of
            Zoology, University of Cambridge, Tennis Court Road, Cambridge CB2
            1QR, UK
FEATURES              Location/Qualifiers
     source           1. .2900
                      /organism="Xenopus laevis"
                      /db_xref="taxon:8355"
     CDS              145. .2223
                      /note="T-box transcription factor"
                      /codon_start=1
                      /product="eomesodermin"
                      /db_xref="PID:g1743869"
                      /translation="MVPGAWHSLLPTTSSASEKEENRSQRMQLGEQLLNSSTPNLPHT
                      FYPLTGDPSSAVHSPSLEFSVGHKGQQHKKYSSGSSRGLQLDSPREAGSSSTMLSETG
                      EGFSVAKTLPDNVRKGSPSAEEELNTAVPTSAPRYLDGSLQAASERYYLQPGQQLQQ
                      TTTELGSPCSIFPPYAPPQHSAVYPAGCAARYPPYGSMLPPAGFSPPVCPSRPQYSSGY
                      QYSQAPGTWYSPYPPAGTCSGLSALCLPGGGAGVRAQVFLCNRPLWLKFHRHQTEMII
                      TKQGRRMFPFLSFNITGLNPTAHYNVFVEVVLADPNHWRPQGGKWVTCGKADNNMQGN
                      KVYVHPESPNTGAHWMRQEISFGKLKLTNNKGANNNSTQMIVLQSLHKYQPRLHIVEV
                      SEDGVEDLNDSAKNQTFTPPENQFIAVTAYQNTDITQLKIDHNPFAKGFRDNYDSMYT
                      ASESDRLTPSPADSPRSHQIVPGTRYSVQPFFQDQFVWNLPPARYYSGERTVPQANGL
                      LSPQTNEEVANVPPQRWFVTPVQQAAAANKLDMGAYETDYSSGSLLTYGIKSLPIQTSH
                      PMAYYPDAAFASMAGWGSRGSTYQRHMTTSLPWSSRSSPSGFSEDLLPKDKVKEEMSS
```



```
                      SWVETPPSIKSLDSNDSGVYTGACKRRRLSPSTSSNENSPPIKCEDIGTEDYKDATKG
                      LGYTSPTSSS"
BASE COUNT     795 a    756 c    650 g    699 t
ORIGIN

  Query Match           5.7%;  Score 17;  DB 20;  Length 2900;
  Best Local Similarity 100.0%;  Pred. No. 1.61e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2449 TTTATAAATGTATAAAT 2465
           |||||||||||||||||
Qy     99 TTTATAAATGTATAAAT 115


RESULT   33
LOCUS       PPB2LR       3195 bp    DNA             VRL       15-SEP-1989
DEFINITION  Bovine papillomavirus type 2 late region DNA encoding structural
            proteins L1 and L2, complete cds.
ACCESSION   M24326
NID         g332997
KEYWORDS    structural protein.
SOURCE      Bovine papillomavirus type 2 DNA.
  ORGANISM  Bovine papillomavirus
            Viruses; dsDNA viruses, no RNA stage; Papovaviridae;
            Papillomavirus.
REFERENCE   1  (bases 1 to 3195)
  AUTHORS   Potter,H.L.Jr. and Meinke,W.J.
  TITLE     Nucleotide sequence of bovine papillomavirus type 2 late region
  JOURNAL   J. Gen. Virol. 66, 187-193 (1985)
  MEDLINE   85107102
FEATURES              Location/Qualifiers
     source           1. .3195
                      /organism="Bovine papillomavirus"
                      /db_xref="taxon:10571"
     CDS              109. .1512
                      /note="structural protein L2"
                      /codon_start=1
                      /db_xref="PID:g332998"
                      /translation="MSARKRVKRAHVYDLYRTCKQAGTCPPDVIPKVEGDTIADKILK
                      LGGLAIYLGGLGIGTWSTGRVAAGGSPRYVPLRTSGSTTSLASVGSRAGAATGTRSSI
                      TGIPLDTLETIGALRPGAYEDTVLPEAPAIVTPDAVPADTGIDGLSIGTDSSTETLIT
                      LLEPEGPEDVAVLELQPLDHANWQVSNAVHQGSAYHAPLQLQSSIAETSGLENIPVGG
                      AGLGDTGGENIELTFPGSPRTSTPRNLPQTARGILNWFSKRYTYTQIPTEDPDVFSSQT
                      FSNPVTDPEPAVLKGPSGRVGLSQVYRPDYIETRGGGQVGPQLHVRYSLSTITEDVEA
                      IPIAVDEDTQGLAFPLHEPGDPEEIELDDLGEEHALLPKSSTAPIGSGVRRALIPG
                      QGFSATRPTGVVTYGSPDMYPASPVGPDSTSPSLVIDDNTTPPIIIDGHTVDLYSNH
                      YSLHPSLLRKRKKRKRHA"
     CDS              1525. .3018
                      /note="structural protein L1"
                      /codon_start=1
                      /db_xref="PID:g332999"
```



```
                      /translation="MALWQQGQKLYLPPTPVSKVLCSETYVQRKSIFYHAETERLLTV
                      GHPYTQVTVGDKTVPKVSANQFRVFKIQLPDPNQFALPDRTVHNPSKERLVWAVIGVQ
                      VSRGQPLGGTVTGHPTFNALLDAENVNRKVTAQTTDDRKQTGLDAKQQQKLLLGCTPA
                      EGEYWTTARPCVTDRIENGAGCPPLELKNKHIEDGDMMEIGFGAADFKTLNASKSDLPL
                      DIQNEICLYPDYLEMAEDAAGHSMFFPARKEQVTVRHIWTRGGSEKEAPSKDFYLKNG
                      RGEETLKIPSVHPGSPSGSLVSTDNQIFNRPYWLFRAQGMNNGIAWDNLLLFLTVGDNT
                      RGTNLSISVAADGNALSEYDTGKFNLYHRHMEEYKLAFILELCSVEITAQTVSHLQGL
                      MPSVLQNWEIGVQPPASSILEDTYRYIESPATKCASNVIPPKEDPYAGLRFWSIDLKE
                      KLSLDLDQFPLGRRFLAQQGAGCSTVRKRAVATRNSSKPAKRKKIKA"
BASE COUNT     950 a    695 c    738 g    812 t
ORIGIN      3009 bp upstream of unique HindIII site.

  Query Match           5.7%;  Score 17;  DB 33;  Length 3195;
  Best Local Similarity 100.0%;  Pred. No. 1.61e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   3020 TTTTTATAAATGTATAA 3036
           |||||||||||||||||
Qy     97 TTTTTATAAATGTATAA 113


RESULT   34
LOCUS       PABPV2L      3195 bp    DNA             VRL       01-OCT-1996
DEFINITION  Bovine papillomavirus type 2 (BPV-2) late region DNA sequence.
ACCESSION   X01760 M24326
NID         g60859
KEYWORDS    oncogene.
SOURCE      Bovine papillomavirus.
  ORGANISM  Bovine papillomavirus
            Viruses; dsDNA viruses, no RNA stage; Papovaviridae;
            Papillomavirus.
REFERENCE   1  (bases 1 to 3195)
  AUTHORS   Potter,H.L. Jr. and Meinke,W.J.
  TITLE     Nucleotide sequence of bovine papillomavirus type 2 late region
  JOURNAL   J. Gen. Virol. 66 (Pt 1), 187-193 (1985)
  MEDLINE   85107102
COMMENT     Data kindly reviewed (08-JAN-1986) by W. Meinke.
FEATURES              Location/Qualifiers
     source           1. .3195
                      /organism="Bovine papillomavirus"
                      /strain="type 2"
                      /db_xref="taxon:10571"
     misc_feature     <1. .108
                      /note="5' untranslated region"
     promoter         88. .91
                      /note="pot. TATA-box"
     CDS              109. .1512
                      /note="open reading frame L2"
                      /codon_start=1
                      /db_xref="PID:g50860"
                      /db_xref="SWISS-PROT:P06457"
```

```
                     /translation="MSARKRVKRANVYDLYRTCKQAGTCPPDVIPKVEGDTIADKILK
                     LGGLAIYILGGLGIGTWSTGRVAAGGSPRYVPLRTSGSTTSLASVGSRAGAATGTRSSI
                     TGIPLDTLETIGALRPGAYEDTVLPEAPAIVTPDAVPADTGIDGLSIGTDSSTETLIT
                     LLEPEGPEDVAVLELQPLDHANWQVSNAVHQGSAYHAPLQLQSSIAETSGLENIFVGG
                     AGLGDTGGENIELTFFGSPRTSTPRNLPQTARGILNWFSKRYYTQIPTEDPDVPSSQT
                     FSNPVYDPEPAVLKGPSGRVGLSQVYRPDYIETRGGGQVGPQLHVRYSLSTITEDVEA
                     IPIAVDEDTQGLAFLPLHEEPGDFEEIELDDLGEEHALLPKSSTAPIGSGVRRALIPG
                     QGFSATRPTGVVTYGSPDMYPASPVGPDSTSPSLVIDDNTTTPIIIIDGHTVDLYSNN
                     YSLHPSLLRKRKKRKHA"
     CDS             1525. .3018
                     /note="open reading frame L1"
                     /codon_start=1
                     /db_xref="PID:g60861"
                     /db_xref="SWISS-PROT:P06458"
                     /translation="MALWQQGQKLYLPPTPVSKVLCSETTVQRKSIFYHAETERLLTV
                     GHPYYQVTVGDKTVPKVSANQFRVFKIQLPDPNQFALPDRTVHNPSKERLVWAVIGVQ
                     VSRGQPLGGTVTGHPTFNALLDAENVNRKVTAQTTDDRKQTGLDAKQQQILLLGCTPA
                     EGEYWTTARPCVTDRIENGACPPLELKNKHIEDGDMMEIGFGAADFKTLNASKSDLPL
                     DIQNEICLYPDYLKMAEDAAGNSMFFFARKEQVYVRHIWTRGGSEKEAPSKDFYLKNG
                     RGEETLKIPSVHPGSPSGSLVSTDNQIFNRPYWLFRAQGGMNGIAWDNLLFLTVGDNT
                     RGTNLSISVAADGNALSEYDTGKFNLYHRHMEEYKLAFILELCSVEITAQTVSHLQGL
                     MPSVLQNWEIGVQPPASSILEDTKRYIESPATKCASNVIPPKEDFAGLRFWSIDLKE
                     KLSLDLDQFPLGRRPLAQQGAGCSTVRKRAVATRNSSKPAKRKKIKA"
     misc_feature    3035. .3040
                     /note="pot. polyadenylation signal"
     misc_feature    3071. .3076
                     /note="pot. polyadenylation signal"
BASE COUNT      949 a    694 c    737 g    815 t
ORIGIN

  Query Match           5.7%;  Score 17;  DB 33;  Length 3195;
  Best Local Similarity 100.0%;  Pred. No. 1.61e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  3020 TTTTTATAAATGTATAA 3036
         |||||||||||||||||
Qy    97 TTTTTATAAATGTATAA 113


RESULT   35
LOCUS       HSAC001477   3547 bp    DNA             PRI       18-APR-1997
DEFINITION  Homo sapiens (subclone 2_a4 from BAC H111) DNA sequence, complete
            sequence.
ACCESSION   AC001477
NID         g1944797
KEYWORDS    HTG.
SOURCE      Homo sapiens (Subclones in pOT2 from BAC clone H111) DNA.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
```



```
REFERENCE   1  (bases 1 to 3547)
  AUTHORS   Martin,C.H., Arcaina,T.T., Bondoc,M.M., Chiang,A., Critz,P.A.,
            Davis,C.A., Doyle,C.M., Ericsson,C.L., Farfan,D.E., Gunning,K.M.,
            Houston,K.A., Jaklevic,M.A., Kadner,K.E., Kim,K., Kim,S.F.,
            Ko,C.L., Lewis,K.D., Li,M., Lindquist,K.J., Lomotan,M.A.,
            Lustre,V.M., Machrus,M.U., Mayeda,C.A., Miguel,T.M., Miller,C.A.,
            Mok,M.S., Pacleb,J.M., Patel,S.G., Santos,R.F., Subramanian,S.,
            Wan,K.H., Whitelaw,K.R., Yee,A., Yeh,R.T., Yu,C. and Palazzolo,M.J.
  TITLE     Sequencing of
  JOURNAL   Unpublished (1997)
REFERENCE   2  (bases 1 to 3547)
  AUTHORS   Martin,C.H., Arcaina,T.T., Bondoc,M.M., Chiang,A., Critz,P.A.,
            Davis,C.A., Doyle,C.M., Ericsson,C.L., Farfan,D.E., Gunning,K.M.,
            Houston,K.A., Jaklevic,M.A., Kadner,K.E., Kim,K., Kim,S.F.,
            Ko,C.L., Lewis,K.D., Li,M., Lindquist,K.J., Lomotan,M.A.,
            Lustre,V.M., Machrus,M.U., Mayeda,C.A., Miguel,T.M., Miller,C.A.,
            Mok,M.S., Pacleb,J.M., Patel,S.G., Santos,R.F., Subramanian,S.,
            Wan,K.H., Whitelaw,K.R., Yee,A., Yeh,R.T., Yu,C. and Palazzolo,M.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-APR-1997)
COMMENT     Sequence submitted by:
            Human Genome Center
            Lawrence Berkeley National Laboratory, MS 74-157
            Berkeley, CA 94720
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Website (http://www-hgc.lbl.gov/sequence-archive.html) or
            send email to human@genome.lbl.gov.
FEATURES              Location/Qualifiers
     source           1. .3547
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome=""
                      /clone="BAC H111 (bac187F3)"
                      /sub_clone="2_a4"
BASE COUNT     1184 a    657 c    576 g   1130 t
ORIGIN

  Query Match           5.7%;  Score 17;  DB 26;  Length 3547;
  Best Local Similarity 100.0%;  Pred. No. 1.61e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   734 TTTATAAATGTATAAAT 750
         |||||||||||||||||
Qy    99 TTTATAAATGTATAAAT 115


RESULT   36
LOCUS       AF029395     5127 bp    mRNA            INV       16-JAN-1998
DEFINITION  Drosophila melanogaster Cdc42-binding kinase Genghis Khan (gek)
            mRNA, complete cds.
```



```
ACCESSION   AF029395
NID         g2772929
KEYWORDS    .
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryotae; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea;
            Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 5127)
  AUTHORS   Luo,L., Lee,T., Tsai,L., Tang,G., Jan,L.Y. and Jan,Y.N.
  TITLE     Genghis Khan (Gek) as a putative effector for Drosophila Cdc42 and
            regulator of actin polymerization
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 94 (24), 12963-12968 (1997)
  MEDLINE   98058752
REFERENCE   2  (bases 1 to 5127)
  AUTHORS   Luo,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-OCT-1997) Biological Sciences, Stanford University,
            385 Serra Mall, Stanford, CA 94305, USA
FEATURES              Location/Qualifiers
     source           1. .5127
                      /organism="Drosophila melanogaster"
                      /db_xref="taxon:7227"
                      /chromosome="II"
                      /map="60B"
                      /dev_stage="4-8 hour embryo"
                      /clone_lib="4-8 hour embryonic cDNA library made by N.
                      Brown (J. Mol. Biol. 203: 425-427, 1988)"
     gene             1. .5127
                      /note="genghis khan"
                      /gene="gek"
     CDS              189. .5030
                      /gene="gek"
                      /note="protein kinase with additional domains including
                      coiled coil, Cys-rich domain, plekstrin homology domain,
                      and Cdc42/Rac interactive binding domain; putative Cdc42
                      effector"
                      /codon_start=1
                      /product="Genghis Khan"
                      /db_xref="PID:g2772930"
                      /translation="MYSKRHHKRSKWAAEFDEDTEGHQFSLDYLLDTFIVLYDECSNS
                      SLRREKGVSDFLKLSKPFVHIVRKLRLSRDDFDILKIIGRGAPGEVCVVQMISTEKVY
                      AMKILNKWEMLKRAETACFREERDVLVFGDRQWITNLHYAFQDNINLYLVMDYYCGGD
                      LLTLLSKFEDKLPEDMAKFYITEMILAINSIHQIRYVHRDIKPDNVLLDKRGHVRLAD
                      FGSCLRLDKDGTVQSNVAVGTPDYISPEILRAMEDGKGRYGTECDWWSLGVCMYEMLY
                      GETPFYAESLVETYGKIMNHQNCFNLPSQETLNYKVSETSQDLLCKLICIPENRLGQN
                      GIQDFMDHPWFVGIDWKNIRQGPAPYVPEVSSPTDTSNFDVDDNDVRLTDSIPPSANP
                      AFSGFHLPFIGFTFSLTSSSTLDSKKNQSSGFGDDTLDTISSPQLAILPSNNSETPVD
                      SVQLKALNDQLAALKQEKAELSKQHNEVFERLKTQDSELQDAISQRNIAMMEYSEVTE
                      KLSELRNQKQKLSRQVRDKEEELDGAMQKNDSLRNELRKSDKTRRELELHIEDAVIEA
                      AKEKKLREHAEDCCRQLQMELRKGSSSVETTMPLSISSEMSSYEIERLELQFSEKLSH
```



```
                      QQTRHNMELEALREQFSELENANLALTKELQQTQERLKYTQMESITDSAETLLELKKQ
                      HDLEKSSWFEEKQRLSSEVNLKSKSLKELQAEDDEIFKELRMKREAITLWERQMAEII
                      QWVSDEKDARGYLQALATKMTEELEYLKHVGTFNNNGVDNKNWRNRRSQKLDKMELLN
                      LQSALQREIQAKNMISDELSQTRSDLISTQKEVRDYKKRYDSILHDFQKKETELRDFE
                      KGGLEYSESPLNKSTHHGLSSAFFRDMSKNSEIIDSAESPGNESGDNFTPHFPQSGNS
                      GMLFNYEPKYAGKNNKDHSSMKEASVSDLSREESDQLVKESQKKVPGNTAIHQFLVRT
                      PSSPTKCNHCTSLMVCLTRQGVVCEICGPACHTICQRVVPTTCPVPMDQTKRPLGIDP
                      TRGIGTAYEGTVKVPKSGVIKRGWIRQFVVVCDFKLFLYDISPDRCALPSVSVSQVLD
                      MRDPEFSVGSVRESDVIHAAKKDVPCIFKITALIDGGLSLNTLMLADNESEKSKWVI
                      ALGELHRILKRNSLPNTAIFKVNEILDNTLSLIRNALCSVIIYPNQILLGTEDGLFYI
                      NLDQYEIARIGESKKILQLWYIEEEQILVILCGKQRNLRLLPIRALEASDVEWIKVVE
                      SKNCISACTGIIRGFPNIVYSFIIALKRPNNHTQIVVYEINRTRTRHQKTCEFTIGYM
                      AQHLQILSDMRLVVAHQSGFTAYFLRGEATAMSLVHPENQLCAFLNYSGVDAVRVIEI
                      LCPSGGNFGEYLLVFQTLAIYVDLQGRKSRDREIMYPAFPTYITFCDGHLLVFSDTHL
                      DIFNTQTAEWVQSIGLKQSLPLNNLGNVVLSSVNDTPLIVYLSNIHTKGLLQYRDGNR
                      KGLPSIKRRFSIREINKTIKSDRRSKMISAPTNFNHISHMGPGDGIQNQRLLDLPTTL
                      ETADQACSPIIHSLSCIPQSRKSNFLEQVDANSDDYGNDNIISRTPSPMASSFMDGLS
                      NND"
BASE COUNT     1451 a   1233 c   1304 g   1139 t
ORIGIN

  Query Match           5.7%;  Score 17;  DB 18;  Length 5127;
  Best Local Similarity  94.4%;  Pred. No. 1.61e+01;
  Matches   17;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db   150 AACCACAGGAAGCTGCAA 167
         |||||||||| |||||||
Qy   249 AACCACAGGANGCTGCAA 266


RESULT   37
LOCUS       PPB2CG       7937 bp    DNA             VRL       17-MAY-1995
DEFINITION  Bovine papillomavirus type 2, complete genome.
ACCESSION   M20219 M19551
NID         g332996
KEYWORDS    complete genome.
SOURCE      Bovine papillomavirus type 2 DNA.
  ORGANISM  Bovine papillomavirus type 2
            Viruses; dsDNA viruses, no RNA stage; Papovaviridae;
            Papillomavirus.
REFERENCE   1  (bases 1 to 7937)
  AUTHORS   Groff,D.E., Mitra,R. and Lancaster,W.D.
  JOURNAL   Unpublished (1986)
COMMENT     Draft entry and printed copy of sequence for [1] kindly provided by
            D.E.Groff and W.D.Lancaster, 30-MAY-1988).
FEATURES              Location/Qualifiers
     source           1. .7937
                      /organism="Bovine papillomavirus type 2"
                      /db_xref="taxon:10560"
     gene             91. .504
                      /gene="E6"
```

```
     CDS             91. .504
                     /gene="E6"
                     /note="Open reading frame begins at bp 85.; putative"
                     /codon_start=1
                     /product="transforming protein"
                     /db_xref="PID:g808778"
                     /translation="MDLQSPSRGNPPSGLACVWCREPLTEVDAFRCMIKDFHVVYRDG
                     VKFGACTTCLENCLDKERRLWKGVPVTGEEAQLLHGKSLDRLCIRCCYCGGKLTKNEK
                     QRHVLYNEPFCKTRSNIIRGRCYDCCRHGSRSNYP"
     gene            479. .862
                     /gene="E7"
     CDS             479. .862
                     /gene="E7"
                     /note="Open reading frame begins at bp 452.; putative"
                     /codon_start=1
                     /product="transforming protein"
                     /db_xref="PID:g808779"
                     /translation="MVQGPTTHRNLDDSPAGPLLILSPCAGTPTRVPAAPDAPDFRLP
                     CHFGRPTRKRGPSTPPLSSPGKVCATGPRRVYSVTVCCGHCGKDLTPAVKTGSTTLLG
                     FEHLLNSDLDLLCPRCESRERHGKR"
     gene            849. .2663
                     /gene="E1"
     CDS             849. .2663
                     /gene="E1"
                     /note="Open reading frame begins at bp 801.; putative"
                     /codon_start=1
                     /product="replication protein"
                     /db_xref="PID:g808780"
                     /translation="MANDKGSNWDSALGCSYLLTEAECESDKENEEPGAGVELSVESD
                     RYDSQDEDFLDNASVFQGNHLEVFQALEKKAGEEQLLNLKRKVLGSSENSSGSEASET
                     PAKRQKAGAKRRLFSENEANRVLTPLQVQGGEWRQGFNEDQAISHRLLQLVKSKNATV
                     FKLGLFKSLFLCSFHDLTRLFKNDKTTNQQWVLAVFGIAEVFFEASLELLKKQCSFVQ
                     NQKRSHEGGTCAVYLLCFNTARSRETVRNLMARWLNVREECLLMQPPRIRGLSAALFW
                     FKSSLSPATLHGALPEWIRAQYTLHDSLATEKFDFGTWVQWAYDHKYAFESKIAYEY
                     ALAAGSDSNARAFLATNSQAKHVKDCATMVRHYLRAETQALSMPAYIKTRCKLATGDG
                     SWKSILTFFNYQNIELITPINALKLWLNGIPKKNCLAFIGPPKTGKSMLCNSLIHFLG
                     GSVLSFANHKSHFWLASLADARAALVDDATHACWRTFDTYLRNALDGTPVSIDRKHKA
                     AVQIKAPPLLVTSNIDVQAEERYLYLHSRVQTFRFEQPCTDESGEQPPTIITDADWKSF
                     FVRLWGRLDLVDEEEDSEEDGDSMRTFTCSARNTNAVD"
     gene            1204. .1428
                     /gene="E8"
     CDS             1204. .1428
                     /gene="E8"
                     /note="Open reading frame begins at bp 1099.; putative"
                     /codon_start=1
                     /db_xref="PID:g808781"
                     /translation="MKLTVPLRPCRYREGNGGKDLMRTRQLVIDFCSSSNLKMLQFLS
                     WGFLSLYFCAAPTILLDCLRMIKLLISSGC"
     gene            2605. .3840
                     /gene="E25"
     CDS             2605. .3840
```



```
                     /gene="E25"
                     /note="Open reading frame begins at bp 2572.; putative"
                     /codon_start=1
                     /product="regulatory protein"
                     /db_xref="PID:g808782"
                     /translation="METACERLHVAQETQMQLIEKASDKLQDHILYWTAVRTENTLLY
                     AARRKGVTVLGHCRVPHSVVCQERAKQAIEMQLSLQELSKTEPGNEPWCLLDTSWDRY
                     MSEPKRCFKKGARVVEVEFDGNASNTNWYTVYSKLYMRTEDGWQLAKAGADGTGLYYC
                     TMAGAGRIYYSRFGEEAARFSTTGHYSVRDQDRVYAGVSSTSSDFRDRPDGVSASEGP
                     EGDPAGKEAEPAQPVSSLLGSPACVPIRAGLGWVRDGPRPHPYHFPAGSGGSLLRSAS
                     TPVQGPVPVDLAPRQEEEENQSPDSTEEEPVTVPRHTSDADGFHLLKAGQSCFALISG
                     SANQVKCYRFRVKKNHRHRYENCTTTSFTVADNGAERQGQAQILITPGSPGQRQDFLK
                     HVPLPPGMNISGFTASLDF"
     gene            3188. .3526
                     /gene="E4"
     CDS             3188. .3526
                     /gene="E4"
                     /note="Open reading frame begins at bp 3170.; putative"
                     /codon_start=1
                     /db_xref="PID:g808783"
                     /translation="MLVSHPPLLILEIAQTEFPHQKDLKETLQEKKPSQPSLSLLCSA
                     PPPAYPSEQASVGYGTVLARTPTIFLQARGALFSALPPPRCRARIRWTWHQGRKKKKI
                     NRPTQQRRNL"
     gene            3261. .3548
                     /gene="E3"
     CDS             <3261. .3548
                     /gene="E3"
                     /note="Open reading frame begins at bp 3261. No methionine
                     present in ORF.; putative"
                     /codon_start=1
                     /db_xref="PID:g808784"
                     /translation="RRPCRKRSRASPACLFFARLPRLRTHQSRPRLGTGRSSPAPLPF
                     SCRLGGLSSPLCLHPGAGPGTGGLGTKAGRRRRKSIARLNRGGTCDCTKAHL"
     gene            3882. .4016
                     /gene="E5"
     CDS             3882. .4016
                     /gene="E5"
                     /note="Open reading frame begins at bp 3714.; putative"
                     /codon_start=1
                     /product="transforming protein"
                     /db_xref="PID:g808785"
                     /translation="MPNLWFLLFLGLVAAMQLLLLLFLLLFFLVYWDHPECSCTGLPF
                     "
     gene            4185. .5588
                     /gene="L2"
     CDS             4185. .5588
                     /gene="L2"
                     /note="Open reading frame begins at bp 4101.; putative"
                     /codon_start=1
                     /product="minor capsid protein"
                     /db_xref="PID:g808786"
```



```
                     /translation="MSARRRVKRANVYDLYRTCKQAGTCPPDVIPKVEGDTIADKILK
                     LGGLAIYLGGLGIGTWSTGRVAAGGSPRTVPLRTSGSTTSLASVGSRAGAATGTRSSI
                     TGIPLDTLETIGALRPGAYEDTVLPEAPAIVTPDAVPADTGIDGLSIGTDSSTETLIT
                     LLEPEGPEDVAVLELQPLDHANWQVSNAVHQGSATHAPLQLQSSIAETSGLENIFVGG
                     AGLDTGGNIELTFTGSPRTSTPRNLPQTARGILLNWFSKRYTYQIPTEDDPVFSSQT
                     PSNPLFDPEPAVLKGPSGRVGLSQVTRPDPIETRGGGQVGPQLHVRYSLSTITEDVEA
                     IPIAVDEDTQGLAFIPLHEEPGDFEEIELDDLGEEHALLPKSTTAPIGSGVRRALIPG
                     QGFSATRPTGVVTYGSPDMYPASPVGPDSTSPSLVIDDNTTTPIIIIDGHTVDLYSNN
                     YSLHPSLLRKRKKRKHA"
     gene            5601. .7094
                     /gene="L1"
     CDS             5601. .7094
                     /gene="L1"
                     /note="Open reading frame begins at bp 5589.; putative"
                     /codon_start=1
                     /product="major capsid protein"
                     /db_xref="PID:g808787"
                     /translation="MALWQQGQKLYLPPTPVSKVLCSETTVQRKSIFYHAETERLLTV
                     GHPYTQVTVGDKTVPKVSANQFRVFKIQLPDPNQFALPDRTVHNPSKERLVWAVIGVQ
                     VSRGQPLGGTVTGHPTFNALLDAENVNRKVTAQTTDDRKQTGLDAKQQQILLLGCTPA
                     EGEYWTTARPCVTDRLENGACPPLELKNKHIEDGDMKEIGFGAADFKTLNASKSDLPL
                     DIQNEICLYPDYLKMAEDAAGNSMFFFARKEQVYVRHIWTRGGSEKEAPSKDFYLKNG
                     RGEETLKIPSVHFGSPSGSLVSTDNQIFNRPYWLFRAQGMNNGIAWNNLLFLTVGDNT
                     RGTNLSISVAADGNALSEYDTGKFNLYHRHMEEYKLAFILELCSVEITAQTLSHLQGL
                     MPSVLQNWEIGVQPPASSILEDTYRYIESPATKCASNVIPPKEDPYAGLKFWSIDLKE
                     KLSLDLDQFPLGRRFLAQQGAGCSTVRRARAVATRNSSKPARRKKIKA"
BASE COUNT     2252 a   1711 c   1923 g   2051 t
ORIGIN

  Query Match           5.7%;  Score 17;  DB 33;  Length 7937;
  Best Local Similarity 100.0%;  Pred. No. 1.61e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    7096 TTTTTATAAATGTATAA 7112
           |||||||||||||||||
Qy      97 TTTTTATAAATGTATAA 113


RESULT   38
LOCUS       AF022713     8595 bp    DNA             INV       13-JUN-1998
DEFINITION  Drosophila melanogaster putative inorganic phosphate cotransporter
            (Picot) gene, partial cds; putative sodium channel (Nach) and
            putative amylase-related protein (Amyrel) genes, complete cds; and
            putative serine-enriched protein (gprs) gene, partial cds.
ACCESSION   AF022713
NID         g3025664
KEYWORDS    .
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea;
```



```
            Drosophilidae; Drosophila.
  REFERENCE 1  (bases 1 to 8595)
    AUTHORS   Da Lage,J.L., Renard,E., Chartois,F., Lemeunier,F. and Cariou,M.L.
    TITLE     Amyrel, a paralogous gene of the amylase gene family in Drosophila
              melanogaster and the Sophophora subgenus
    JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 95 (12), 6848-6853 (1998)
    MEDLINE   98284020
  REFERENCE 2  (bases 1 to 8595)
    AUTHORS   Da Lage,J.-L.
    TITLE     Direct Submission
    JOURNAL   Submitted (03-SEP-1997) Populations, genetique et evolution, Centre
              National de la Recherche Scientifique, Avenue de la Terrasse, Gif
              sur Yvette 91198, France
  COMMENT     On Apr 7, 1998 this sequence version replaced gi:3004813.
  FEATURES             Location/Qualifiers
       source          1. .8595
                       /organism="Drosophila melanogaster"
                       /strain="Canton S"
                       /db_xref="taxon:7227"
                       /map="53D1-3"
                       /chromosome="2"
       mRNA            join(<1. .32,95. .>310)
                       /gene="Picot"
                       /product="putative inorganic phosphate cotransporter"
       gene            <1. .>310
                       /gene="Picot"
       CDS             join(<1. .32,95. .310)
                       /gene="Picot"
                       /note="PICOT; similar to Na dependant inorganic phosphate
                       cotransporter"
                       /codon_start=3
                       /product="putative inorganic phosphate cotransporter"
                       /db_xref="PID:g3004815"
                       /translation="AGHLISDPSKPHMGQWQIVFFIAAFVYIICGTFYNIPGSGERQY
                       WDNPEDDEQKPALQTTVTTSPARLSNGSTAPAAISSS"
       mRNA            join(<1022. .1383,1443. .1733,1791. .2239,2292. .2418,
                       2485. .2588,2644. .>2861)
                       /gene="Nach"
                       /product="putative sodium channel"
       gene            <1022. .>2861
                       /gene="Nach"
       CDS             join(1022. .1383,1443. .1733,1791. .2239,2292. .2418,
                       2485. .2588,2644. .2861)
                       /gene="Nach"
                       /note="NACH; similar to epithelial sodium channels"
                       /codon_start=1
                       /product="putative sodium channel"
                       /db_xref="PID:g3004816"
                       /translation="MGHQEELKPEQVDLKVTPFVGSLRRTWSDPCATSSIHGLKYTRD
                       EDTNKIVHLVWLLISVVMFICAVVMARTFYMDYRSSPTRNNVESDNTPVNRLYFPPVT
                       ICPDVLFNMQKSEAFLNTLRLPKGAELRGILRKLHIFYGFMLDDERYSAEDIEQMEAL
```

```
                     LFLNNLTIPEFVEHLRWNCDEILYRCRFNGEIMDCSKIPQLSKTFFGHCCSFNLRQKG
                     LNFTAQRAIGGLRYGLSVVVRYKDDNYDPLQLYSYGVKLLIQEADAFPSAHFAAKFIA
                     FNSETFAAVRPQETFCFPAVKALIIEERNCVFQNEFPNRYFSDYVYPNCELNCRVTNM
                     VKFCGCHTTYFFAFNRTSDRICTFRDIPCLVDNFANIITRKKSTQCYCPLTCEHIDYDV
                     QLTNFPLELNMPVADKFYSGLAKNDGVLHVFINSFSYRRLRRDLLSNMVTLVSNLGSA
                     FSLFVGMSMLSVVEIIYYFSVILRKNYKLECETRSQMLHKKPKFAWPKANDTHSKEQK
                     SVFIIHKS"
     mRNA            join(<3408. .4061,4118. .>4945)
                     /gene="Amyrel"
                     /product="putative amylase-related protein"
     gene            <3408. .>4945
                     /gene="Amyrel"
     CDS             join(3408. .4061,4118. .4945)
                     /gene="Amyrel"
                     /note="AMYREL"
                     /codon_start=1
                     /product="putative amylase-related protein"
                     /db_xref="PID:g3004814"
                     /translation="MSKFALALTLCLAGSLSVAQHNPHWWGNRNTIVHLFEWKWSDIA
                     QECESFLGPRGFAGVQVSPVNENIISAGRPWWERYQPISYKLTTRSGNEEEFGDMVRR
                     CNDVGVRIYVDVLLNHMSGDFDGVVVGTAGTEAEPSKKSFPGVPYTAQDFHPTCVITD
                     WNDRFQVQCELVGLKDLDQSSDWVRSKLIEFLDHLIELGVAGFRVDAAKHMASEDLE
                     YIYSSLSNLNIDHGFPHNSRPPIPQEVIDHGHETDSRDEYKDLGAVTEFRPSEEIGNA
                     FRGNNALKWLQSWGTDWGFLPSGQALTFVDNHDSQRDAGAVLNYKSPRQYKMATAFHL
                     AYPYGTSRVMSSFAFDDHDTPPPRDAQERIISPEFDADGACVNGWICEHRWRQIYAMV
                     GFKNAVRDTEITGWWDNGDNQISFCRGNKGFLAINNNLYDLSQDLNTCLPAGTYCDVI
                     SGSLIDGSCTGKSVTMSENGYGYIHIGSDDFDGVLALHVDAKV"
     mRNA            complement(7339. .>8595)
                     /gene="gprs"
                     /product="putative serine-enriched protein"
     gene            complement(<7339. .>8595)
                     /gene="gprs"
     CDS             complement(7339. .>8595)
                     /gene="gprs"
                     /codon_start=1
                     /product="putative serine-enriched protein"
                     /db_xref="PID:g3004817"
                     /translation="LQALRRETEVADLAIAQAEAAALAAAEPAKKPLQCPTIEFEPPS
                     RRSSFDPPRSPFLEQLRSPGVDTETDLINLQRLDSGGDSFELVESKWSKSSRGESSFD
                     CPYSSRDTSFDVSISRYQSTSYEDQTSSFEIVDTDEKGQGRRAVDLRKSSIELVDAET
                     FQRSGSSCGGRKSSLETHFDYTFTEGGGRSPSLPFPAMSKKQRTENFRQLHVSQFSAF
                     SRARSPLSQQTSSNYSSRDSYDSSGSYPHGYGYPPEPQRSPYGDPSRKHFPLTVRQKG
                     EEEREVGTFLCTDQRCASIPEPRPSSVLTQQLSTGSMSTPSGTTNGTPRIPGAAPVGP
                     VPLPHPSAPLSSCGFSSGSEPEPPSPRRAASASPKHTFTPRIVMKKVDSSPEALCPER
                     HRSRIIDRYRRRDSRRKRIHDAGKNF"
BASE COUNT     2349 a   1849 c   1953 g   2444 t
ORIGIN

  Query Match           5.7%;  Score 17;  DB 18;  Length 8595;
  Best Local Similarity 100.0%;  Pred. No. 1.61e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```



```
Db    6417 TACATTTATACATTTAT 6433
           ||||||||||||||||
Cp     118 TACATTTATACATTTAT 102


RESULT   39
LOCUS       ATT3T7POL     9208 bp    DNA             PLN       29-AUG-1997
DEFINITION  Arabidopsis thaliana rpoMt gene.
ACCESSION   Y09006
NID         g2330565
KEYWORDS    DNA-dependent RNA polymerase; rpoMt gene.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryotae; Viridiplantae; Charophyta/Embryophyta group;
            Embryophyta; vascular plants; seed plants; Magnoliophyta;
            Magnoliopsida; Capparales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 9208)
  AUTHORS   Hedtke,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-OCT-1996) B. Hedtke, Humboldt University Berlin,
            Institute of Biology (Genetics), Chausseestr.117, D-10115 Berlin,
            FRG
REFERENCE   2  (bases 1 to 9208)
  AUTHORS   Hedtke,B., Borner,T. and Weihe,A.
  TITLE     Mitochondrial and chloroplast phage-type RNA polymerases in
            Arabidopsis
  JOURNAL   Science 277 (5327), 809-811 (1997)
  MEDLINE   97390506
COMMENT     Related sequence: Y08137.
FEATURES             Location/Qualifiers
     source          1. .9208
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
     exon            2503. .3451
                     /number=1
                     /evidence=experimental
     mRNA            join(2503. .3451,3597. .3821,3937. .4070,4177. .4243,
                     4375. .4433,4591. .4708,4851. .4964,5117. .5248,5330. .5746,
                     5907. .5957,6081. .6151,6242. .6326,6500. .6652,6781. .6855,
                     6963. .7052,7164. .7211,7412. .7526,7615. .7718,7836. .8412)
     gene            2690. .7946
                     /gene="rpoMt"
     CDS             join(2690. .3451,3597. .3821,3937. .4070,4177. .4243,
                     4375. .4433,4591. .4708,4851. .4964,5117. .5248,5330. .5746,
                     5907. .5957,6081. .6151,6242. .6326,6500. .6652,6781. .6855,
                     6963. .7052,7164. .7211,7412. .7526,7615. .7718,7836. .7946)
                     /gene="rpoMt"
                     /codon_start=1
                     /product="mitochondrial single-subunit DNA-dependent RNA
```



```
                     polymerase"
                     /db_xref="PID:e283109"
                     /db_xref="PID:g2330566"
                     /translation="MWRNILGRASLRKVKPLSDSSSSGTHYPVNRVRGILSSVNLSGV
                     RNGLSINPVNEMGGLSSFRHGQCYVFEGYATAAQAIDSTDPEDESSGSDEVNELITEM
                     EKETERIRKKARLAAIPPKRVIAGMGAQKFYMLKQRQVKMETEEWERAARECREILAD
                     MCEQKLAPNLPTMKSLFLGWFEPVRNAIQDDLDTFKIKKGKIPTAPFMEQLPADKMAV
                     ITMHKMMGLLMTNAEGVGIVKLVNAATQSEAVEQEVRINSFLQKKNKNATDKTINT
                     EAENVSEEIVAKETKARKQVTVLMEKNKLRQVKALVRKHDSFKPWGQEAQVKVGARL
                     IQLLMENAYIQPPAEQFDDGPPDIRPAFKQNPRTVTLENTKTSRRYGCIECDPLVLKG
                     LDKSARHMVIPYLPMLIPPQNWTGYDQGAHFFLPSYVMRTHGAKQQRTVMKRTPKEQL
                     EPVYEALDTLGNTKWKINKKVLSLVDRIWANGGRIGGLVDREDVPIPEEPEREDQEKF
                     KNWRWESKKAIKQNNERHSQRCDIELKLEVARKMKDEEGFYYPHNVDFRGRAYPIHPY
                     LNHLGSDLCRGILEFCEGKPLGKSGLRWLKIHIANLYAGGVDKLAYEDRIAPTESHLE
                     DIFDSSDRPLEGKRWWLNAEDPFQCLAACINLSEALRSPFPEAAISHIPIHQDGSCNG
                     LQHYAALGRDKLGADAVNLVTGEKPADVYTEIAARVLKIMQQDAEEDPETFPNATAK
                     LMLDQVDRKLVKQTVMTSVYGVTYSGARDQIKKRLKERGTPEDDSLTFHASCYAAKIT
                     LKALEEMFEAARAIKSWFGDCAKIIASENNAVCWTTPLGLPVVQPYRKPGRHLVKTTL
                     QVLTLSRETDKVMARQMTAFAPNFIHSLDGSHMMMTAVACNRAGLSFAGVHDSFWTH
                     ACDVDVMNTILREKFVELYEKPILENLLESFQKSFPDISFPPLPERGDFDLRKVLEST
                     YFFN"
     intron          3452. .3596
                     /gene="rpoMt"
                     /number=1
     exon            3597. .3821
                     /gene="rpoMt"
                     /number=2
                     /evidence=experimental
     intron          3822. .3936
                     /gene="rpoMt"
                     /number=2
     exon            3937. .4070
                     /gene="rpoMt"
                     /number=3
                     /evidence=experimental
     intron          4071. .4176
                     /gene="rpoMt"
                     /number=3
     exon            4177. .4243
                     /gene="rpoMt"
                     /number=4
                     /evidence=experimental
     intron          4244. .4374
                     /gene="rpoMt"
                     /number=4
     exon            4375. .4433
                     /gene="rpoMt"
                     /number=5
                     /evidence=experimental
     intron          4434. .4590
                     /gene="rpoMt"
```



```
                     /number=5
     exon            4591. .4708
                     /gene="rpoMt"
                     /number=6
                     /evidence=experimental
     intron          4709. .4850
                     /gene="rpoMt"
                     /number=6
     exon            4851. .4964
                     /gene="rpoMt"
                     /number=7
                     /evidence=experimental
     intron          4965. .5116
                     /gene="rpoMt"
                     /number=7
     exon            5117. .5248
                     /gene="rpoMt"
                     /number=8
                     /evidence=experimental
     intron          5249. .5329
                     /gene="rpoMt"
                     /number=8
     exon            5330. .5746
                     /gene="rpoMt"
                     /number=9
                     /evidence=experimental
     intron          5747. .5906
                     /gene="rpoMt"
                     /number=9
     exon            5907. .5957
                     /gene="rpoMt"
                     /number=10
                     /evidence=experimental
     intron          5958. .6080
                     /gene="rpoMt"
                     /number=10
     exon            6081. .6151
                     /gene="rpoMt"
                     /number=11
                     /evidence=experimental
     intron          6152. .6241
                     /gene="rpoMt"
                     /number=11
     exon            6242. .6326
                     /gene="rpoMt"
                     /number=12
                     /evidence=experimental
     intron          6327. .6499
                     /gene="rpoMt"
                     /number=12
     exon            6500. .6652
```

```
                     /gene="rpoMt"
                     /number=13
                     /evidence=experimental
     intron          6653. .6780
                     /gene="rpoMt"
                     /number=13
     exon            6781. .6855
                     /gene="rpoMt"
                     /number=14
                     /evidence=experimental
     intron          6856. .6962
                     /gene="rpoMt"
                     /number=14
     exon            6963. .7052
                     /gene="rpoMt"
                     /number=15
                     /evidence=experimental
     intron          7053. .7163
                     /gene="rpoMt"
                     /number=15
     exon            7164. .7211
                     /gene="rpoMt"
                     /number=16
                     /evidence=experimental
     intron          7212. .7411
                     /gene="rpoMt"
                     /number=16
     exon            7412. .7526
                     /gene="rpoMt"
                     /number=17
                     /evidence=experimental
     intron          7527. .7614
                     /gene="rpoMt"
                     /number=17
     exon            7615. .7718
                     /gene="rpoMt"
                     /number=18
                     /evidence=experimental
     intron          7719. .7835
                     /gene="rpoMt"
                     /number=18
     exon            7836. .8412
                     /number=19
                     /evidence=experimental
BASE COUNT     2688 a    1774 c    1826 g    2920 t
ORIGIN

  Query Match           5.7%;  Score 17;  DB 23;  Length 9208;
  Best Local Similarity  94.4%;  Pred. No. 1.61e+01;
  Matches   17;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;
```



```
Db    5061 CCTACTTCTCTTTTAGAA 5078
           || ||||||||||||||||
Cp     219 CCNACTTCTCTTTTAGAA 202


RESULT   40
LOCUS       ATSSDDRPL    9764 bp    DNA             PLN       13-DEC-1996
DEFINITION  A.thaliana genes encoding mitochondrial DNA-directed RNA
            polymerase, tRNA-Met initiator and ORF with unknown function.
ACCESSION   Y09432
NID         g1742970
KEYWORDS    mitochondrial DNA-directed RNA polymerase; tRNA-Met initiator.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryotae; mitochondrial eukaryotes; Viridiplantae;
            Charophyta/Embryophyta group; Embryophyta; Magnoliophyta;
            Magnoliopsida; Capparales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 9764)
  AUTHORS   Schuster,W.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 9764)
  AUTHORS   Schuster,W.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-NOV-1996) W. Schuster, Institut fuer Genbiologische
            Forschung, Berlin GmbH, Ihnestrasse 63, D- 14195 Berlin, FRG
FEATURES             Location/Qualifiers
     source          1. .9764
                     /organism="Arabidopsis thaliana"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /dev_stage="three weeks after germination"
                     /tissue_type="whole plants"
     exon            complement(<762. .1055)
                     /gene="ORF"
                     /number=4
     mRNA            complement(join(<762. .1055,1139. .1268,1396. .1757,
                     1893. .>1982))
                     /gene="ORF"
                     /note="unknown function"
     gene            complement(762. .1982)
                     /gene="ORF"
     intron          complement(1056. .1138)
                     /gene="ORF"
                     /number=3
     exon            complement(1139. .1268)
                     /gene="ORF"
                     /number=3
     intron          complement(1269. .1395)
                     /gene="ORF"
                     /number=2
     exon            complement(1396. .1757)
```



```
                     /gene="ORF"
                     /number=2
     intron          complement(1758. .1892)
                     /gene="ORF"
                     /number=1
     exon            complement(1893. .>1982)
                     /gene="ORF"
                     /number=1
     terminator      complement(3305. .3318)
     tRNA            complement(<3332. .>3407)
                     /product="tRNA-Met initiator"
     TATA_signal     3426. .3436
                     /note="putative"
     exon            <3812. .4573
                     /number=1
     mRNA            join(<3812. .4573,4722. .4946,5062. .5195,5302. .5368,
                     5500. .5558,5717. .5834,5977. .6090,6242. .6373,6455. .6871,
                     7032. .7082,7206. .7276,7367. .7451,7623. .7775,7904. .7978,
                     8086. .8175,8287. .8334,8535. .8649,8738. .8841,8959. .>9069)
                     /product="mitochondrial DNA-directed RNA polymerase"
     CDS             join(3812. .4573,4722. .4946,5062. .5195,5302. .5368,
                     5500. .5558,5717. .5834,5977. .6090,6242. .6373,6455. .6871,
                     7032. .7082,7206. .7276,7367. .7451,7623. .7775,7904. .7978,
                     8086. .8175,8287. .8334,8535. .8649,8738. .8841,8959. .9069)
                     /note="putative"
                     /codon_start=1
                     /product="mitochondrial singel subunit DNA-directed RNA
                     polymerase"
                     /db_xref="PID:e285318"
                     /db_xref="PID:g1742971"
                     /translation="MWRNILGRASLRKVKPLSDSSSSGTHYPVNRVRGILSSVNLSGV
                     RNGLSINPVNEMGGLSSFRHGQCYVPEGYATAAQAIDSTDPEDESSGSDEVNELITEM
                     EKETERIRKKARLAAIPPKRVIAGMGAQKFTMLKQRQVKMETEEWERAARECREILAD
                     MCEQKLAPNLPYMKSLFLGWFEPVRNAIQDDLDTFKIKKGKIPYAPPMEQLPADKMAV
                     ITHKKMMGLLMTNAEGVGIVKLVNAATQIGEAVEQEVRINSPLQKNNKNATDKTINI
                     EAENVSEEIVAKETEKARKQVTVLKEKNKLRQVKALVRKHDSFKPWGQEAQVEVGARL
                     IQLLMENAYIQPPAEQFDDQGPPDIRPAFKQNFRTVTLENTKTSRRYGCIECDPLVLKG
                     LDKSARHMVIPYLPMLIPPQNWTGTDQGAHFFLPSTVKRTHGAKQQRTVMKRTPKEQL
                     EPVYEALDTLGNTKWKINKKVLSLVDRIWANGGRIGGLVDREDVPIPEEPEREDQEKF
                     KNWRWESKKAIKQNNERHSQRCDIELKLEVARKMKDEEGFYTPHNVDPRGRAYPIHPY
                     LNHLGSDLCRGILEPCEGKPLGKSGLRWLKIHIANLYAGGVDKLAYEDRIAFTESHLE
                     DIFDSSDRPLEGKRWWLNAEDPFQCLAACINLSEALRSPFPEAAISHIPIHQDGSCNG
                     LQHYAALGRDKLGADAVNLVTGEKPADVTTEIAARVLKIMQQDAEEDPETFPNATYAK
                     LMLDQVDRKLVKQTVMTSVYGVTYSGARDQIKKRLKERGTFEDDSLTFHASCYAAKIT
                     LKALEEMKFEAARAIKSWFGDCAKIIASENNAVCWTTPLGLPVVQPYRKPGRHLVKTTL
                     QVLTLSRETDKVHARRQMTAFAPNFIHSLDGSHMMMTAVACNRAGLSFAGVHDSFWTH
                     ACDVDVMNTILREKFVELYEKPILENLLESFQKSFPDISFPPLPERGDFDLRKVLEST
                     YFFR"
     intron          4574. .4721
                     /number=1
     exon            4722. .4946
```



```
                     /number=2
     intron          4947. .5061
                     /number=2
     exon            5062. .5195
                     /number=3
     intron          5196. .5301
                     /number=3
     exon            5302. .5368
                     /number=4
     intron          5369. .5499
                     /number=4
     exon            5500. .5558
                     /number=5
     intron          5559. .5716
                     /number=5
     exon            5717. .5834
                     /number=6
     intron          5835. .5976
                     /number=6
     exon            5977. .6090
                     /number=7
     intron          6091. .6241
                     /number=7
     exon            6242. .6373
                     /number=8
     intron          6374. .6454
                     /number=8
     exon            6455. .6871
                     /number=9
     intron          6872. .7031
                     /number=9
     exon            7032. .7082
                     /number=10
     intron          7083. .7205
                     /number=10
     exon            7206. .7276
                     /number=11
     intron          7277. .7366
                     /number=11
     exon            7367. .7451
                     /number=12
     intron          7452. .7622
                     /number=12
     exon            7623. .7775
                     /number=13
     intron          7776. .7903
                     /number=13
     exon            7904. .7978
                     /number=14
     intron          7979. .8085
                     /number=14
```

```
     exon            8086. .8175
                     /number=15
     intron          8176. .8286
                     /number=15
     exon            8287. .8334
                     /number=16
     intron          8335. .8534
                     /number=16
     exon            8535. .8649
                     /number=17
     intron          8650. .8737
                     /number=17
     exon            8738. .8841
                     /number=18
     intron          8842. .8958
                     /number=18
     exon            8959. .9066
                     /number=19
BASE COUNT     2902 a    1836 c    1955 g    3071 t
ORIGIN

  Query Match           5.7%;  Score 17;  DB 23;  Length 9764;
  Best Local Similarity  94.4%;  Pred. No. 1.6le+01;
  Matches    17;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db   6186 CCTACTTCTCTTTTAGAA 6203
          || ||||||||||||||||
Cp    219 CCNACTTCTCTTTTAGAA 202


RESULT   41
LOCUS       CEC34E7     27008 bp    DNA             INV       21-SEP-1998
DEFINITION  Caenorhabditis elegans cosmid C34E7, complete sequence.
ACCESSION   Z67880
NID         g1065688
KEYWORDS    HTG; Dif.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 27008)
  AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
```



```
            Wilkinson-Sproat,J. and Wohldman,P.
  TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
  JOURNAL   Nature 368 (6466), 32-38 (1994)
  MEDLINE   94150718
REFERENCE   2  (bases 1 to 27008)
  AUTHORS   Lloyd,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-NOV-1996) Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     Coding sequences below are predicted from computer analysis, using
            predictions from Genefinder (P. Green, U. Washington), and other
            available information.
            For a graphical representation of this sequence and its analysis
            see:-
            http://webace.sanger.ac.uk/cgi-
            bin/display?db=wormace&class=Sequence &object=C34E7
            Current sequence finishing criteria for the C. elegans genome
            sequencing consortium are that all bases are either sequenced
            unambiguously on both strands, or on a single strand with both a
            dye primer and dye terminator reaction, from distinct subclones.
            Exceptions are indicated by an explicit note.
            IMPORTANT:  This sequence is NOT necessarily the entire insert of
            the specified clone.  It may be shorter because we only sequence
            overlapping sections once, or longer because we arrange for a small
            overlap between neighbouring submissions.
            IMPORTANT: This sequence is not the entire insert of clone C34E7.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we arrange for a small overlap between
            neighbouring submissions.
            The true left end of clone C34E7 is at 1 in this sequence. The true
            right end of clone C34E7 is at 14303 in
            sequence Z67879.
            The true left end of clone C05E7 is at 26908 in this sequence. The
            true right end of clone K02A4 is at 6791 in this sequence. The
            start of this sequence (1. .100) overlaps with the end of sequence
            Z67883.
            The end of this sequence (26908. .27008) overlaps with the start of
            sequence Z67879.
FEATURES                Location/Qualifiers
     source             1. .27008
                        /organism="Caenorhabditis elegans"
                        /db_xref="taxon:6239"
                        /chromosome="X"
                        /clone="C34E7"
     gene               6184. .9607
                        /gene="C34E7.1"
     CDS                join(6184. .6698,8270. .8384,9020. .9211,9260. .9387,
                        9487. .9607)
                        /gene="C34E7.1"
                        /note="Similarity to Drospohila Rel-containing protein Dif
```



```
                        (PIR Acc. No. A49435)"
                        /codon_start=1
                        /db_xref="PID:g1065689"
                        /db_xref="SPTREMBL:Q18456"
                        /translation="MYSATPPSQPQPKQSRQQQYPSQQPPAYNYPHQGYGQMGQMGHG
                        SLPMDTRGIQQQPQNMGTMMGRQVPLQQQAAAGQMNAHGMPMDNRSMNSGQMQHLDNR
                        MGSQQMTTNQRPMGMESHNMGQLGNMQMGSMGSMNSSMSNSMASMGGMMSDRTMGTQQ
                        HQMMGQHQQQQSTSNPGNNQQMQRMMNPPAYSQQEKQRMEQYQRQQHQSQMQSYQSQM
                        QQSMTPVVQAPAPAKPAPKNNSRRKKTVNKANLDEPTQVPAPNYHPQPPPSYGYQQNN
                        NMMGNPQVSGHQMAGGQQQLQHHAMMHGGPGGSQSQYPNTKSSEMVRLELRNSIQAKQ
                        QQQKQKEQPPMDQQRKFCPRRF"
     gene               11852. .12399
                        /gene="C34E7.2"
     CDS                join(11852. .12151,12203. .12292,12343. .12399)
                        /gene="C34E7.2"
                        /codon_start=1
                        /db_xref="PID:g1065690"
                        /db_xref="SPTREMBL:Q18457"
                        /translation="MQRAMQQDSSMMMGNQMGQYQYSQMGQPPHYHQSSSMGNGMGQA
                        PQMGQGHPQQMQEMGSFNTPQPQVQINQNTPISSVASTSSASSLDFPNELGFFDMQDP
                        ELPVSCDKLEENLKSLTVHDFAILGRCAFDFSNQQTIDFVRQILLM"
     gene               23197. .24886
                        /gene="C34E7.3"
     CDS                join(23197. .23213,23783. .23891,23975. .24101,24270. .24427,
                        24479. .24742,24791. .24886)
                        /gene="C34E7.3"
                        /codon_start=1
                        /db_xref="PID:g1065691"
                        /db_xref="SPTREMBL:Q18458"
                        /translation="MEFWKSAFNNLNTKYIIAIPEPYLPTDLSSEHLFGIYVSLSDAN
                        ANDTKMPKGEREKRQKKLQYQLKRRNSVLMANFKEELFQPIAAGMRRSWSCVELGNIS
                        CHETTQTKRVSFSVTVSRPAQQDIEKTTYDLCPPKISARPILVRRNTTFSGKDNLVHM
                        DFDSFTAWRRGSKMSDDVEKQRKKSMKEKEKEANGGSRMKIRKDSNSSKKEDMEHGDW
                        WIVRVAKMGFESCFQRRRITPSPPPTNETDLKSPVNAN"
     CDS                join(complement(Z67879:738. .849),complement(26380. .26501),
                        complement(26203. .26330),complement(25712. .25805),
                        complement(25371. .25661))
                        /gene="C34E7.4"
                        /note="cDNA EST EMBL:C07750 comes from this gene; cDNA EST
                        EMBL:C09239 comes from this gene; cDNA EST yk215d9.3 comes
                        from this gene; cDNA EST yk215d9.5 comes from this gene;
                        cDNA EST yk327d8.3 comes from this gene; cDNA EST
                        yk327d8.5 comes from this gene; cDNA EST yk330d12.3 comes
                        from this gene; cDNA EST yk330d12.5 comes from this gene;
                        cDNA EST yk331b5.5 comes from this gene; cDNA EST
                        yk363h12.5 comes from this gene; cDNA EST yk364a5.5 comes
                        from this gene; cDNA EST yk409d5.5 comes from this gene;
                        cDNA EST yk442e1.3 comes from this gene; cDNA EST
                        yk442e1.5 comes from this gene; cDNA EST yk453b6.5 comes
                        from this gene"
                        /codon_start=1
```



```
                        /db_xref="PID:e1321839"
                        /db_xref="PID:g3642079"
                        /translation="MTTRRIVLLCAVAACVSAASSEISVRIERHFPCSASSGPKKDSL
                        LIKFPSYKTAGVQFHEEIGENGHKCFRMSGGKVEVTAPGLNGSKKYIVHLETRIGIHG
                        KPERCVHADANGCGGIGSCVHCDICHTMGGSLKNFVQIFQADKPAQCSTTGLSAGSYS
                        DLSLRVCLPTKNELLPFLDQNASRAEQLWDLFVSSRARSGEIPLVIAARLFDRPINNL
                        DAKTLNTILHDSKEGMIGCHWIYATVSQPN"
BASE COUNT     8430 a    4983 c    4446 g    9149 t
ORIGIN

  Query Match           5.7%;  Score 17;  DB 18;  Length 27008;
  Best Local Similarity 100.0%;  Pred. No. 1.6le+01;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  19934 ATTTATACATTTATAAA 19950
          |||||||||||||||||
Cp    115 ATTTATACATTTATAAA 99


RESULT   42
LOCUS       CEM01B2     27013 bp    DNA             INV       21-SEP-1998
DEFINITION  Caenorhabditis elegans cosmid M01B2, complete sequence.
ACCESSION   Z83116
NID         g2645034
KEYWORDS    HTG.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 27013)
  AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
  TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
  JOURNAL   Nature 368 (6466), 32-38 (1994)
  MEDLINE   94150718
REFERENCE   2  (bases 1 to 27013)
  AUTHORS   Lloyd,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-NOV-1996) Nematode Sequencing Project, Sanger Centre,
            Hinxton, Cambridge CB10 1RQ, England and Department of Genetics,
            Washington University, St. Louis, MO 63110, USA. E-mail:
```

```
            jes@sanger.ac.uk or rw6nematode.wustl.edu
  COMMENT   On Nov 26, 1997 this sequence version replaced gi:1752968.
            Coding sequences below are predicted from computer analysis, using
            predictions from Genefinder (P. Green, U. Washington), and other
            available information.
            For a graphical representation of this sequence and its analysis
            see:-
            http://webace.sanger.ac.uk/cgi-
            bin/display?db=wormace&class=Sequence &object=M01B2
            Current sequence finishing criteria for the C. elegans genome
            sequencing consortium are that all bases are either sequenced
            unambiguously on both strands, or on a single strand with both a
            dye primer and dye terminator reaction, from distinct subclones.
            Exceptions are indicated by an explicit note.
            IMPORTANT:  This sequence is NOT necessarily the entire insert of
            the specified clone.  It may be shorter because we only sequence
            overlapping sections once, or longer because we arrange for a small
            overlap between neighbouring submissions.
            IMPORTANT: This sequence is not the entire insert of clone M01B2.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we arrange for a small overlap between
            neighbouring submissions.
            The true left end of clone M01B2 is at 1 in this sequence. The true
            left end of clone T10H4 is at 26910 in this sequence. The true
            right end of clone R10G11 is at 18127 in this sequence. The start
            of this sequence (1. .6199) overlaps with the end of sequence
            Z81047.
            The end of this sequence (26910. .27013) overlaps with the start of
            sequence Z81119.
  FEATURES             Location/Qualifiers
     source            1. .27013
                       /organism="Caenorhabditis elegans"
                       /db_xref="taxon:6239"
                       /chromosome="V"
                       /clone="M01B2"
     gene              3252. .5611
                       /gene="M01B2.6"
     CDS               join(3252. .3417,3931. .4406,4795. .5611)
                       /gene="M01B2.6"
                       /codon_start=1
                       /pseudo
                       /db_xref="PID:e1322832"
     gene              6431. .9155
                       /gene="M01B2.1"
     CDS               join(6431. .6627,6847. .7160,7216. .7351,7395. .7677,
                       8187. .8296,8341. .8577,8942. .9050,9090. .9155)
                       /gene="M01B2.1"
                       /note="similar to tyrosine-protein kinase (KIN15/KIN16
                       subfamily)"
                       /codon_start=1
                       /db_xref="PID:e1186479"
```



```
                       /db_xref="PID:g2645035"
                       /db_xref="SPTREMBL:O17955"
                       /translation="MNPPFPNRFARFVRSVENHDESEKPNLIIVFLMCALFIYGVLSI
                       IFVTICFLRRIYRSSATKNGNRRYLVQNTYVKAQLDTNSSSSLPVDKEIIETISINEE
                       TPIITDVQPLNERAEYLPYNPAFEIHSEHLDVFEKQPGQGNFGKVNKALLKLINQKTG
                       EVVRMDVAVKEPADSTDKSQDKLLDEIKLMCAIGKHPNVLAIVGAITKQEKVIGREH
                       NLVVTEFVEGGDLRSVLKTPTTFHDELTSTDRTGQQMVNSDIFDELSTSDLYSFATQ
                       IANGMEYLAAKPVRFFWGKSGANFQPLKKSEFPKCVHRDLALRNVFVKRNKMIRIGDF
                       GLARHHSKKSYYRMQCNPDTPLPIFWLAPECFNESKFTEMTDVWSYGVCLPELFSLGE
                       SPYKKLHNSPSYDVVHYLKKGYRLSAPRYCNAEIYEFMLYCWNIDATLRPKFTECKDF
                       CKSLITRKKVGHLKKIESRLQMEEPQLQNELSLA"
     gene              9843. .11125
                       /gene="M01B2.5"
     CDS               join(9843. .10114,10147. .10296,10452. .10674,10689. .10909,
                       10942. .11125)
                       /gene="M01B2.5"
                       /codon_start=1
                       /db_xref="PID:e1186483"
                       /db_xref="PID:g2645039"
                       /db_xref="SPTREMBL:O17959"
                       /translation="MPENILSLIDNLLTPIFLYLNNLNFTVSVVGVIANLFHLSVLFR
                       KPMRILTINIFMIGIALCDLCKMISVIFTAIINYTQSVIQSQMPTECFQRSISNIYRT
                       YNNQFMLKGIRFSKKLSVWLAVSLAILRAVVIKYPINTKIQDFTSSAHGIRIFLLICL
                       MNSPLWYLQNRFSIDEQEKWTTPIECLNFTENSTQIIYLENSSGLDYSTSLLIDGIIP
                       SILLPLSTIVLLYELKKIKKLTASTNSRSTDNWTRTTKLVCPVTITFLISTVPLGILF
                       LIEMIPFDTDGLDIIRKISSIFKLISLINGTIHFLICFFMSTLYRKTVKENFGMQPHN
                       GNLINVQPTFRKSEI"
     gene              12272. .13339
                       /gene="M01B2.2"
     CDS               join(12272. .12450,12506. .12995,13046. .13339)
                       /gene="M01B2.2"
                       /codon_start=1
                       /db_xref="PID:e1186480"
                       /db_xref="PID:g2645036"
                       /db_xref="SPTREMBL:O17956"
                       /translation="MLYQLLHTVLSVVGVSLNAFMMYLALTKSPKIMLPCSAIITIKT
                       FTDILTSAMSFFVMQRIVTDGSSILVIPTGPCTHLGPTACYIGHMFMLCFLECNLIWM
                       ISSYIFRYYILYVRDPSIKSLVFVAICLSIPSFIHMATWISNYDPTVAIAIPENVGIE
                       SRDMVLGGTIVTWSALTLIIQLFITSILVLIAYAWIRNTLLSFAVKMGSDKNDVKKLN
                       TRLVKVINFQVFLPSFIFLGVFVFVGMPTQLIDPKISQYLVSVIFMFSPICSPFSYIL
                       FVPHYLNVILGNKKVSEAKSTTEGCTFRHVNMAASTSNTPECSA"
     gene              14050. .14959
                       /gene="M01B2.4"
     CDS               join(14050. .14682,14729. .14959)
                       /gene="M01B2.4"
                       /codon_start=1
                       /db_xref="PID:e1186482"
                       /db_xref="PID:g2645038"
                       /db_xref="SPTREMBL:O17958"
                       /translation="MPWVLLFNYIISIVFATTAFFLNGFLLLSIFYYKKITVNPSMTF
                       IYWKFGADLFYTLSLSICISYYLLMSLSANFIIKNLTFWLVWASTSIGSMRSILALSI
```



```
                       SIERVFATFIPVYFHKYRSKFPNLLVLCIILAKGFLDQYILFGFCGNDIDAPLECNGI
                       PCAVNSCYNSPYFSSELGVYFVNGAFSATLFFRFPIWNYFSKMQKNCIISRATRIALL
                       DFFILLFFNVLSSHFVNNLRIVNIREVGPLTVATKMPGFVIEGLITCRVLFGGKLVST
                       FVATSNSLKLI"
     gene              15497. .16473
                       /gene="M01B2.3"
     CDS               join(15497. .16129,16252. .16473)
                       /gene="M01B2.3"
                       /codon_start=1
                       /db_xref="PID:e1186481"
                       /db_xref="PID:g2645037"
                       /db_xref="SPTREMBL:O17957"
                       /translation="MILLSLLNYLLSILFATAAFFLNCYILFSIFYQKRIPINQSMTF
                       IYWKFSVDVVYTLNLTTLMFYYLLISLSANFIIKNLTFWLVWASTSIGSMRSLLALLI
                       SFERVLATYFPIIFHKYRQRFPNFIVLLIIVTSGFFDQYTLFGYCGNDIDTPLECKGF
                       SCAVNTCYYSYYFSNERVVCFWNGVFSLTLFFRFPVWKYVSKTQTNERISRATRIALL
                       DFFINIFFNILASYITTIFPNLQNVGPLTIVTKTLGFVIEGVITCRVLFGAKNVVTVV
                       TTNSQKSS"
     gene              complement(17357. .18437)
                       /gene="M01B2.9"
     CDS               complement(join(17357. .17589,17675. .17817,18018. .18126,
                       18173. .18285,18337. .18437))
                       /gene="M01B2.9"
                       /note="similar to 7 transmembrane receptor (rhodopsin
                       family)"
                       /codon_start=1
                       /db_xref="PID:e1186485"
                       /db_xref="PID:g2645040"
                       /db_xref="SPTREMBL:O17960"
                       /translation="MEPLLIILCCIYCTVFVVGTTGNLIMVMVTFYCKNLRSICNILI
                       GPCCFCDLLLLFTNIIAFMISMFMPITQEPCTFMSIPADPGAFASSACVLDVGIDRLIA
                       VALPARNFIFVVCLVPESLGHAYDLFALTSFFINLFVLPIYVYVYLKIKKMGLNSSMK
                       AVFKSLMITVCLVLCGWMATDTVGALSLTLPIDKNVARMMQLYAGVFILSSSSFMAFV
                       YYRIRYDMGEHLFR"
     gene              complement(19249. .20817)
                       /gene="M01B2.7"
     CDS               complement(join(19249. .19384,19554. .19755,19803. .19936,
                       19988. .20143,20398. .20506,20560. .20672,20717. .20817))
                       /gene="M01B2.7"
                       /note="similar to 7 transmembrane receptor (rhodopsin
                       family)"
                       /codon_start=1
                       /db_xref="PID:e1186486"
                       /db_xref="PID:g2645041"
                       /db_xref="SPTREMBL:O17961"
                       /translation="MELLLVILLFLYSTIFIVGFTGNLLMVLVTLHSNKLRSICNILI
                       CVCCFCDLLLLFTDVIAFVVSMFTPSTQEPCFYISIPADPGAFASHACVLAIGIDRLIA
                       VAAFTRIIFSSRYSYHNCFRYKLLENERKYLFLQMTPPVIYSSTLLIMGPGQRDPRR
                       NVVCLLPESLGHAYDMFALSSFGINLFVPPLYAYVYFKVKNMSDNNSIKTVFKSLSVT
                       VCLVVCGWMTTDLIGALTVILPMDRSVARMVQLYCGTFIFTSSAFNAVVYYKFSRDYR
                       SAVRTMLGLSDEVSIANINIYQQASDPPKSSNQRPFTNTS"
```



```
     gene              21906. .22454
                       /gene="M01B2.8"
     CDS               join(21906. .22002,22057. .22210,22253. .22454)
                       /gene="M01B2.8"
                       /codon_start=1
                       /db_xref="PID:e1186487"
                       /db_xref="PID:g2645042"
                       /db_xref="SPTREMBL:O17962"
                       /translation="MFQQLLIFLVLIEALLACIPTQQIEPPTEAPFPCNVCSKIYNSG
                       CQGFGLPSASNWCSTAAQVPVSYTLGVGPSEASSLPDVCSSQFTCPAGTFIKVTLING
                       VTVISGNTNGAPQVVYCFETGAYAATWWVHIDDDDHSYDISSIECKNL"
     gene              complement(22652. .24134)
                       /gene="M01B2.10"
     CDS               complement(join(22652. .23319,23519. .23713,23804. .24134))
                       /gene="M01B2.10"
                       /codon_start=1
...
Note: remainder of annotations omitted.

  Query Match          5.7%;  Score 17;  DB 18;  Length 27013;
  Best Local Similarity 100.0%;  Pred. No. 1.61e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  24302 ATAAATGTATAAATGTA 24318
          |||||||||||||||||
Qy    102 ATAAATGTATAAATGTA 118


RESULT   43
LOCUS       FRHDGENE    27470 bp    DNA             VRT       17-FEB-1997
DEFINITION  F.rubripes Huntington's disease gene homologue.
ACCESSION   X82939
NID         g790997
KETWORDS    homologue; Huntington's disease gene.
SOURCE      Fugu rubripes.
  ORGANISM  Fugu rubripes
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Actinopterygii; Neopterygii; Teleostei; Euteleostei;
            Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontoidei;
            Tetraodontidae; Fugu.
REFERENCE   1  (bases 1 to 27470)
  AUTHORS   Baxendale,S., Abdulla,S., Elgar,G., Buck,D., Berks,M., Micklem,G.,
            Durbin,R., Bates,G., Brenner,S., Beck,S. and Lehrach,H.
  TITLE     Comparative sequence analysis of the human and pufferfish
            Huntington's disease genes
  JOURNAL   Nat. Genet. 10 (1), 67-76 (1995)
  MEDLINE   95375788
REFERENCE   2  (bases 1 to 27470)
  AUTHORS   Beck,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-NOV-1994) S. Beck, Imperial Cancer Research Fund, 44
```

```
            Lincoln's Inn Fields, London WC2A 3PX, UK
  FEATURES             Location/Qualifiers
       source          1. .27470
                       /organism="Fugu rubripes"
                       /db_xref="taxon:31033"
       repeat_unit     101. .108
                       /note="DRPT1a"
       repeat_unit     109. .116
                       /note="DRPT1b"
       repeat_unit     117. .124
                       /note="DRPT1c"
       repeat_region   258. .289
                       /note="MVR"
       repeat_unit     762. .775
                       /note="IRPT1a"
       repeat_region   948. .1008
                       /note="MVR"
       CDS             join(3223. .3320,3858. .3941,4079. .4199,4325. .4384,
                       4588. .4667,4771. .4909,5140. .5281,5363. .5541,5657. .5861,
                       6084. .6131,6626. .6703,6781. .7163,8438. .8564,8637. .8752,
                       8846. .8957,9359. .9496,9654. .9812,9894. .9991,10065. .10204,
                       10408. .10471,10556. .10656,10766. .10912,11039. .11159,
                       11292. .11431,12908. .13089,13167. .13390,13921. .14047,
                       14131. .14255,14345. .14455,14548. .14625,14882. .15105,
                       15207. .15285,15367. .15528,15620. .15675,15771. .15919,
                       16017. .16153,16222. .16338,16465. .16587,16793. .17052,
                       17133. .17275,17407. .17626,17718. .17856,17981. .18160,
                       18354. .18530,18712. .18788,18991. .19126,19213. .19335,
                       19482. .19689,19756. .19901,19975. .20152,20240. .20344,
                       20447. .20631,20701. .20827,20935. .21038,21167. .21321,
                       21408. .21547,21629. .21711,22073. .22203,22279. .22411,
                       22495. .22650,22727. .22917,23007. .23121,23248. .23461,
                       24484. .24592,24703. .24865,25382. .25542,25698. .25911)
                       /gene="Huntington's Disease Gene - Fugu homologue FrHD"
                       /codon_start=1
                       /db_xref="PID:g804981"
                       /db_xref="SWISS-PROT:P51112"
                       /translation="MATMEKLMKAFESLKSFQQQQGPPTAEEIVQRQKKEQATTKKDR
                       VSHCLTICENIVAQSLRTSPEFQKLLGIAMEMFLLCSDDSESDVRMVADECLNRIIKA
                       LMDSNLPRLQLELYKEIKKNGASRSLRAALWRFAELAHLIRPQKCRPYLVNLLPCLTR
                       ITKRQEETIQETLAAAMPKIMAALGHFANDGEIKMLLKSFVANLKSSSPTIRRTAASS
                       AVSVCQHSRRTSYFYTWLLNVLGLLVPVDEEHSHLILGVLLTLRYLMPLLQQQVNT
                       ISLKGSFGVMQKEADVQPAPSQLQVYELTLHYTQHWDHNVVTAALELLQQTLRTPPP
                       ELLHVLITAGSIQRASVFRQDIESRARSGSILELIAGGGSTCSPLLRRKHRGKMLSGE
                       EDALEDDPEKTDVTTGYFTAVGADNSSAAQVDIITQQPRSSQHTIQPGDSVDLSASSE
                       QGGRGGGASASDTPESPNDEEDMLSRSSSCGANITPETVEDATPENPAQEGRPVGGSG
                       AYDHSLPPSDSSQTTTEGPDSAVTPSDVAELVLDGSESQYSGMQIGTLQDEEDEGTAT
                       SSQEDPPDPPFLRSALALSKPHLFESRGHNRQGSDSSVDRFIPKDEPPEPEPDNKMSRI
                       KGAIGHYTDRGAEPVVRCVRLLSASFLLTGQKNGLTPDRDVRVSVKALAVSCVGAAAA
                       LHPEAFFNSLYLEPLDGLRAEEQQYISDVLGFIDHGDPQIRGATAILCAAIIQAALSK
                       MRYNIHSWLASVQSKTGNPLSLVDLVPLLQKALKDESSVTCKMACSAVRHCIMSLCGS
```



```
                       TLSELGLRLVVDLFALKDSSYWLVRTELLETLAEMDFRLVNFLERKSEALHKGEHHYT
                       GRLRLQERVLNDVVIQLLGDDDPRVRHVAASAVSRLVSRLFFDCDQGQADPVVAIARD
                       QSSVYLQLLMHETQPPSQLTVSTITRTYRGFNLSNNVADVTVENNLSRVVTAVSHAPT
                       SSTSRALTFGCCEALCLLAVHFPICTWTTGWHCGHISSQSSFSSRVGRSRGRTLSVSQ
                       SGSTPASSTTSSAVDPERRTLTVGTANMVLSLLSSAWFPLDLSAHQDALLLCGNLLAA
                       VAPKCLRNPWAGEDDSSSSSTNTSGGTHKMEEPWAALSDRAFVAMVEQLFSHLLKVLN
                       ICAHVLDDTPPGPPVKATLPSLTNTPSLSPIRRKGKDKDAVDSSSAPLSPKKGNEANT
                       GRPTESTGSTAVHKSTTLGSFYHLPPYLKLYDVLKATHANFKVMLDLHSHQSKFGSFL
                       RAALDVLSQLLELATLNDINKCVEELLGYLKSCFSREPTWATVCVQQLLKTLFGTNLA
                       SQYEQFLSGPSRSQGKALRLGSSSLRPGLYHYCFMAPYTHFTQALADASLRNMVQAEH
                       EQDTSGWFDVMQKTSNQLRSNIANAARHRGDKNAIHNHIRLFEPLVIKALKQYTTSTS
                       VALQRQVLDLLAQLVQLRVNYCLLDSDQVPIGFVLKQFEYIEVGQFRDSEAIIPNIFF
                       FLVLLSYERYHSKQIISIPKIIQLCDGIMASGRKAVTHAIPALQPIVHDLFVLRGSNK
                       ADAGKELETQKEVVVSMLLRLVQTHQVLEMFILVLQQCHKENEDKWKRLSRQIADVIL
                       PMIAKQQMHLDSPEALGVLNTLFETVAPSSLRPVDMLLKSMFTTPVTMASVATVQLWV
                       SGILAVLRVLVSQSTEDIVLSRIHELSLSPHLLSCHTIKRLQQPNLSPSDQPAGDGQQ
                       NQEPNGEAQKSLPEETFARFLIQLVGVLLDDISSRHVKVDITEQQHTFYCQQLGTLLM
                       CLIHVFKSGMFRRITVAASRLLKGESGSGSHSGIEFYPLEGLNSMVHCLITTHPSLVLL
                       WCQVLLIDYTNYSWWTEVHQTPKGHSLSCTKLLSPHSSGEGEEKPETRLAMINREIV
                       RRGALILPCDTVCQNLHDSEHLTWLIVNHVRDLIDLSHEPPVQDFISAVHRNSAASGL
                       FIQAIQSRCDNLNSPTMLKKTLQCLEGIHLSQSGSLLMLYVDKLLSTPFRVLARMVDT
                       LACRRVEMLLAETLQNSVAQLPLEELHRIQEYLQTSGLAQRHQRFTSLLDRFRATVSD
                       TSSPSTPVTSHPLDGDPPPAPELVIADKEWYVALVKSQCCLHGDVSLLETTELLTKLP
                       PADLLSVMSCKEFNLSLLCPCLSLGVQRLLRGQGSLLLETALQVTLEQLAGATGLLPV
                       PHHSFIPTSHPQSHWKQLAEVYGDPGFYSRVLSLCRALSQYLLTVKQLPSSLRIPSDK
                       EHLITTFTCAATEVVVWHLLQDQLPLSVDLQWALSCLCLALQQPCVWNKLSTPEYNTH
                       TCSLIYCLHHIILAVAVSPGDQLLHPERKKTKALRHSDDEDQVDSVHDNHTLEWQACE
                       IMAELVEGLQSVLSLGHHRNTAFPAFLTPTLRNIIISLSRLPLVNSHTRVPPLVWKLG
                       WSPQPGGEFGTTLPEIPVDFPLQEKDVFREFLYRINTLGWSNRTQFEETWATLLGVLVT
                       QPITMDQEEETQQEEDLERTQLNVLAVQAITSLVLSAMTLPTAGNPAVSCLEQQPRNK
                       SLKALETRFGRKLAVIRGEVEREIQALVSKRDNVHTYHPYHAWDPVPSLSAASPGTLI
                       SHEKLLLQINTERELGNMDYKLGQVSIHSVWLGNNITPLREEEWGEDEDDEADPPAPT
                       SPPLSPINSRKHRAGVDIHSCSQFLLELYSQWVIPGSPSNRKTPTILISEVVRSLLAV
                       SDLFTERNQFDMMFSTLMELQKLHPPEDEILNQYLVPAICKAAAVLGMDKAIAEPVCR
                       LLETTLRSTHLPSRMGALHGVLYVLECDLLDDTAKQLIPTVSEYLLSNLRAIAHCVNL
                       HNQQHVLVMCAVAFTMMENYPLDVGTEFMAGIIQLCGVWVSASEDSTPSIIYSCVLRG
                       LERLLLSEQLSRVDGEALVKLSVDRVNMPSPHRAMAALGLMLTCMYTGKEKASPAARS
                       AHSDPQVPDSESIIVAMERVSVLFDRIRKGLPSEARVVARILPQFLDDFFPPQDIMNK
                       VIGEFLSNQQPYPQFMATVVYKVFQTLHATGQSSMVRDWVLLSLSNFTQRTPVAMAMW
                       SLSCFPVSASTSQWISALLPHVISRMGSSDVVDVNLFCLVAMDFYRHQIDEELDRRAF
                       QSVFETVASPGSPYFQLLACLQSIHQDKSL"
       exon            3223. .3320
                       /gene="Huntington's Disease Gene - Fugu homologue FrHD"
                       /number=1
       gene            3223. .25911
                       /gene="Huntington's Disease Gene - Fugu homologue FrHD"
       intron          3321. .3857
                       /gene="Huntington's Disease Gene - Fugu homologue FrHD"
                       /number=1
       exon            3858. .3941
                       /gene="Huntington's Disease Gene - Fugu homologue FrHD"
```



```
                       /number=2
       intron          3942. .4078
                       /gene="Huntington's Disease Gene - Fugu homologue FrHD"
                       /number=2
       repeat_region   4019. .4044
                       /note="MVR"
       exon            4079. .4199
                       /gene="Huntington's Disease Gene - Fugu homologue FrHD"
                       /number=3
       intron          4200. .4324
                       /gene="Huntington's Disease Gene - Fugu homologue FrHD"
                       /number=3
       exon            4325. .4384
                       /gene="Huntington's Disease Gene - Fugu homologue FrHD"
                       /number=4
       intron          4385. .4587
                       /gene="Huntington's Disease Gene - Fugu homologue FrHD"
                       /number=4
       repeat_unit     4436. .4451
                       /gene="Huntington's Disease Gene - Fugu homologue FrHD"
                       /note="PRPT"
       exon            4588. .4667
                       /gene="Huntington's Disease Gene - Fugu homologue FrHD"
                       /number=5
       intron          4668. .4770
                       /gene="Huntington's Disease Gene - Fugu homologue FrHD"
                       /number=5
       exon            4771. .4909
                       /gene="Huntington's Disease Gene - Fugu homologue FrHD"
                       /number=6
       intron          4910. .5139
                       /gene="Huntington's Disease Gene - Fugu homologue FrHD"
                       /number=6
       exon            5140. .5281
                       /gene="Huntington's Disease Gene - Fugu homologue FrHD"
                       /number=7
       intron          5282. .5362
                       /gene="Huntington's Disease Gene - Fugu homologue FrHD"
                       /number=7
       exon            5363. .5541
                       /gene="Huntington's Disease Gene - Fugu homologue FrHD"
                       /number=8
       intron          5542. .5656
                       /gene="Huntington's Disease Gene - Fugu homologue FrHD"
                       /number=8
       exon            5657. .5861
                       /gene="Huntington's Disease Gene - Fugu homologue FrHD"
                       /number=9
       intron          5862. .6083
                       /gene="Huntington's Disease Gene - Fugu homologue FrHD"
                       /number=9
```



```
       exon            6084. .6131
                       /gene="Huntington's Disease Gene - Fugu homologue FrHD"
                       /number=10
       intron          6132. .6625
                       /gene="Huntington's Disease Gene - Fugu homologue FrHD"
                       /number=10
       exon            6626. .6703
                       /gene="Huntington's Disease Gene - Fugu homologue FrHD"
                       /number=11
       intron          6704. .6780
                       /gene="Huntington's Disease Gene - Fugu homologue FrHD"
                       /number=11
       exon            6781. .7163
                       /gene="Huntington's Disease Gene - Fugu homologue FrHD"
                       /number=12
       intron          7164. .8437
                       /gene="Huntington's Disease Gene - Fugu homologue FrHD"
                       /number=12
       exon            8438. .8564
                       /gene="Huntington's Disease Gene - Fugu homologue FrHD"
                       /number=13
       intron          8565. .8636
                       /gene="Huntington's Disease Gene - Fugu homologue FrHD"
                       /number=13
       exon            8637. .8752
...
Note: remainder of annotations omitted.

  Query Match           5.7%;  Score 17;  DB 20;  Length 27470;
  Best Local Similarity 100.0%;  Pred. No. 1.6le+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2744 TAAATGTATAAATGTAT 2760
           |||||||||||||||||
Qy    103 TAAATGTATAAATGTAT 119


RESULT   44
LOCUS       CAU55001     30288 bp    DNA             VRL       02-AUG-1996
DEFINITION  Canine adenovirus type 1 E1A protein, small T-antigen, large
            T-antigen, protein IX, IVa2 protein, DNA polymerase, pTP, 52-55K
            protein, pIIIa, penton base protein, pVII protein, pV protein, Mu
            peptide precursor, pVI protein, viral protein II, endoproteinase,
            E2A-DBP, 100 kD protein, L4-22kD protein, L4-33 kD protein, pVIII,
            virion component IV, U-exon, E4 orf5, E4 orf4, E4 orf3, E4 orf2,
            and E4 orf1 genes, complete cds.
ACCESSION   U55001
NID         g1477654
KEYWORDS    .
SOURCE      Canine adenovirus type 1.
  ORGANISM  Canine adenovirus type 1
```

```
          Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
REFERENCE 1  (bases 1 to 199)
  AUTHORS   Sira,S., Abouhaidar,M.G., Liu,Y.C. and Campbell,J.B.
  TITLE     Multiple reiteration of a 40-bp nucleotide sequence in the inverted
            terminal repeat of the genome of a canine adenovirus
  JOURNAL   Virology 159 (1), 76-83 (1987)
  MEDLINE   87265479
REFERENCE 2  (bases 24577 to 27568)
  AUTHORS   Dragulev,B.P., Sira,S., Abouhaidar,M.G. and Campbell,J.B.
  TITLE     Sequence analysis of putative E3 and fiber genomic regions of two
            strains of canine adenovirus type 1
  JOURNAL   Virology 183 (1), 298-305 (1991)
  MEDLINE   91272490
REFERENCE 3  (bases 1 to 30288)
  AUTHORS   Campbell,J.B. and Zhao,Y.
  TITLE     DNA sequence and genomic organization of canine adenovirus type 1
  JOURNAL   Unpublished
REFERENCE 4  (bases 1 to 30288)
  AUTHORS   Campbell,J.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-APR-1996) James B. Campbell, University of Toronto,
            Microbiology, 150 College Street, Toronto, Ontario, Canada, M5S 1A8
COMMENT     Nomenclature of proteins has largely followed accepted terminology
            of the human adenoviruses.
FEATURES             Location/Qualifiers
     source          1. .30288
                     /organism="Canine adenovirus type 1"
                     /strain="CLL"
                     /db_xref="taxon:10512"
                     /lab_host="dog kidney"
     repeat_unit     1. .199
                     /note="left ITR"
     repeat_unit     54. .93
                     /note="40-base pair repeat not found in other strains of
                     canine adenovirus type 1; field isolate: Spibey et al.,
                     Virus Research 14 (1989) 241-256."
                     /citation=[3]
                     /evidence=experimental
     CDS             join(493. .1003,1101. .1282)
                     /function="transcription activation"
                     /note="similar to human Ad E1A 298R protein; splice site
                     by comparison with canine adenovirus type 2: Shibata et
                     al., Virology 172 (1989) 460-467."
                     /codon_start=1
                     /product="E1A protein"
                     /db_xref="PID:g1477655"
                     /translation="MKLTLPEAPRCLHEYVSQLLEDWQPECLSCEYSRGGSSPPSLHD
                     LFDLELENSRSPSPLLCDWCAEADSESTISTETDVGPTLWTPPVSPLPSYSTSPASIP
                     EDMLLCLEEMPTFDDGDEVRSATTSFPEHWENNFDPNVGSFPGCLRCAYYQEQGENSIC
                     GLCYLKALAEEPVDADAGEDDEVIPVSAKPGSRKRSAVTSRGSVESSKRPCLPEPEQT
                     EPLDLSLKPRPQ"
```



```
     CDS             1406. .1915
                     /function="inhibition of apoptosis"
                     /note="E1B 19kD protein"
                     /codon_start=1
                     /product="small T-antigen"
                     /db_xref="PID:g1477656"
                     /translation="MDPLKICENYLTFRSIIKGSTFSPGVFRRWRFHALADVVGNIVE
                     REEGRFWEIVPETHTLWALFRGGFTVAPFTEILTSLQLENRGRQLAFLAFLSFLLRNW
                     PSDSVVSEDARLLDLVCAPAWSRIQIWSQAARLINDLPESVPEGQGSVVEEECGEEHLA
                     RDSDDPLFD"
     CDS             1768. .3102
                     /function="anti-oncogene"
                     /note="E1B 55kD protein"
                     /codon_start=1
                     /product="large T-antigen"
                     /db_xref="PID:g1477657"
                     /translation="MEQDSDLESGRATNQRPPRVRVRGAGVRGRGRVRRRALSEGQRR
                     SPFRLDDLQLPDSLYVTRALQRDHALEMPRGQVDFSLIEAEERRAGPTDEWYFESVRT
                     YRAPGDDLQTLIKNYAKISLECGAVYEINSKIVVTGACYIIGNCAVLRANLPVGTAM
                     FEVLNVDVIPSIGFMERIVFSNILFDCRSTTAVVCCISERNTLFHNCVFSGPHMLCLD
                     IRAGAEVRGCHFVGAVCALRSKGLYSVRVRNSIFEKCAFGVVSGSKASISHSMFKDCA
                     CCIMFGGQGTIAHSHFMATTCTDTPMNLQLCTCEGNGSHVVPLGNIHFASNREAPWPT
                     FNANVLVRVRLYMGRRRGVFHPKQSTFSMCVIAAPRGVVQRIYLFSVYDATCAILQLG
                     EAGDAATERLCTRGMRHNTPSLRAAAYVTDTRIDREINSQDTAEFFSSDEDNL"
     CDS             3174. .3485
                     /function="virion stability"
                     /codon_start=1
                     /product="protein IX"
                     /db_xref="PID:g1477658"
                     /translation="MDPQQKGIVNTCFLTTRIPSWAGARQNVTGSDLGGKPVPSDVLE
                     SGRPLAAPRVRTLYEEQQLNMLTVNVILDDLKTQVAAMQNSVTAIQEELKDLKQRVAA
                     R"
     CDS             complement(join(3502. .4829,5108. .5120))
                     /function="virion assembly"
                     /note="splice site tentative, by comparison with human
                     Ad2"
                     /codon_start=1
                     /product="IVa2 protein"
                     /db_xref="PID:g1477659"
                     /translation="MEEKAGLRHLQNQQNEPSQDPHQCADPCPALHSDRNHLNKEAEA
                     MEGQNPTCSRHESAYPIQSQVSKSKKQRNTVDGAAVDDLKNLWDRLQTLQQSLTEMPT
                     AEGLKPLKNFTSFEELLSMGGDSLLNDLLDIQDSITQCCIRASKYLNKEGNCTSLNYY
                     KQPPIAAVYGPTGCGKSQLLRNLMSAQLIVPTPETVFFITPQVDMIPPQEIAAWETQI
                     CEGNFLAGPENTVVPQSGSIMPRFIQMSYAQLTKDENYDVTSPNNIFARAASKGPIAI
                     VMDECMEDLGGNRGIAKFFHAPPSKLLDRFPKCTGYSVIVVLHNMMPRRDQGGNIANL
                     KIQAKVHMISPKVHPSQLRRFINIITKGQPTAISLLLKDIFWTHRLNTNPDWIVTNTE
                     PIDNCMHWNTLSPQEGLIPHYLNIQGKLVQALERIHRTLTDRQRWTRYYHSKKK"
     CDS             complement(join(4593. .8033,12508. .12516))
                     /function="DNA synthesis"
                     /note="splice site tentative, by comparison with human
                     Ad2"
```



```
                     /codon_start=1
                     /product="DNA polymerase"
                     /db_xref="PID:g1477660"
                     /translation="MSLVQSHGTSGLFTEPPNSINQQESSGPSLPAQDATQASASSAR
                     AGATPAINSTKRKYRGAVVAQRATLSISAILDNGQCVQIKYHSNLASALTNLCNTNLY
                     DLPACLNKPITVHNLPTLIEEAAASYSLIYYFQRGTVRKVEFRAEIPLLSFPLKFLVK
                     HGKVFLIKDISFNQRCEPCGSPFKVTHTCTLRRRDFYFHHVASHSGDWWEKISFSPIG
                     APANTERLFIVYDVETTWHGKFGKQLVPFMLVPQLLGLDEHLVNAAKNLATTONWDTW
                     NSNEQTALYYCITPEKRAIGVKFKTFRDTLQQHFANNLWSHVLCQNPKLMEEAAALGL
                     ENPEDITANQLKKFKLQGTPRFIEVYVVGHNITGFDEILLAAQVVSTRAEIPPVFDIS
                     RNFMPRAGRLLFNDITTSLPNPSYVPSKDYRHWEQGQVLASDLKTQYIKFMVRDTFSL
                     THTSLKNAAKAYSLTVSKGCCPYQAVNEFTMLGSYQQDADGFPDLKYWKDQEEYCFNK
                     DLWKKEKKGAYDIIQQTLDYCALDVQVTAQLVNKLIESTQIFIKNSVNLPETYFNVPQ
                     RPTISSNSHAIFKQILYRAEKPNAPHLNTIIMAPSNEMYETVYRLSIRGGRCYPTYIGV
                     LQEPVFVYDICGMYASALTHPFPAGSPLNPYERAVAIKAYEHKMQEHKTISYFDEDLL
                     PGIFTIDADPPAEEFLDVLPPFCSRKGGRLCWTNEPLRGEVTTSIDVITLHNRGWKVT
                     LIPVTRTTVFPEWKCLARETVQLHNAKEADESKHQTMRSIAKLLSNALYGSFATKL
                     DNKKTVFSDQIESNIAKEIASGAYVVKSSSTIETDNLCAEIMPEFVVAYPPVNFDVHR
                     LAPPSYSEEYPTENPHAEGPFMQNFNMTSYRTKPIMPIDAEDDDFCLHTLEKSTPLIT
                     NNRYASQIASFVLAWTRAFVSEWSQFLYENDAGIPLEKRILKSVYGDTDSLFTTMEGY
                     RLMEEKGKRRLKKNGGNLVFDPNNPELTWLVECETKCEKCGADAYSSESVYLAPKLYA
                     LKDTTCPECQTVGKGKLRAKGHATSTLSYDVLKACYYADLQQGSDIFKTSRMSLRRTL
                     TSVQTHVQPPTVTETTLTRKLRPWKDKTLHALDMNRLIPYSRKYPNPRNNETTWMELQ
                     WMT"
     CDS             complement(join(7889. .9706,12508. .12516))
                     /function="DNA replication"
                     /note="terminal protein precursor; splice site tentative,
                     by comparison with human Ad2"
                     /codon_start=1
                     /product="pTP"
                     /db_xref="PID:g1477661"
                     /translation="MSLNALDCARLTGQTPYTVEVFRPIRNIFNRVRETTRASTTSVG
                     LAWMSKYIYQYHRLMLMNLSPREPATEGWPLFLYPPPHLLVGYQYLVRTCNDYVFDTR
                     SYSRLKYTEIHLPLQQKLNWTVMANCSYTINTGAYHRFIDFENFEETLAQVQQAVLAE
                     RVVADLALIRPMRGYGTTNMAGDRQVPVEGLLQDHYKNLSQCQNHAWGLADRMRIQNA
                     GNKDIVILTTIRKLETAPFNFLVSPRNPTYISLSPCDCLWDLAFMQKFTDPSLSQFQT
                     IQSLPSQSVTKSIISALSLPQPAPCTPLSGGAFELRPRENGRAVTEEMKRRRGEMIER
                     FIDRLPMRRRRRRRAPPPPPHSEELSEPEVEAFPPASPPRRSPEEEVRDTIVEAIRLLQ
                     EELTVSARNEQFFNFAINFYEVIQRLEMLGNINELTIRRWVMTFPVAEHVATTLNYLH
                     HNLRLYPPCSRWVDLELAQVVMRARDHEGQVVYSRVWNEMGENAPSQLMARVSGDLAA
                     TVERAGLGELEEEEMEQFMADIAYHENSGDVSEILRQVAINDTEVDSMELSFRFKVTG
                     PVVFSQNRQIQSINRRVVALASQLRMQHRPLPAQHEQVQLPP"
     CDS             9756. .10925
                     /function="virion assembly"
                     /codon_start=1
                     /product="52-55 KD protein"
                     /db_xref="PID:g1477662"
                     /translation="MHPVLRQMKPLPASVVKLEAESEGLARLQGGAAPEVHPRVQMKQ
                     EAAEAYIPASNVFRDNEGEEAEGLRHLKYESGRMLRKDHPGKRVLDERDPEKSPENGI
                     SAAEAHLKSADLVTAYEHTVKAEVNFQTTFNNNVRTLIAREEVVIGLMHLWDFVEAPL
                     ENPVSKALTAQLFLIVQHCRDEGVLRESLLNIAEPESRWLVDLLNLLQTIVVQERSLA
```



```
                     VGEKVAAINYSVITLSKHYARKIFNSVFVPIDKEAKINTFTMRTVVQILVLSDDLGMY
                     RNERIERAVSGARQRELNDRELMYRLREALTSNGLSEIEGGEDSARVRSKENWGAGAG
                     VGAASARYPHLLDYEEKENPDGSVFSQQHERGTQSHENGGHAEPAYSRRQLGRFY"
     CDS             10798. .12489
                     /function="structural protein"
                     /note="IIIa protein precursor"
                     /codon_start=1
                     /product="pIIIa"
                     /db_xref="PID:g1477663"
                     /translation="MRKRRRTLTAPSHFNSMSAALNPMKMAAMQSQPTVDDSWAASISR
                     IMSLTAGDRHKFSSQPFANRLDAILEAVVPSRKDPTHEKVLTIVNALIENGAIRRDEG
                     AGVYDALLHRVSKYNSINTQSNLERLAGDVREAVAQQVRIAAGNLGSLTALNSFLARL
                     PANVERGQDNYTGFLSALKLLVSEVPNTEVYQSGPHYFLQSSRNGTQTVNLTNAFENL
                     KPLWGVKAPTMERLSISALLTPNTRLLLLLVSPFTDSVSISRDSYLGYLTLTLREALG
                     RNHLDERTLEEVTEVSKAMGNENINNLQATLNFLLTNRQKKIPKDYSLTPEEERIVRF
                     IQQAVSLFMMQENLSPTEALDVTAANMEPSFTANNRDFINKLMDYFHRAAAIAPDYFL
                     GAVKNPRWLPPEGFFTGVFDPFERDNYIWDGPDSSSLDLTRQDAMRPLEEKFIDDDQRT
                     ESRSVSRVPTPASSRRSSVAMASDSLIRPMNNDKNSLREIEVLADKLARWKTYKRESE
                     EARESLPVVVRPKKYSSAISSDESDDGMSKPDKPLKFPEGSGNPFAHLRPKLGRCL"
     CDS             12540. .13973
                     /function="structural protein; virion component III"
                     /codon_start=1
                     /product="penton base protein"
                     /db_xref="PID:g1477664"
                     /translation="MEFPSSPPPSYETVMAQVPSILAPLVPPRYKGATEGRNSIRYSQ
                     LPPLFTTKLYLIDNKSSDIQALNYQNDHSNFLTTVVQNANYTPMEASTQSIQLDERS
                     RWGGDFRSILHMNMPNVTETMFSNSFKAYLPATADASGKVLTYEWYTLTIPEGNYSEV
                     MLLDLLNNAVVENYLAHGRQHNVKEEDIGLKFDTRNFYLGFDPETELVMPGFYTNEAF
                     HPDIILSPGCAVDFTHSRLNNFLGIRKRLPYQEGFIIKWEDLQGGNIPALLDLEIYNP
                     DTPGDNITPLLQDSKARSYHVGEDPSAGSTFTSYRSWFLAYNYGPVDGIRSKTVLVAP
                     DITCGVEQIYWSLPDHAVDPVTFTSSHNPSSYPVVGTELLPLLPRSFYNGSSVYSQLL
                     QESTAQTHVFNRFPENAILKRPPAPTIISISENVPALSNHGTLPLKNNIPGVQRVTIT
                     DARRRVCPYVYKSLGVVVPRVLSSKTF"
     CDS             13891. .14403
...
Note: remainder of annotations omitted.

  Query Match           5.7%;  Score 17;  DB 33;  Length 30288;
  Best Local Similarity 100.0%;  Pred. No. 1.61e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   6379 ACAGAGTTTTTTATAAA 6395
         |||||||||||||||||
Qy     90 ACAGAGTTTTTTATAAA 106


RESULT   45
LOCUS       CAV1GENOM   30536 bp    DNA             VRL       08-SEP-1997
DEFINITION  Canine adenovirus type 1 complete genome.
ACCESSION   Y07760
NID         g1526575
```

```
KEYWORDS    inverted repeat; open reading frame.
SOURCE      Canine adenovirus type 1.
  ORGANISM  Canine adenovirus type 1
            Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
REFERENCE   1  (bases 1 to 30536)
  AUTHORS   Morrison,M.D., Onions,D.E. and Nicolson,L.
  TITLE     Complete DNA sequence of canine adenovirus type 1
  JOURNAL   J. Gen. Virol. 78 (Pt 4), 873-878 (1997)
  MEDLINE   97275900
REFERENCE   2  (bases 1 to 30536)
  AUTHORS   Morrison,M.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-SEP-1996) M.D. Morrison, University of Glasgow Vet
            School, Department of Veterinary Pathology, Garscube Estate,
            Bearsden Road, Glasgow G61 1QH, UK
COMMENT     Related sequences; M16830, M16831, M60937, S38212, S38238 and
            U55001.
FEATURES             Location/Qualifiers
     source          1. .30536
                     /organism="Canine adenovirus type 1"
                     /isolate="RI261"
                     /specific_host="dog"
                     /db_xref="taxon:10512"
     repeat_region   1. .161
                     /rpt_type=INVERTED
     gene            454. .978
                     /gene="orf1"
     CDS             454. .978
                     /gene="orf1"
                     /note="putative"
                     /codon_start=1
                     /db_xref="PID:e264811"
                     /db_xref="PID:g1526576"
                     /db_xref="SPTREMBL:Q96676"
                     /translation="MKLTLEPAPRCLHEYVSQLLEDWQPECLSCEYSHGGSSPPSLHD
                     LFDLELENSRSPSPLLCDWCAEADSESTISTETDVGFTLNTPPVSPLPSYSTSPASIP
                     EDMLLCLEEMPTFDDGDEVRSATTSFEHWENNFDPNVGSFPGCLRCAYYQEQGENSIC
                     GLCYLKALAEGKIF"
     gene            1046. .1243
                     /gene="orf2"
     CDS             1046. .1243
                     /gene="orf2"
                     /note="putative"
                     /codon_start=1
                     /db_xref="PID:e264812"
                     /db_xref="PID:g1526577"
                     /db_xref="SPTREMBL:Q96677"
                     /translation="MPVRAEPVDADAGEDDEVIFVSAKPGSRKRSAVTSRGSVESSKR
                     PCLPEPEQTEPLDLSLKPRPQ"
     polyA_signal    1239. .1244
                     /note="putative"
```



```
     gene            1367. .1876
                     /gene="orf3"
     CDS             1367. .1876
                     /gene="orf3"
                     /note="putative"
                     /codon_start=1
                     /db_xref="PID:e264813"
                     /db_xref="PID:g1526578"
                     /db_xref="SPTREMBL:Q96678"
                     /translation="MDPLKICENYLTFRSIIKGSTPSPGVFRRWRFHALADVVGNIVE
                     REEGRFWEIVPETHTLWALPRGGFTVAPFTEILTSLQLENRGRQLAFLAFLSFLLRNW
                     PSDSVVSEDARLDLVCAPAWSRIQIWSQAARLINDLPESVFEGQGSVVEEECGEEHLA
                     RDSDDPFFD"
     gene            1729. .3063
                     /gene="orf4"
     CDS             1729. .3063
                     /gene="orf4"
                     /note="putative"
                     /codon_start=1
                     /db_xref="PID:e264814"
                     /db_xref="PID:g1526579"
                     /db_xref="SPTREMBL:Q96679"
                     /translation="MEQDSDLESGRATNQRPPRVRVRGAGVRGRGRVRRRALSEGQRR
                     SLPRLDDLQLPDSLYVTRALQRDHALEMPRGQVDFSLIEAEERRAGPTDEWYFESVKT
                     YRAKPGDDLQTLIKNYAKISLECGAVYEINSKIVVTGACYIIGNCAVLRANLPVGTAM
                     FEVLNVDVIPSIGFMERIVFSNILFDCRSTTAVVCCISERNTLFHNCVFSGPHMLCLD
                     IRAGAEVRGCHFVGAVCALRSKGLYSVRVRNSIFEKCAFGVVSGSKASISHSMFKDCA
                     CCIMLGGGQTIAHSHFMATTCTDTPMNLQLCTCEGNGSHVVPLGNIHFASNREAPWPT
                     FNANWLVRVRLYMGRRRGVFHPKQSTFSMCVIAAPRGVVQRIILPSVTDATCAIQLG
                     EAGDAATERLCTCGMRHNTPSLRAAYVTDTRIDREINSQDTAEFFSSDEDNL"
     gene            3135. .3446
                     /gene="orf5"
     CDS             3135. .3446
                     /gene="orf5"
                     /note="putative"
                     /codon_start=1
                     /db_xref="PID:e264815"
                     /db_xref="PID:g1526580"
                     /db_xref="SWISS-PROT:Q65944"
                     /translation="MDPQQKGIVNTCPLTTRIPSWAGARQNVTGSDLGGKPVPSDVLE
                     SGRPLAAPRVRTLYEEQQLNMLTVNVILDDLKTQVAAMQNSVTAIQEELKDLKQRVAA
                     R"
     polyA_signal    3454. .3459
                     /note="putative"
     polyA_signal    complement(3462. .3467)
     gene            complement(3463. .4971)
                     /gene="orf6"
     CDS             complement(3463. .4971)
                     /gene="orf6"
                     /note="putative"
                     /codon_start=1
```



```
                     /db_xref="PID:e264816"
                     /db_xref="PID:g1526581"
                     /db_xref="SPTREMBL:Q96680"
                     /translation="MWGGCIQFRISVPGAKIICFKGHHMSRMSIRGQGQVTGKRSCHL
                     HPLLRCAKGLLLCRPAAGLRHLQNQQNEPSQDPHQCADSCPALHSDRNHLNKEAEAME
                     GQNPTCSRHESAYPIQSQVSKSKKQRNYVDGAAVDDLKNLWDRLQTLQQSLTEMPYAE
                     GLKPLKNFTSFEELLSMGGDSLLNDLLDIQDSITQCCIRASKYLNKEGNCTSLNYYKQ
                     PFIAAVYGPTGCGKSQLLRNLMSAQLIVPTPETVFPITPQVDMIPPQEIAAWETQICE
                     GNFLAGPENTVVPQSGSSIMPRFIQMSYAQLTKDENYDVTSPNNIFANAASKGPIAIVM
                     DECMEDLGGNRGIAKFFHAFPSKLLDRFPKCTGYSVIVVLKNMNPRRDQGGNIANLKI
                     QAKVHMISPKVHPSQLNRPINIYTKGQPTAISLLLKDIFNYHRLNTNFDWIVYNTEPI
                     DNCMHWMYLSPQEGLIPMYLNIQGKLYQALERIHRTLTDRQRWTRYYHSKKK"
     gene            complement(4554. .7487)
                     /gene="orf7"
     CDS             complement(4554. .7487)
                     /gene="orf7"
                     /note="putative"
                     /codon_start=1
                     /db_xref="PID:e264817"
                     /db_xref="PID:g1785626"
                     /db_xref="SPTREMBL:Q96681"
                     /translation="MQKCEFCGSPFKVTHTCTLRRRDFYFHHVASHSADWWEKISFSP
                     IGAPANTERLPIVYDVETYTWHGKFGKQLVPFMLVFQLLGDEHLVNAAKNLATTQNWD
                     TWNSNEQTALYYCITPEKRAIGVKFKTFRDTLQQHFANNLWSHVLCQNPKLMEEAAAL
                     GLENPEDITANQLKKFKLQGTPRFIEVYVVGHNITGFDEILLAAQVVSTRAEIPPVFD
                     ISRNFMPRAGRLLFNDITYSLPNPSYVPSKDYRHWEQGQVLASDLKTQYIKFMVRDTF
                     SLTHTSLKNAAKAYSLTVSKGCCPYQAVNEFTMLGSYQQDADGFPDLKYWKDQEEYCF
                     NKDLWKKEKKGAYDIIQQTLDYCALDVQVTAQLVNKLIESTQIFIRNSVNLPETYFNV
                     FQRPTISSNSHAIFKQILYRAEKPNAPHLNTIIMAPSNEMYEYVRLSIRGGRCYPTYI
                     GVLQEPVFVYDICGMYASALTHPFPAGSPLNPYERAVAIKAYEHKMQEHKTISYPDED
                     LLDGIFTIDADPPAEEFLDVLPPFCSRKGGRLCWTNEPLRGEVTTSIDVITLHNRGWK
                     VTLIPDTRTTVPFEWKCLAREYVQLNINAKEKADKSKNQTMRSIAKLLSNALYGSFAT
                     KLDNKKTVFSDQIESNIAKEIASGAYVVKSSSYIETDNLCAEIMPEFVVAYPPVNFDV
                     HRLAPPSYSEEYPTENPHAEGPFMQNFNMTSYRYKPIMPIDAEDDDPCLHTLEKSTPL
                     ITNNRYASQIASFVLAWTRAFVSEWSQPLYENDAGIPLEKRILKSVYGDTDSLFTTME
                     GYRLMEEKGKRRLKKNGGNLVFDPNNPELTWLVECETKCEKCGADAYSSESVYLAPKL
                     YALKDTTCPECQYVGKGKLRAKGHATSTLSYDVLRACYYADLQQGSDIFKTSRMSLRR
                     TLTSVQTHVQPFTVTETTLTRKLRPWKDKTLHALDMNRLIPYSRKYPNPRNNETTWME
                     LQWMT"
     gene            complement(7850. .10021)
                     /gene="orf8"
     CDS             complement(7850. .10021)
                     /gene="orf8"
                     /note="putative"
                     /codon_start=1
                     /db_xref="PID:e264818"
                     /db_xref="PID:g1526583"
                     /db_xref="SPTREMBL:Q96682"
                     /translation="MPFSGLFSKSRSSSTRLEGWSLRNILPDSYFKCRNPSASSPSLS
                     LNTLLAGMYASAASCFICTRGCTSGAAPPCNRAKPSLSASSLTTEAGRGFICRKTGCI
                     CNGANESAIDSLFVLQNALDCARLTGQTPYTVEVFRPIRNIFNRVREYTRASTTSVGL
```



```
                     AWMSKYIYQYHRLMLMNLSPREPATEGWPLFLYPPPHLLVGYQYLVRTCNDYVFDTRS
                     YSRLKYTEIHLPLQQKLNWTVMANCSTTINTGAYHRFIDFENFEETLAQVQQAVLAER
                     VVADLALIRPMRGYGTTNMAGDRQVPVEGLLQDHYKNLSQCQNHAWGLADRMRIQNAG
                     NKDIVILTTIRKLKTAFFNFLVSPRNPHTILSLPCDCLWLDAFMQKFTDPTLSQFQTI
                     QSLPSQSVTKSIISALSLPGPAPCTPLSGGAFELRPRENGRAVTEEMRRRRGEMIERF
                     IDRLPMRRRRRRAPPPPPMSEELSEPEVEAFPPASPPRRSPEEEVRDTIVEAIRLLQE
                     ELTVSARNEQFFPFAINFTEVIQRLEMLGNIMELTIRRWVMYFFVAEHVATTLNYLHH
                     NLRLYPCSRWVDLELAQVVMRARDHEGQVVYSRVWNEMGENAFSQLMARVSGDLAAT
                     VERAGLGELEEEEMEQFMADIAYHENSGDVSEILRQVAINDTEVDSMELSFRFKVTGP
                     VVFSQNRQIQSINRRVVALASQLRMQHRPLPAQHEQVQLPP"
     gene            9501. .10886
                     /gene="orf9"
     CDS             9501. .10886
                     /gene="orf9"
                     /note="putative"
                     /codon_start=1
                     /db_xref="PID:e264819"
                     /db_xref="PID:g1526584"
                     /db_xref="SPTREMBL:Q96683"
                     /translation="MRRWYWYMYLDIQASPTEVVEALVYSRTRLKMPLIGLNTSTVYG
                     VCPVRRAQSSAFCKTNKESMALSFAPLQMHPVLRQMKPLPASVVKLEAESEGLARLQG
                     GAAPEVHPRVQMKQEAAEAYIPASNVFRDNEGEEAEGLRHLKYESGRMLRKDHPSKRV
                     LDERDFEKSPENGISAAEAHLKSADLVTAYEHTVKAEVNFQTTFNNNVRTLIAREEVV
                     IGLMHLWDFVEAFLENPVSKALTAQLFLIVQHCRDEGVLRESLLNIAEPESRWLVDLL
                     NLLQTIVVQERSLAVGEKVAAINYSVITLSKHYARKIFNSVFVPIDKEAKINTFYMRT
                     VVQILVLSDDLGMYRNERIERAVSGARQRELNDRELMYRLREALTSNGLSEIEGGEDS
                     ARVRSKENWGAGAGVGAASARYPHLLDYEEKENPDGSVSPQQHERGTQSHENGGHAEP
                     AYSRRQLGRFT"
     gene            10759. .12450
                     /gene="orf10"
     CDS             10759. .12450
                     /gene="orf10"
                     /note="putative"
                     /codon_start=1
                     /db_xref="PID:e264820"
                     /db_xref="PID:g1526585"
                     /db_xref="SPTREMBL:Q96684"
                     /translation="MRKRRTLTAPSHFNSMSAALNPMKMAAMQSQPTVDDSWAASISR
                     IMSLTAGDRHKFSSQPFANRLDAILEAVVPSRKDPTHEKVLTIVNALIENGAIRRDEG
                     AGVYDALLHRVSKYNSINTQSNLERLAGDVREAVAQQVRIAAGNLGSLTALNSFLARL
                     PANVERGQDNYTGFLSALKLLVSEVPNTEVYQSGPHYFLQSSRNGTQTVNLTNAFENL
                     KPLWGVKAPTMERLSISALLTPNTRLLLLLVSPPTDSVSISRDSYLGYLLTLYREALG
                     RNHLDERTLEEVTEVSKAMGNENINNLQATLNFLLTNRQKKIPKDYSLTPEEERIVRF
```

Note: remainder of annotations omitted.

```
  Query Match           5.7%;  Score 17;  DB 33;  Length 30536;
  Best Local Similarity 100.0%;  Pred. No. 1.61e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   6340 ACAGAGTTTTTTATAAA 6356
```

```
          !!!!!!!!!!!!!!!!
Qy     90 ACAGAGTTTTTTATAAA 106


RESULT   46
LOCUS       CET18D3     32601 bp    DNA             INV       22-SEP-1998
DEFINITION  Caenorhabditis elegans cosmid T18D3, complete sequence.
ACCESSION   Z68119
NID         g1082143
KEYWORDS    HTG; Cobalt uptake like; Myosin heavy chain; Steroid hormone like;
            Transfer RNA.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 32601)
  AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sprat,J. and Wohldman,P.
  TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
  JOURNAL   Nature 368 (6466), 32-38 (1994)
  MEDLINE   94150718
REFERENCE   2  (bases 1 to 32601)
  AUTHORS   McMurray,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-NOV-1996) Nematode Sequencing Project, Sanger Centre,
            Hinxton, Cambridge CB10 1RQ, England and Department of Genetics,
            Washington University, St. Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     Coding sequences below are predicted from computer analysis, using
            predictions from Genefinder (P. Green, U. Washington), and other
            available information.
            For a graphical representation of this sequence and its analysis
            see:-
            http://webace.sanger.ac.uk/cgi-
            bin/display?db=wormace&class=Sequence &object=T18D3
            Current sequence finishing criteria for the C. elegans genome
            sequencing consortium are that all bases are either sequenced
            unambiguously on both strands, or on a single strand with both a
            dye primer and dye terminator reaction, from distinct subclones.
            Exceptions are indicated by an explicit note.
            IMPORTANT:  This sequence is NOT necessarily the entire insert of
```



```
            the specified clone.  It may be shorter because we only sequence
            overlapping sections once, or longer because we arrange for a small
            overlap between neighbouring submissions.
            IMPORTANT: This sequence is not the entire insert of clone T18D3.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we arrange for a small overlap between
            neighbouring submissions.
            The true left end of clone T18D3 is at 1 in this sequence. The true
            right end of clone T18D3 is at 4855 in
            sequence Z68117.
            The true left end of clone F45E6 is at 32502 in this sequence. The
            true right end of clone T24D3 is at 19338 in this sequence. The
            start of this sequence (1. .100) overlaps with the end of sequence
            Z68121.
            The end of this sequence (32502. .32601) overlaps with the start of
            sequence Z68117.
FEATURES              Location/Qualifiers
     source           1. .32601
                      /organism="Caenorhabditis elegans"
                      /db_xref="taxon:6239"
                      /chromosome="X"
                      /clone="T18D3"
     gene             4964. .5341
                      /gene="T18D3.7"
     CDS              join(4964. .5176,5246. .5341)
                      /gene="T18D3.7"
                      /note="similar to TSC-22 / dip / bun family"
                      /codon_start=1
                      /db_xref="PID:g1082149"
                      /db_xref="SWISS-PROT:Q22544"
                      /translation="MTSSPSSVALLSGDPTKMEDSAKLATATSNTVVAIDNKIVQAMD
                      LVRTHLTFAVREEVETLRTTITDLEARLNALREENRLLRENVSPEVLAFITANKQLCE
                      "
     gene             12018. .12918
                      /gene="T18D3.6"
     CDS              join(12018. .12758,12844. .12918)
                      /gene="T18D3.6"
                      /codon_start=1
                      /db_xref="PID:g1082148"
                      /db_xref="SPTREMBL:Q22543"
                      /translation="MEEPHPAPRNDRDAEGEEARHALRMERQLAIREGRYLPPLPQLP
                      PLIQLIPVHRAQANEAVDRAIEFQLNQRRQFLLQHLRPEHFQPGGFFDPQPEGNTHPL
                      SPQHSDEELDDAEEEIERQRRIRQNWRNTMNQRVDPIPQEIVDRLLREYDNWARAALN
                      NNNGNFRNYNAMLSPSPSPSPNASPAPSDDEDPMEHVNSDDEDVEMLNAQMMEMRPNQ
                      PRPRKRHNSQGPLPIKRRRVTFKEDSSDEAINSFDEDSSDCNDNVPVNRKREE"
     gene             16866. .19738
                      /gene="T18D3.1"
     CDS              join(16866. .17055,17104. .17375,18030. .18334,18390. .18489,
                      18544. .18704,19195. .19738)
                      /gene="T18D3.1"
                      /note="cDNA EST EMBL:Z14951 comes from this gene; cDNA EST
```



```
                      EMBL:T00215 comes from this gene; cDNA EST EMBL:D71583
                      comes from this gene; cDNA EST EMBL:D74238 comes from this
                      gene; cDNA EST yk376f7.3 comes from this gene; cDNA EST
                      yk376f7.5 comes from this gene"
                      /codon_start=1
                      /db_xref="PID:g1082144"
                      /db_xref="SPTREMBL:Q22540"
                      /translation="MPVRIDLSSNASVRLLSVNDENTACAYLANLYIPIFSLDSGKVI
                      RRYEVPNQIAGIHPSFNKKSVIYAVDDSPGLHTPDPRTDSKKHRMNPESENHNVVCT
                      ALSSNSTTIAVASSEPGGGILDVRGSRRREHHHHVVSLYDLRYPNDPITSYFKRLKSS
                      VTSMEFYLNGNDLILGDESGSIKSFDCQLRDEDRAVRWERWATGPISKVGVINKRIVY
                      GISRQSQASVFVDSKGKLEVLYGSVTESADDPVVGTSKAYYKDPEWCIGLIKGVNDDI
                      PLVGVHGVKKKKIISMTAMNQDGVRQEKPMLSYKGHKGEVKCMAATKDLLLTGGEDGK
                      VIVQIANFKNPKFGDDNINHHRLTLNRGNLAELKDKGLEKDDEAISESDDSESDDGSK
                      PSTSASASVSSSSKHKKDKHDIPSTSSAASTSSSSSSKHETKEERRARKEKERERAEK
                      KAKKEKDGHREEEKPKKKDKKHETEEERAERRKEKKKEKERQERKEKEHDEQKKRHRD
                      NEEGAGSSSKRSKAP"
     tRNA             complement(20277. .20348)
                      /gene="T18D3.t1"
                      /note="AGC Ala A-tRNA; predicted using tRNAscan-SE-1.11;
                      preliminary prediction; similar to tRNA-Ala"
     gene             complement(20277. .20348)
                      /gene="T18D3.t1"
     gene             complement(21170. .23740)
                      /gene="odr-7"
     CDS              complement(join(21170. .21333,21381. .21523,22030. .22196,
                      22242. .22329,22379. .22856,23407. .23740))
                      /gene="odr-7"
                      /note="Similarity to steroid hormone receptors but lacks
                      the appropriate prosite."
                      /codon_start=1
                      /db_xref="PID:g1082145"
                      /db_xref="SWISS-PROT:P41933"
                      /translation="MIVPDTEGLLIYSYGLMYGSYCMACQMLIPHFQCIPGIFPNFRI
                      STELIKTMTDKLEQPNNNVPQQPWGPFPPAPGGRPSGEQTDGNPGEFDNDAAHQQTAP
                      FMTHFFPRIGLQFPDFTEYQRFNGFQRNAFFPNFPGSQFTGQAFAQSFPLHNSMTTMD
                      GFNLTHAPHPFSTNTNSTKPKDIENTVQSTIKHSSENIQDKPPVLSVETPVKYDSELK
                      FDANVDFTAVPKQESSDDSTLKWLKKSDQQLQQPQQFTFPPPLLAEKSPEQPRMREDV
                      LPPHPQFYPAPLDMGTNFKQEMRTPPIDGHIDYRRFDASGKRMEPQPPGALHDCQVCL
                      STHANGLHFGARTCAACAAFFRRTISDDKRTVCKRNQRCNNASRDGTGYRKICRSCRM
                      KRCLEIGMLPENVQHKRNRRDSGSPPRKTPFDTFFNGFYPSFQPSGSAAQPITVSSSE
                      SPRHTTN"
     gene             25919. .26478
                      /gene="T18D3.5"
     CDS              join(25919. .26013,26129. .26237,26287. .26478)
                      /gene="T18D3.5"
                      /codon_start=1
                      /db_xref="PID:g1082147"
                      /db_xref="SPTREMBL:Q22542"
                      /translation="MRKLCNCLTCLLIIIILSITLANAQDEQHSPRENEASRDNSDVW
                      LVVILCLLFCFALRHMNDFTVGESVTLVAEGEPSHRFSRKQTNKHLERVEKDFNDMFQ
```



```
                      KLCEKYERDLDSDGCCDYYEENKPSEKIP"
     gene             complement(26870. .29276)
                      /gene="T18D3.3"
     CDS              complement(join(26870. .27043,27088. .27194,27506. .27611,
                      27693. .27876,28147. .28300,28348. .28557,28625. .28737,
                      29176. .29276))
                      /gene="T18D3.3"
                      /note="Similarity to Yeast cobalt uptake protein cot1 (PIR
                      Acc. No. S31302)"
                      /codon_start=1
                      /db_xref="PID:g1082146"
                      /db_xref="SPTREMBL:Q22541"
                      /translation="MYDVFNTFTFHPCSCSLWGCHNKTSTAQNPQCTRVPHIYNTDEG
                      RKTTVTYDNFHCHDEADSTDSHDSNRRATRILWLTVVLCLFFMVCEVIGGVLAGSLAI
                      VTDAAHLLTDFASVLISLFSLYIARRPPSQKMSFGFHRAEVLGAFFSVFLIWIVTGVL
                      VVLAIMRIVSGDTEVEGGIMALTAALGVVVNLVMLALLYFGGHSHSHGGGSSSHGHSHG
                      GGNGDNINVRAAFIHVLGDLLQSLGVLVAALFIYFQPSWVIIDPICTLVFSVIVLCTT
                      IYILRDAMIVLLEGRPSNIDFAKVFSSLEDIDGVKEVHDLRIWSLTMDFIALSVHLEI
                      DANSQSQSILRETRKMLKQTYNVHEITIQIEEFGANRSDCGKCDFPTK"
     CDS              join(30250. .30312,30423. .30551,30608. .30748,30798. .31272,
                      31328. .31509,31558. .31717,31978. .32601,Z68117:101. .768,
                      Z68117:821. .3357,Z68117:3404. .3554,Z68117:3602. .3892,
                      Z68117:3961. .4023,Z68117:4078. .4359,Z68117:4406. .4546)
                      /gene="myo-1"
                      /note="Identity to myosin heavy chain C (SW:MYSC_CAEEL);
                      cDNA EST EMBL:M89007 comes from this gene; cDNA EST
                      EMBL:Z14591 comes from this gene; cDNA EST EMBL:D33348
                      comes from this gene; cDNA EST EMBL:D65585 comes from this
                      gene; cDNA EST EMBL:D64088 comes from this gene; cDNA EST
                      EMBL:D69073 comes from this gene; cDNA EST EMBL:C11379
                      comes from this gene; cDNA EST EMBL:C13479 comes from this
                      gene; cDNA EST EMBL:C11537 comes from this gene; cDNA EST
                      EMBL:C13791 comes from this gene; cDNA EST EMBL:C11506
                      comes from this gene; cDNA EST EMBL:C13728 comes from this
                      gene; cDNA EST EMBL:C11245 comes from this gene; cDNA EST
                      EMBL:C13288 comes from this gene; cDNA EST EMBL:C10670
                      comes from this gene; cDNA EST EMBL:C12497 comes from this
                      gene; cDNA EST EMBL:C10222 comes from this gene; cDNA EST
                      EMBL:C11779 comes from this gene; cDNA EST EMBL:C12848
                      comes from this gene"
                      /codon_start=1
                      /db_xref="PID:e1323092"
                      /db_xref="PID:g3646582"
                      /translation="MDYENDPGWKYLRRSREEMLQDQSRAYDSKKNVWIPDSEDGYIE
                      GVITKTAGDNVTVSIGQGAEKTVKKDVVQEMNPPKFEKTEDMSNLTFLNDASVLYNLK
                      ARYAAMLIYTYSGLFCVVINPYKRLPIYTDSVARMFMGKRRTEMPPHLFAVSDEATRN
                      MLQNHENQSMLITGESGAGKTENTKKVIGYFAAVGAAQQFTPGAKKAATEEDKNKKEV
                      TLEDQIVQTNPVLEAFGNAKTVRNNNSSRFGKFIRIHFSKQGRVASCDIEHYLLEKSR
                      VIRQAPGERCTYHIFYQVFSDYLPNLKKDLLLNKPVKDYWFIAQAELIIDGINDKEEHQ
                      LTDEAFDILKPTPTEKMECYRLVAAMMHMGNMKFKQRPREEQAEPDGTDDAERAARCF
                      GIDSEEFLKALTRPRVKVGNEWVNKGQNIEQVNWAVGAMAKGLYSRIFNWLVKKCNQT
```

```
                    LDQKGISRDHFIGVLDIAGFEIFDFNSFEQLWINFVNEKLQQFFNHHMFVLEQEEYAR
...
Note: remainder of annotations omitted.

  Query Match           5.7%;  Score 17;  DB 18;  Length 32601;
  Best Local Similarity 100.0%;  Pred. No. 1.61e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  26036 ATACATTTATACATTTA 26052
          |||||||||||||||||
Cp    119 ATACATTTATACATTTA 103


RESULT   47
LOCUS       CEK08E5     32828 bp    DNA             INV       21-SEP-1998
DEFINITION  Caenorhabditis elegans cosmid K08E5, complete sequence.
ACCESSION   Z30974
NID         g1067000
KEYWORDS    HTG.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 32828)
  AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
  TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
  JOURNAL   Nature 368 (6466), 32-38 (1994)
  MEDLINE   94150718
REFERENCE   2  (bases 1 to 32828)
  AUTHORS   Kershaw,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-NOV-1996) Nematode Sequencing Project, Sanger Centre,
            Hinxton, Cambridge CB10 1RQ, England and Department of Genetics,
            Washington University, St. Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     On Nov 21, 1995 this sequence version replaced gi:463205.
            Coding sequences below are predicted from computer analysis, using
            predictions from Genefinder (P. Green, U. Washington), and other
            available information.
            For a graphical representation of this sequence and its analysis
```



```
            see:-
            http://weback.sanger.ac.uk/cgi-
            bin/display?db=wormace&class=Sequence &object=K08E5
            Current sequence finishing criteria for the C. elegans genome
            sequencing consortium are that all bases are either sequenced
            unambiguously on both strands, or on a single strand with both a
            dye primer and dye terminator reaction, from distinct subclones.
            Exceptions are indicated by an explicit note.
            IMPORTANT:  This sequence is NOT necessarily the entire insert of
            the specified clone.  It may be shorter because we only sequence
            overlapping sections once, or longer because we arrange for a small
            overlap between neighbouring submissions.
            IMPORTANT: This sequence is not the entire insert of clone K08E5.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we arrange for a small overlap between
            neighbouring submissions.
            The true left end of clone K08E5 is at 1 in this sequence. The true
            left end of clone T20G5 is at 32729 in this sequence. The true
            right end of clone ZK1128 is at 11463 in this sequence. The start
            of this sequence (1. .101) overlaps with the end of sequence Z47357.
            The end of this sequence (32729. .32828) overlaps with the start of
            sequence Z30423.
FEATURES                    Location/Qualifiers
     source                 1. .32828
                            /organism="Caenorhabditis elegans"
                            /db_xref="taxon:6239"
                            /chromosome="III"
                            /clone="K08E5"
     gene                   15432. .18782
                            /gene="K08E5.2"
     CDS                    join(15432. .15554,15633. .15898,15983. .16123,16714. .16848,
                            16896. .17012,17081. .17765,18645. .18782)
                            /gene="K08E5.2"
                            /note="similar to Yeast ORF YCR37C; cDNA EST yk250f6.5
                            comes from this gene"
                            /codon_start=1
                            /db_xref="PID:g1067001"
                            /db_xref="SWISS-PROT:Q21339"
                            /translation="MQIRHVRKMDPLEILAVAKNEKPWKIGSPPTIRKNPANWQLEWK
                            CAYCVCIIAVYWMSEVMPLAVTAMLPVVLFPLVGVLDANTTAKEYMNDTNFLFIGGLI
                            MAAAVEKCDLHERVALSVLRCVGSEPKWIMLGFMTVTALLSSFISNTATTAMMVPIGQ
                            SVVQQLISSFQHHPTNGERGRLGCKKMATGLVLSICFAANIGGTGTATGTPSNLVMLG
                            QLSALFPKVDGSLNYVTWIFFAYPLMLLCLFVAWMTLVSFFLRDAPEKDEAVTEMLKT
                            RYNELPRMTYAEKSVFVCFCILLSLWVFRNPGVVPGPGVFFKKGAYTDATSAMIVAFL
                            LFVLPSERPDLATYIKKEDLKKRGCLMDWKTMQETFPWSVVLLGGGFALAAGVTESG
                            LSLLIGNSLSSIEHLPLWILQLLTMLIAMVITWICSNTVTASIFVPIVATLAQRAGHH
                            PPTLMLPTTLASSFAFIPPVGTPPNAIVFGSGMVKVSDMAFVGGIISLELLVLTVLYM
                            NSIAYLTLPLLEPPTWAIIANSTMQQ"
     gene                   22030. .27635
                            /gene="K08E5.3"
     CDS                    join(22030. .22210,22574. .22686,24261. .24486,24743. .25141,
```



```
                    25202. .25317,25576. .25821,26285. .26465,26516. .26605,
                    26667. .26819,27182. .27266,27317. .27466,27515. .27635)
                    /gene="K08E5.3"
                    /note="similar to LDL receptor related protein"
                    /codon_start=1
                    /db_xref="PID:g463207"
                    /db_xref="SPTREMBL:Q21340"
                    /translation="MQAGISIFFLFLHIPIFFVNCSNSTSCVAREEFQCKMDDSCISM
                    KKWQDGVDDCYDGSDEVTSYSQKKLSRFSETMYKTKLITVFPSFFSHFSFFSKPCACH
                    GNSTVSLDHLDVFQKTNCTTKRSIVILVSTKDVLLTILCVEIEVHALNHRNILMVLLI
                    VKISRMNVSDDIQDTKKNPKNICQKNKKVAKSGHTFLKIFFSSLSRHCATPKIQKTPP
                    EYKNQTPKKCFSACAQNQFQCSDGTKCIPKAQFQDGKEDCDDGSDEECTTSQFACQCG
                    TIKCVSDTFIMDGNWDCEDGSDEFINKTLTANCTRNNKVNIATNSLSLGKLKFCSGKN
                    PCKPELGQVCVIIGGTWRCVCKLGTFRPLGSDKCIPIELLTRYRNAPSSKCSNSHEPQ
                    FGLLQGFQKSLSSRKELERWSHIRRAPPRTLSGTTPSGSDGFLKSGAEQVFPWFQEDK
                    NYVSVIPDEIIDSYGTIMTPPEFHFNRDDIRCGNKTCGLHESCQKNSESKYECICREG
                    FTIFEGTCRELIDECAQGKHDCHPEARCVDALIGYECLCREGYLDTSIDPKARPGRKC
                    RKLINECTNALWNDCSQNARCLDKPIGYTCRCQDDYVDVSREGARKPGRNCTQAINEC
                    ASNLHNCDTHAICQDQPVGYSCRCPPGFIDSSPTALEPGRKCVQGECSKHQKIEKKSC
                    SRIV"
     gene           join(Z47357:26136. .26323,Z47357:26373. .26421,
                    Z47357:27277. .27485,91. .301)
                    /gene="ZK1128.7"
     CDS            join(Z47357:26136. .26323,Z47357:26373. .26421,
                    Z47357:27277. .27485,91. .301)
                    /gene="ZK1128.7"
                    /note="similar to alpha-B-crystallin"
                    /codon_start=1
                    /db_xref="PID:e1322774"
                    /db_xref="PID:g3643236"
                    /translation="MHHVDFDDDDLDSYLEGSVWSEKPSASEIPSSKPGDVISVFPRS
                    SHSNMIYGYPINYKETVFPTHRGPPDQSYDSYYTKITESRPRSPGPVAGAGEITNTSH
                    GFTIEIDVFHFMPEEIKVVLTDDTLSISGERFESTGDGHTLRRSFSRKYSIPDDVHLD
                    TIRSHLTNSGVLIINVSFSKNLWISILQNTYSRVLVKDGVRQASQAITQQLNETWQEV
                    "
BASE COUNT    10368 a   5925 c   5850 g  10685 t
ORIGIN

  Query Match           5.7%;  Score 17;  DB 18;  Length 32828;
  Best Local Similarity 100.0%;  Pred. No. 1.61e+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  19902 TAAATGTATAAATGTAT 19918
          |||||||||||||||||
Qy    103 TAAATGTATAAATGTAT 119


RESULT   48
LOCUS       CELC34B2    33448 bp    DNA             INV       23-JAN-1998
DEFINITION  Caenorhabditis elegans cosmid C34B2.
ACCESSION   AF043693
```



```
NID         g2804410
KEYWORDS    .
SOURCE      Caenorhabditis elegans strain=Bristol N2.
  ORGANISM  Caenorhabditis elegans
            Eukaryotae; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 33448)
  AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
  TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
  JOURNAL   Nature 368 (6466), 32-38 (1994)
  MEDLINE   94150718
REFERENCE   2  (bases 1 to 33448)
  AUTHORS   Graves,T., Sutterer,C., Hawkins,M. and Wilson,R.
  TITLE     The sequence of C. elegans cosmid C34B2
  JOURNAL   Unpublished (1998)
REFERENCE   3  (bases 1 to 33448)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JAN-1998) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     Submitted by:
                    Genome Sequencing Center
                    Department of Genetics, Washington University,
                    St. Louis, MO  63110, USA, and
                    Sanger Centre, Hinxton Hall
                    Cambridge CB10 1RQ, England
                    e-mail: rw@nematode.wustl.edu and jes@sanger.ac.uk


            NOTICE:  This sequence may not be the entire insert of this clone.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we provide a small overlap between
            neighboring submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded or sequenced with an alternate
            chemistry; an attempt was made to resolve all sequencing problems,
            such as compressions and repeats; all regions were covered by
            sequence from more than one subclone
```

NEIGHBORING COSMID INFORMATION:

The 5' cosmid is 80511, 200 bp overlap. Actual start of this cosmid is at base position 38872 of CELB0511; actual end is at 33448 of CELC34B2. This cosmid lies in an unanchored cluster, the orientation of which is unknown.

NOTES:

Coding sequences below are predicted from computer analysis, using the program Genefinder(P. Green and L. Hillier, ms in preparation).

```
FEATURES             Location/Qualifiers
     source          1. .33448
                     /organism="Caenorhabditis elegans"
                     /strain="Bristol N2"
                     /db_xref="taxon:6239"
                     /chromosome="I"
                     /clone="C34B2"
     gene            complement(828. .1872)
                     /gene="C34B2.4"
     CDS             complement(join(828. .945,993. .1071,1116. .1186,1231. .1292,
                     1807. .1872))
                     /gene="C34B2.4"
                     /note="contains similarity to LIM domains"
                     /codon_start=1
                     /evidence=not_experimental
                     /db_xref="PID:g2804415"
                     /translation="MNECGHCSVKIGEESAILANGKVWHVHHLLCDLCKCRINDGERC
                     VPQNGVILCSECHIKTTRPICKGCGEFIKTNFCEALNSTWHPTCPQCSVCQKPLKVDF
                     HQLPNRMCVHSDCFWDLQLRMIVAKDLPK"
     gene            2096. .4107
                     /gene="C34B2.3"
     CDS             join(2096. .2215,2257. .2373,2586. .2680,2730. .2799,
                     2991. .3086,3131. .3214,3268. .3486,3739. .3843,3893. .3979,
                     4027. .4107)
                     /gene="C34B2.3"
                     /note="contains similarity to ser/thr protein kinases;
                     most similar to casein kinases"
                     /codon_start=1
                     /evidence=not_experimental
                     /db_xref="PID:g2804414"
                     /translation="MEDRNITEEDRVCPQLPEFDEQANVQTDLKKMKPVRSDVQGDAI
                     GKNWKVVKAIQSDKGFNTIYVAEHVQKKKMAAVRVERKTEAIKMLQFELFVLLTVEK
                     KNQCKQFCNLFEKGWEEEYWWIAITLCGKSLRALRKNQPTGKLTVACGLSVAQQCLKG
                     LEELHRMGFIHRNVAPSVFAIGRYTGDMQSDLRNIYIILDFGFAHQYMIKDGTLKPPSA
                     HPWKYVGSLRHMPRAAYSKVEFSRMEDLEMWFYMSVELVKGCLPWAHLKKPKEVHDYQ
                     KLCRNGLQMREMLGGLPPEFVDIMQIVDKLSFTDTPNYTEIYGLLTNAILFSGKNEPP
                     YDWEEAEINEFKNPQKPSVEQAT"
     gene            complement(4561. .6086)
```



```
                     /gene="C34B2.5"
     CDS             complement(join(4561. .4827,4879. .4980,5492. .5620,
                     5961. .6086))
                     /gene="C34B2.5"
                     /note="contains similarity to peptidylprolyl isomerases"
                     /codon_start=1
                     /evidence=not_experimental
                     /db_xref="PID:g2804416"
                     /translation="MAIIEEIFEDEEDQITKVDSLKKEGNNFPANGEFEKANEKYQEA
                     IASCPPTSTEVQSILLSNSAAALIKLRKWESAVEAASKSIEIGATNEKALERRAFAYS
                     NMSEKYENSIEDYKQLQESLPKRRVELDRKIAEINEKINNRNEAMKADVMEKLKGFGN
                     FCLSPFGLSTDSFEMVPNGNGGFSVQMKGATNKQEGASQEKKEEENA"
     gene            complement(6346. .11655)
                     /gene="C34B2.6"
     CDS             complement(join(6346. .6489,6537. .6701,6751. .6993,
                     7040. .7279,7672. .7930,7982. .8250,8306. .8558,9169. .9482,
                     9536. .9698,10359. .10808,11139. .11230,11280. .11381,
                     11434. .11655))
                     /gene="C34B2.6"
                     /note="similar to the LON family of ATP-dependent
                     proteases; coded for by C. elegans cDNA yk33b9.3; coded
                     for by C. elegans cDNA yk43g11.3; coded for by C. elegans
                     cDNA yk2f3.3; coded for by C. elegans cDNA yk131g7.3;
                     coded for by C. elegans cDNA CEESQ02F; coded for by C.
                     elegans cDNA yk180g11.3; coded for by C. elegans cDNA
                     yk180g11.5; coded for by C. elegans cDNA yk43g11.5; coded
                     for by C. elegans cDNA cm18e11; coded for by C. elegans
                     cDNA yk33b9.5; coded for by C. elegans cDNA yk131g7.5;
                     coded for by C. elegans cDNA yk2f3.5"
                     /codon_start=1
                     /db_xref="PID:g2804417"
                     /translation="MYRAGAVLLRGATRTRLLAAASAHQSFATFSQRNQSILMMKSME
                     LAGNSGERRFYSTHDDPIAVDDSLELYKDLGGMSPIQVPADMPNVPMLAINRYPLPPG
                     FIKKVDIVKDDNLKALIRRQLSLKQPYAGVFVKRDDENKEETITSLSEVTPTGSFVQI
                     IEVRDQGSVLELVLSAHRRIRALEPIDEITPKNETPLNGRRARGKRAASATSPLTPPP
                     SPPPLAPSVLNAVAPEISATEEKEETTPPSATGEKQKKGIINVRTENVVAEPVPKNNE
                     TKATMMAIVQTIRDVVQFNQLFGQQINLLLHPSQNVIDNPVYLCDLVATLVQSAETKD
                     LQEMMDEIDVSKRLKIALLLIQKEKAVAKLKYDINKDVEKKVQDHHRKYLLNEQLKVI
                     KKELGIEKDEKTTIIEKIDERIKTLAVPEYALKVINEEKTKLQFLDPHSSEFSVTRNY
                     LEWLTSVPWGLTSPENRRLSVAKKALDEGHYGMKDVKERIMEFIAVNLLRKSIGGKIL
                     CFHGPPGVGKTSIAKSIATALNREYFRFSVGGMTDVAEIKGHRRTYVGAMPGKMIQCM
                     KKVKTENPLVLIDEVDKIGGAGFHGDPASALLELLDPEQNANFNDHFLDVPVDLSRVL
                     FICTANEISKIPGPLRDRMEMIDVSGYLAEEKVEIAHQHLIPQLRKDTSLATEQLKIE
                     DSALEELIKHYCRESGVRNLQQHIERIFRKAALQIAEQQNEDEEPAEKATTAITENSE
                     AEPITSTSSADCLKSSABQIVVCTENLQKFVGRPKPTSDRMYEVTPPGVIMGLAWTAM
                     GGSALYIETVLKRPVDLTNDKDGSIETTGNLGDVMKESVRTALTVAKGILAREQPDNK
                     FFDKAHIHIHVPEGATPKDGPSAGVTLVSSLLSALDRPVVQDMAMTGEISLTGKVLP
                     VGGIREKVIAARRVGAKRVFLPNENRRDFDDLPEFMKSELDIRFVSHYDELYEHLFQ"
     gene            complement(11905. .12323)
                     /gene="C34B2.10"
     CDS             complement(join(11905. .11970,12016. .12171,12228. .12323))
```



```
                     /gene="C34B2.10"
                     /note="coded for by C. elegans cDNA yk342g1.3; coded for
                     by C. elegans cDNA yk342g1.5"
                     /codon_start=1
                     /db_xref="PID:g2804411"
                     /translation="MDGMIAMLPAPLQKLSSHIDFQGQKVAERTYQVILTIAGIIGFL
                     VGFWTQQLSYAMFTVLGASAFTALIILPPWPFLFRKNPIVWHTPAEPQESGDKKKETK
                     KTK"
     gene            complement(12661. .14866)
                     /gene="C34B2.7"
     CDS             complement(join(12661. .13059,13111. .13300,13344. .13611,
                     13656. .14330,14376. .14528,14576. .14754,14808. .14866))
                     /gene="C34B2.7"
                     /note="similar to the flavoprotein subunit of succinate
                     dehydrogenase; coded for by C. elegans cDNA yk98a10.5;
                     coded for by C. elegans cDNA yk98a10.3"
                     /codon_start=1
                     /db_xref="PID:g2804418"
                     /translation="MLNVVKSINRAKTPVRTTMKKQVSATTNFDVVDHTFDAVVVGAG
                     GAGLRAAMGLSEGGMKTAVITKLFPTRSHTVAAQGGVNAALGNMNPDNWRWHFYDTVK
                     GSDWLGDQDAIHYMTREAERAIIELENYGMPFSRTTDGKIYQRAFGGQSNDFGRGGQA
                     HRTCCVADRTGHSLLHTLYGASLQYDCNYFVEYFALDLIMDKGKCIGVVALDIETGQI
                     HRFRAKNTVLATGGYGRAYFSCTSAHTCTGDGTALTARAGIRNSDMEFVQFHPTGIYG
                     VGCLITEGSRGEGGYLVNSQGERFMERYAPNAKDLASRDVVSRAMTMEINEGRGVGPN
                     KDHIYLQLHHLPAEQLQQRLPGISETAQIFAGVDVTKEPIPVIPTVHYNMGGVPTNYK
                     GQVLDFTPEGGDKVIPGLYAAGECAAHSVHGANRLGANSLLDLVIFGRSCALTILNEN
                     KPGDSIPELPVNCEEKSCDNLNSLLNSKGDISSIELRQKMQMTMQKHAAVFRRGDLLK
                     EGVDKMSSIYKEQQNLKACADSGKVWNSELVETLELQNLLINANQTIVAAENRTESRG
                     AHARDDFQERIDEYDYSNPLEGQQKKPFDQHWRKHSIIGIDTKTGAVDLTYRPVIDKT
                     LDKSETDWVPPKVRSY"
     misc_feature    complement(14870. .14871)
                     /note="SL1 trans-splice site; see EST yk98a10.5"
     gene            complement(15324. .16336)
                     /gene="C34B2.8"
     CDS             complement(join(15324. .15449,15493. .15706,15752. .15830,
                     16240. .16336))
                     /gene="C34B2.8"
                     /note="coded for by C. elegans cDNA yk114e9.3; coded for
                     by C. elegans cDNA yk85c8.5; coded for by C. elegans cDNA
                     yk114e9.5"
                     /codon_start=1
                     /db_xref="PID:g2804419"
                     /translation="MATGAEFRQDMPPKGGYRAFNFHRTFPKLVWSPGTVVAAIFGAT
                     AYGVYAALEGKKKDITEKFEDVDINNAMEPFLTAERDRYWLKLLAKNRALEEEIMKDV
...
```

Note: remainder of annotations omitted.

```
  Query Match           5.7%;  Score 17;  DB 18;  Length 33448;
  Best Local Similarity 100.0%;  Pred. No. 1.6le+01;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  10113 AGTTTTTTATAAATGTA 10129
          |||||||||||||||||
Qy     94 AGTTTTTTATAAATGTA 110
```



```
RESULT   49
LOCUS       CELT21F2     37386 bp    DNA             INV       20-DEC-1995
DEFINITION  Caenorhabditis elegans cosmid T21F2.
ACCESSION   U42836
NID         g1125766
KEYWORDS    .
SOURCE      Caenorhabditis elegans strain=Bristol N2.
  ORGANISM  Caenorhabditis elegans
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Nematoda;
            Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea;
            Rhabditidae; Caenorhabditis.
REFERENCE   1  (bases 1 to 37386)
  AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
  TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
  JOURNAL   Nature 368 (6466), 32-38 (1994)
  MEDLINE   94150718
REFERENCE   2  (bases 1 to 37386)
  AUTHORS   Minx,P and Gattung,S.
  TITLE     The sequence of C. elegans cosmid T21F2
  JOURNAL   Unpublished (1995)
REFERENCE   3  (bases 1 to 37386)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-DEC-1995) Robert Waterston
COMMENT               Submitted by:
                      Genome Sequencing Center
                      Department of Genetics, Washington University,
                      St. Louis, MO  63110, USA, and
                      Sanger Centre, Hinxton Hall
                      Cambridge CB10 1RQ, England
                      e-mail: rw@nematode.wustl.edu and jes@sanger.ac.uk
                      NEIGHBORING COSMID INFORMATION:
            5' cosmid is F22H10; 3' cosmid is C10E2. Actual start of this
            cosmid is at base position 1 of CELT21F2; actual end is at 37386 of
            CELT21F2.
```

```
            NOTES:
            Coding sequences below are predicted from computer analysis, using
            the program Genefinder(P. Green and L. Hillier, ms in preparation).
FEATURES                Location/Qualifiers
     source             1. .37386
                        /organism="Caenorhabditis elegans"
                        /db_xref="taxon:6239"
                        /map="X"
     gene               13737. .16413
                        /gene="T21F2.1"
     CDS                join(13737. .13760,13963. .14145,14915. .15100,15195. .15307,
                        15351. .15447,15492. .15554,15681. .15839,15897. .15952,
                        16004. .16098,16141. .16271,16324. .16413)
                        /gene="T21F2.1"
                        /note="Similar to ligand-gated ionic channel."
                        /codon_start=1
                        /evidence=not_experimental
                        /db_xref="PID:g1125767"
                        /translation="MYLQMKDHVIVELTVQSITEISEFSSSFKADVWFSQIWRDPRLD
                        FSDRNYCLKNISLASHKLPTMWSPNVEIHASPSQNILLLIFPNGTVWLNFRVSLTGPC
                        KLDLTYFPMDRQSCNLVFEVSMSIPSSLQSYSYNTAEVRIVWRDWDAVTIPDPDSKNL
                        PDFELVNIEHMNATLVYTAGLWDQLEVKPTFRRLYGYYVLQAYMPTYLSVFISWIAFW
                        IDTKALPARITLGVSSLMALTFQFGNIVKNLPRVSYVKALDIWMFGCVGFIFLSLVEL
                        AVVGFADKLEAKRRRHNRCKEQLWMRSDSEQQWLSRLSGQRPQVSETNSDATHTIQIN
                        DGNGNVRRRKSEEKRQKLLLQMESPMFVNGEKIDEISAKLFPLLFTAFNIFYWFYYIG
                        MSGGFF"
BASE COUNT    11745 a   6618 c   6193 g  12830 t
ORIGIN

  Query Match             5.7%;  Score 17;  DB 18;  Length 37386;
  Best Local Similarity 100.0%;  Pred. No. 1.61e+01;
  Matches    17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  30461 ATAAATGTATAAATGTA 30477
          |||||||||||||||||
Qy    102 ATAAATGTATAAATGTA 118


RESULT   50
LOCUS       AC004177    38587 bp    DNA             HTG       23-FEB-1998
DEFINITION  *** SEQUENCING IN PROGRESS *** Homo sapiens Cosmid C42; HTGS phase
            2, 1 ordered pieces.
ACCESSION   AC004177
NID         g2905883
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 38587)
```



```
  AUTHORS   Janer,M., Guillaudeux,T., Vu,Q., Kutyavin,T., Harter,H. and
            Geraghty,D.E.
  TITLE     Large scale sequence analysis of the human MHC class I region
  JOURNAL   Unpublished (1998)
  REMARK    Fred Hutchinson Cancer Research Center
            The Clinical Research Division
            1100  Fairview Ave. N., P.O. Box 19024
            Seattle, WA 98109-1024

REFERENCE   2  (bases 1 to 38587)
  AUTHORS   Geraghty,D.E. and Olson,M.V.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-FEB-1998) Human Genome Center, University of
            Washington, Box 352145, Seattle, WA 98195, USA
COMMENT      University of Washington Human Genome Center

              Box 352145 Seattle, WA 98195
              Contact: Daniel E. Geraghty (geraghty@fhcrc.org).
            ***                                                      ***
            *** WARNING: Phase 2 High Throughput Genome Sequence ***
            ***                                                      ***
            * This sequence is unfinished. It consists of 1 contigs for
            * which the order is known. The lengths of the gaps have been
            * estimated by the submitter but are not known exactly. When
            * sequencing is complete, the sequence data presented in this
            * record will be replaced by a single finished sequence
            * with the same accession number.
            *       1    38587: contig of 38587 bp in length.
FEATURES                Location/Qualifiers
     source             1. .38587
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone="UWGC:C42"
                        /chromosome="6"
                        /map="p21"
BASE COUNT    10181 a   8633 c   8943 g  10825 t      5 others
ORIGIN

  Query Match             5.7%;  Score 17;  DB 17;  Length 38587;
  Best Local Similarity  94.4%;  Pred. No. 1.61e+01;
  Matches    17;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db  16958 AAGGCACCCCTGGTGCCC 16975
          ||||||| ||||||||||
Cp    179 AAGGCACNCCTGGTGCCC 162


Search completed: Tue Dec 22 22:24:42 1998
Job time : 743 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Tue Dec 22 22:34:52 1998; MasPar time 59.05 Seconds
723.485 Million cell updates/sec

Tabular output not generated.

Title: >US-08-530-112A-4937
Description: (1-314) from US08530112A.seq
Perfect Score: 298
N.A. Sequence: 1 GATCCTCTCTCCCAGGGAAT.....CCTNCNTGTGGCCCCCTNTTN 314
Comp: CTAGGAGAGAGGGTCCCTTA.....GGANGNACACCGGGGANAAN

Scoring table: TABLE jmetric
Gap 60

Nmatch STD : Dbase 0; Query 0

Searched: 188442 seqs, 68026449 bases x 2

Post-processing: Minimum Match 0%
Listing first 1000 summaries

Database: n-geneseq32
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39 40:part40

Statistics: Mean 6.773; Variance 2.783; scale 2.434

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
	1	298	100.0	314	20	T23937	Human gene signature	1.65e-278
	2	18	6.0	390	8	Q59241	Human brain Expressed	2.85e-01
	3	18	6.0	390	6	Q39829	Expressed Sequence Ta	2.85e-01
	4	18	6.0	1641	13	Q80319	Human 38kDa FK-506 bi	2.85e-01
c	5	16	5.4	1146	31	T68945	Mouse p32 11-cis-reti	6.37e+00
c	6	16	5.4	1649	3	Q14638	Human temporal lobe c	6.37e+00
c	7	16	5.4	1649	21	T34379	Plasmid pPDE2RR inser	6.37e+00
	8	16	5.4	1891	1	Q03601	Gene encoding yeast a	6.37e+00
	9	16	5.4	2058	36	T72943	Phaffia carotenoid bi	6.37e+00
	10	16	5.4	2226	12	Q71287	Feline zona pellucida	6.37e+00
c	11	16	5.4	3186	22	T34375	Plasmid pTM72 (ATCC 6	6.37e+00
c	12	16	5.4	3186	3	Q14631	Plasmid pTM72 insert	6.37e+00
	13	16	5.4	3804	39	T99544	Human haematopoietic	6.37e+00
c	14	16	5.4	3890	12	Q71539	Brain low Km, cAMP-sp	6.37e+00
c	15	15	5.0	148	10	Q76548	Human genome fragment	2.78e+01
c	16	15	5.0	151	10	Q76546	Human genome fragment	2.78e+01
c	17	15	5.0	204	10	Q76549	Human genome fragment	2.78e+01
c	18	15	5.0	212	19	T21944	Human gene signature	2.78e+01
	19	15	5.0	252	18	T19597	Human gene signature	2.78e+01
	20	15	5.0	342	8	Q61191	Human brain Expressed	2.78e+01
c	21	15	5.0	370	8	Q60937	Human brain Expressed	2.78e+01
	22	15	5.0	534	19	T12305	Gyrase A probe.	2.78e+01
	23	15	5.0	732	4	Q23090	Antigen ac-1b gene.	2.78e+01
	24	15	5.0	1131	7	Q40029	Sequence encoding apo	2.78e+01
	25	15	5.0	1131	14	Q89633	Human apo B-100 encod	2.78e+01
	26	15	5.0	1471	1	N91115	Portion of clone P4A	2.78e+01
	27	15	5.0	1472	2	N70984	Sequence of flagellin	2.78e+01
	28	15	5.0	1578	4	N40290	Sequence of the alpha	2.78e+01
	29	15	5.0	2155	31	T72149	Murine mortalin cDNA	2.78e+01
	30	15	5.0	2599	1	Q05688	Islets of Langerhans	2.78e+01
	31	15	5.0	4588	38	V02185	Murine type C lectin	2.78e+01
	32	15	5.0	5499	5	Q31189	Alpha 6B integrin sub	2.78e+01
	33	15	5.0	5529	11	Q65673	Sequence encoding alp	2.78e+01
	34	15	5.0	5629	5	Q31188	Alpha 6A integrin sub	2.78e+01
c	35	15	5.0	5760	3	N50530	Sequence encoding the	2.78e+01
	36	14	4.7	18	16	Q95896	Primer B (Group 1, s	1.14e+02
c	37	14	4.7	59	38	V00376	Insecticidal gene seq	1.14e+02
c	38	14	4.7	109	20	T21539	Human gene signature	1.14e+02
	39	14	4.7	187	19	T21176	Human gene signature	1.14e+02
	40	14	4.7	195	22	T24941	Human gene signature	1.14e+02
	41	14	4.7	210	20	T24104	Human gene signature	1.14e+02
c	42	14	4.7	250	20	T23555	Human gene signature	1.14e+02

	43	14	4.7	284	21	T24718	Human gene signature	1.14e+02
	44	14	4.7	301	16	Q95197	Simple tandem repeat	1.14e+02
	45	14	4.7	302	8	Q59747	Human brain Expressed	1.14e+02
c	46	14	4.7	316	6	Q39815	Expressed Sequence Ta	1.14e+02
c	47	14	4.7	316	8	Q59227	Human brain Expressed	1.14e+02
	48	14	4.7	343	8	Q60627	Human brain Expressed	1.14e+02
c	49	14	4.7	345	4	Q28186	Human T lymphocyte re	1.14e+02
c	50	14	4.7	396	14	Q89257	BMP-5 encoding DNA.	1.14e+02
	51	14	4.7	418	6	Q39697	Expressed Sequence Ta	1.14e+02
	52	14	4.7	418	8	Q59109	Human brain Expressed	1.14e+02
c	53	14	4.7	450	9	Q54001	Sequence encoding hum	1.14e+02
c	54	14	4.7	457	7	Q43679	Sequence of intron 29	1.14e+02
c	55	14	4.7	459	4	Q24073	Interleukin-1 beta DN	1.14e+02
c	56	14	4.7	459	9	Q54000	Sequence encoding hum	1.14e+02
c	57	14	4.7	460	2	N60284	Sequence encoding new	1.14e+02
c	58	14	4.7	462	37	V02215	Leaderless protein IL	1.14e+02
c	59	14	4.7	464	4	Q24072	Interleukin-1 beta DN	1.14e+02
c	60	14	4.7	465	4	Q24071	Interleukin-1 beta DN	1.14e+02
c	61	14	4.7	465	14	Q83759	Plasmid 12151.	1.14e+02
c	62	14	4.7	465	31	T72206	DNA encoding mature h	1.14e+02
c	63	14	4.7	469	31	T72208	DNA encoding human in	1.14e+02
c	64	14	4.7	469	14	Q83761	Plasmid pDM25466.	1.14e+02
c	65	14	4.7	472	1	Q03664	Recombinant interleuk	1.14e+02
c	66	14	4.7	475	1	Q04074	Interleukin - 1 beta	1.14e+02
c	67	14	4.7	478	31	T72207	DNA encoding mature h	1.14e+02
c	68	14	4.7	478	2	Q10430	Hybrid IL-1 beta/huma	1.14e+02
c	69	14	4.7	478	14	Q83760	Plasmid pDM15449.	1.14e+02
	70	14	4.7	483	12	Q70414	Pre-pro-vasoactive in	1.14e+02
	71	14	4.7	488	39	V13359	Ragweed pollen IPC Cl	1.14e+02
	72	14	4.7	488	1	Q06220	IPC Clone 5 (Amb a IA	1.14e+02
c	73	14	4.7	510	35	T72980	DNA encoding sugar be	1.14e+02
	74	14	4.7	578	28	T59830	Beta-casein gene frag	1.14e+02
c	75	14	4.7	579	2	N60285	Sequence encoding new	1.14e+02
c	76	14	4.7	594	2	N70849	Plasmid pT9-11 encodi	1.14e+02
c	77	14	4.7	597	2	N60286	Sequence encoding new	1.14e+02
c	78	14	4.7	702	31	T72205	DNA encoding mature h	1.14e+02
c	79	14	4.7	702	14	Q83758	Plasmid 16963.	1.14e+02
c	80	14	4.7	715	40	V10152	Stealth virus nucleic	1.14e+02
c	81	14	4.7	730	40	V15431	Human gene fragment Q	1.14e+02
	82	14	4.7	774	34	T91660	Prepro mistletoe lect	1.14e+02
c	83	14	4.7	810	37	V02214	Leaderless protein IL	1.14e+02
c	84	14	4.7	915	31	T72211	DNA encoding mature h	1.14e+02
c	85	14	4.7	915	14	Q83764	Plasmid 16965.	1.14e+02
	86	14	4.7	981	37	T76682	Methanococcus jannasc	1.14e+02
c	87	14	4.7	986	1	N82421	Sequence of human per	1.14e+02
	88	14	4.7	1112	16	Q89727	Plasmodium falciparum	1.14e+02
c	89	14	4.7	1122	35	T78773	DNA encoding Aquifex	1.14e+02
c	90	14	4.7	1125	2	N60641	Prepro- and mature in	1.14e+02
c	91	14	4.7	1125	2	N60350	Sequence encoding int	1.14e+02
c	92	14	4.7	1125	3	Q14646	Encodes mature Interl	1.14e+02
c	93	14	4.7	1155	16	T01083	2-Alpha-fucosyltransf	1.14e+02

c	94	14	4.7	1155	18	Q98461	GDP-L-fucose-beta-D-g	1.14e+02
c	95	14	4.7	1155	16	T01082	2-Alpha-fucosyltransf	1.14e+02
c	96	14	4.7	1174	17	T12238	Human H-transferase c	1.14e+02
	97	14	4.7	1234	12	Q70415	Full length pre-pro-v	1.14e+02
c	98	14	4.7	1333	27	T48519	Human neuropsin-encod	1.14e+02
c	99	14	4.7	1333	29	T63251	Mouse neuropsin gene.	1.14e+02
c	100	14	4.7	1404	2	N60283	Sequence encoding new	1.14e+02
c	101	14	4.7	1404	2	N70505	Sequence encoding int	1.14e+02
c	102	14	4.7	1404	9	Q51110	Human interleukin-1 b	1.14e+02
c	103	14	4.7	1433	24	Q99243	Insulin like growth f	1.14e+02
c	104	14	4.7	1438	1	Q04550	EcoRI-EcoRI fragment	1.14e+02
	105	14	4.7	1494	34	T73723	Human filamin-like be	1.14e+02
	106	14	4.7	1495	40	V15253	Brassica campestris d	1.14e+02
	107	14	4.7	1495	2	Q13965	B. campestris delta9	1.14e+02
c	108	14	4.7	1507	3	N50060	Sequence encoding int	1.14e+02
c	109	14	4.7	1507	9	Q50981	Human Interleukin-1 c	1.14e+02
c	110	14	4.7	1628	40	V20608	Human kynurenine-3-hy	1.14e+02
c	111	14	4.7	1642	37	T80796	Staphylococcus aureus	1.14e+02
c	112	14	4.7	1722	34	T84982	Human membrane antige	1.14e+02
c	113	14	4.7	1757	9	Q51034	Enhancer for the huma	1.14e+02
	114	14	4.7	1770	18	T14229	Mouse neural cell pro	1.14e+02
	115	14	4.7	1810	38	T84937	Human prostate protei	1.14e+02
	116	14	4.7	1810	4	Q25593	Sequence encoding hum	1.14e+02
	117	14	4.7	1813	12	Q72017	cDNA encoding recombi	1.14e+02
	118	14	4.7	1813	12	Q73382	Recombinant bacterici	1.14e+02
	119	14	4.7	1813	39	V14996	Human bactericidal/pe	1.14e+02
	120	14	4.7	1813	13	Q81445	Holoprotein rBPI50 ge	1.14e+02
	121	14	4.7	1813	39	V10337	Human recombinant BPI	1.14e+02
	122	14	4.7	1813	39	V13961	Human BPI protein enc	1.14e+02
	123	14	4.7	1813	38	V06404	Human bactericidal/pe	1.14e+02
	124	14	4.7	1813	39	T99564	Human bactericidal/pe	1.14e+02
	125	14	4.7	1813	15	Q97483	Recombinant bacterici	1.14e+02
	126	14	4.7	1813	16	Q99816	Recombinant bacterici	1.14e+02
	127	14	4.7	1813	13	Q85583	Human bactericidal-pe	1.14e+02
	128	14	4.7	1813	12	Q67270	Bactericide BPI.	1.14e+02
	129	14	4.7	1813	16	Q97607	Bactericidal/permeabi	1.14e+02
	130	14	4.7	1813	16	Q92640	DNA encoding recombin	1.14e+02
	131	14	4.7	1813	1	N92766	cDNA encoding human b	1.14e+02
	132	14	4.7	1813	16	Q86285	Gene encoding rBPI fo	1.14e+02
	133	14	4.7	1813	25	T31853	Bactericidal permeabi	1.14e+02
	134	14	4.7	1813	24	T39396	Recombinant bacterici	1.14e+02
	135	14	4.7	1813	9	Q52268	Human bactericidal/pe	1.14e+02
	136	14	4.7	1813	34	T67139	Human bactericidal/pe	1.14e+02
	137	14	4.7	1813	34	T67140	Bactericidal/permeabi	1.14e+02
	138	14	4.7	1813	17	T00951	Bactericidal/permeabi	1.14e+02
	139	14	4.7	1813	9	Q52488	pIN4512 coding regio	1.14e+02
	140	14	4.7	1813	14	Q80827	Bactericidal/permeabi	1.14e+02
c	141	14	4.7	1876	38	V12197	Human neuronal nicoti	1.14e+02
c	142	14	4.7	1876	28	T48239	Neuronal nicotinic ac	1.14e+02
c	143	14	4.7	1910	40	V16114	cDNA encoding a trans	1.14e+02
c	144	14	4.7	1910	11	Q63994	Human tissue transglu	1.14e+02

	145	14	4.7	1923	34	T91659	Prepro mistletoe lect	1.14e+02
	146	14	4.7	1974	39	V09703	Porcine retrovirus Ra	1.14e+02
	147	14	4.7	2007	4	Q27287	DNA encoding TAP.	1.14e+02
c	148	14	4.7	2012	12	Q71448	AtT 20 murine prohorm	1.14e+02
c	149	14	4.7	2012	7	Q43249	Sequence of AtT20 mur	1.14e+02
c	150	14	4.7	2020	23	T28190	Feline herpesvirus gl	1.14e+02
c	151	14	4.7	2021	2	Q11301	Leucocyte-plastin gen	1.14e+02
	152	14	4.7	2028	8	Q51230	Human adrenocorticotr	1.14e+02
c	153	14	4.7	2058	20	T27108	Hepatitis E virus (Me	1.14e+02
c	154	14	4.7	2100	8	Q46814	Mexico strain HEV ORF	1.14e+02
c	155	14	4.7	2100	7	Q47130	HEV third and second	1.14e+02
c	156	14	4.7	2153	7	Q41295	Human BMP-5 gene.	1.14e+02
c	157	14	4.7	2153	6	Q32854	BMP5.	1.14e+02
c	158	14	4.7	2153	1	Q06168	Encodes Human Bone Mo	1.14e+02
c	159	14	4.7	2153	4	Q23679	Encodes Human BMP-5.	1.14e+02
	160	14	4.7	2163	40	V24818	H. pylori ORF 14ep119	1.14e+02
c	161	14	4.7	2208	29	T62418	Z. japonica phosphoen	1.14e+02
	162	14	4.7	2214	20	T28807	Human G-protein coupl	1.14e+02
c	163	14	4.7	2222	11	Q68556	Murine pro-hormone co	1.14e+02
c	164	14	4.7	2268	29	T63576	Mouse H2Kb gene promo	1.14e+02
	165	14	4.7	2293	39	V09659	A. cellulolyticus El-	1.14e+02
	166	14	4.7	2326	4	N30174	Sequence of the Ll an	1.14e+02
c	167	14	4.7	2529	25	T42637	Class A starch branch	1.14e+02
c	168	14	4.7	2531	25	T17267	Class A starch branch	1.14e+02
	169	14	4.7	2574	34	T73717	Human filamin-like be	1.14e+02
c	170	14	4.7	2576	25	T42636	Class A starch branch	1.14e+02
c	171	14	4.7	2578	25	T42631	Class A starch branch	1.14e+02
	172	14	4.7	2606	38	V04319	Sugar beet protox-1 p	1.14e+02
	173	14	4.7	2606	36	T86134	Sugarbeet protoporphy	1.14e+02
c	174	14	4.7	2616	40	V09921	Human Arnt receptor c	1.14e+02
	175	14	4.7	2664	10	Q56760	Chitinase derivative	1.14e+02
	176	14	4.7	2734	8	Q47376	hETG mRNA.	1.14e+02
c	177	14	4.7	2734	6	Q39221	Encodes human epiderm	1.14e+02
c	178	14	4.7	2861	24	T28639	RNase L inhibitor *H2	1.14e+02
c	179	14	4.7	2975	25	T42635	Class A starch branch	1.14e+02
c	180	14	4.7	3003	25	T42634	Class A starch branch	1.14e+02
	181	14	4.7	3004	22	T12337	A. cellulolyticus El	1.14e+02
c	182	14	4.7	3074	31	T69587	Potato starch branchi	1.14e+02
c	183	14	4.7	3105	36	T94468	Human Fchd540 gene di	1.14e+02
	184	14	4.7	3215	8	Q48933	mOSF-2.	1.14e+02
c	185	14	4.7	3231	25	T42632	Class A starch branch	1.14e+02
c	186	14	4.7	3244	12	Q70642	L-plastin promoter an	1.14e+02
c	187	14	4.7	3257	4	Q27101	Endothelial tissue tr	1.14e+02
c	188	14	4.7	3373	31	T76768	Human alpha 1,2 fucos	1.14e+02
	189	14	4.7	3435	37	T95250	M. catarrhalis Q8 tra	1.14e+02
c	190	14	4.7	3478	38	V00291	Insecticidal protein	1.14e+02
c	191	14	4.7	3484	38	V00395	Insecticidal gene Cry	1.14e+02
c	192	14	4.7	3568	24	T28638	RNase L inhibitor *RL	1.14e+02
c	193	14	4.7	3791	29	T63575	Chicken beta-actin pr	1.14e+02
c	194	14	4.7	3900	12	Q66950	Cell death hid cDNA.	1.14e+02
	195	14	4.7	3907	35	T79968	Presenilin-interactin	1.14e+02

	196	14	4.7	3958	25	T43596	Glutamate-gated chlor	1.14e+02
	197	14	4.7	3981	1	N91734	Sequence of a macroph	1.14e+02
	198	14	4.7	3981	1	N70695	Macrophage-colony sti	1.14e+02
c	199	14	4.7	4000	33	T86166	Nucleotide sequence o	1.14e+02
	200	14	4.7	4021	2	Q10343	Monocyte macrophage c	1.14e+02
	201	14	4.7	4140	36	T77331	Solanum tuberosum col	1.14e+02
c	202	14	4.7	4171	18	Q74082	Oncostatin M receptor	1.14e+02
c	203	14	4.7	4255	23	T37051	Drosophila Fru 2 cDNA	1.14e+02
c	204	14	4.7	4300	4	Q25812	Clone 45-A.	1.14e+02
c	205	14	4.7	4704	1	Q05264	Sequence encoding bea	1.14e+02
	206	14	4.7	4708	3	N50415	Human acetyl choline	1.14e+02
c	207	14	4.7	5000	40	V20609	Human kynurenine-3-hy	1.14e+02
c	208	14	4.7	5236	4	Q25815	Clone 45-A; subtype 1	1.14e+02
c	209	14	4.7	5931	7	Q43701	Sequence of the human	1.14e+02
c	210	14	4.7	6008	7	Q41289	Ubiquitin-specific pr	1.14e+02
c	211	14	4.7	6154	10	Q56188	Wild-type Feline Herp	1.14e+02
c	212	14	4.7	6208	13	Q81210	Feline herpesvirus-1	1.14e+02
	213	14	4.7	6480	39	T99572	Mouse Friend virus su	1.14e+02
	214	14	4.7	6645	29	T65002	Mouse cell cycle regu	1.14e+02
c	215	14	4.7	7171	3	Q14413	Composite sequence fr	1.14e+02
c	216	14	4.7	7824	16	Q74052	Human interleukin-1-b	1.14e+02
c	217	14	4.7	8174	29	T61677	Human alpha(1,2)-fuco	1.14e+02
c	218	14	4.7	8174	2	Q13332	GDP-Fuc:beta-D-galact	1.14e+02
c	219	14	4.7	8174	9	Q56908	DNA encoding a glycos	1.14e+02
c	220	14	4.7	8789	40	V09076	Mitosin nucleic acid	1.14e+02
c	221	14	4.7	8789	16	Q86851	Human mitosin gene.	1.14e+02
	222	14	4.7	8920	11	Q62924	Carbamoyl-phosphate-s	1.14e+02
c	223	14	4.7	10136	22	T34578	Kinetochore protein C	1.14e+02
c	224	14	4.7	10607	6	Q37894	Beta-casein coding se	1.14e+02
c	225	14	4.7	12412	35	T33537	BCG deletion region 3	1.14e+02
	226	14	4.7	12687	16	Q98814	EAV GL gene.	1.14e+02
	227	14	4.7	12687	32	T70125	Equine arteritis viru	1.14e+02
c	228	14	4.7	12923	3	N90338	Sequence of human mus	1.14e+02
c	229	14	4.7	14507	32	T73568	Expression augmenting	1.14e+02
	230	14	4.7	14985	40	V15693	Human Bin1 gene.	1.14e+02
	231	14	4.7	14985	25	T43574	Human box-dependent m	1.14e+02
	232	14	4.7	16382	28	T63074	IgG-Fc binding protei	1.14e+02
	233	14	4.7	19307	20	T27558	Shuttle vector pADdel	1.14e+02
	234	14	4.7	30967	23	T32454	Calpain large subunit	1.14e+02
	235	13	4.4	20	6	Q39136	HCV Sense primer X(E2	4.37e+02
c	236	13	4.4	20	11	Q63515	NANBHV primer 1421R,	4.37e+02
	237	13	4.4	30	14	Q79797	Hepatitis C virus J1	4.37e+02
c	238	13	4.4	30	24	T32638	Primer MH93 for calci	4.37e+02
c	239	13	4.4	37	18	T03453	Human LAG-3 extracell	4.37e+02
c	240	13	4.4	41	11	Q624Z0	Vector pVAC1 construc	4.37e+02
c	241	13	4.4	42	21	T10349	CAEV env gene TM frag	4.37e+02
	242	13	4.4	48	3	Q14775	Rennin gene probe.	4.37e+02
c	243	13	4.4	50	33	T61564	scFv antibody library	4.37e+02
c	244	13	4.4	50	33	T61580	VH and scFv antibody	4.37e+02
c	245	13	4.4	56	6	Q39405	Human heavy chain var	4.37e+02
c	246	13	4.4	56	8	Q48995	Multimeric (SBP) anti	4.37e+02

c	247	13	4.4	56	7	Q42928	Primer HuVH2aBACKSfi.	4.37e+02
c	248	13	4.4	56	3	Q23707	Primer HuVH2aBACKSfi	4.37e+02
c	249	13	4.4	56	5	Q32344	HUVH2aBACKSfi, a VH b	4.37e+02
	250	13	4.4	83	5	Q33875	Sequence upstream fro	4.37e+02
c	251	13	4.4	90	6	Q36804	Mycobacterium 16S rRN	4.37e+02
c	252	13	4.4	91	8	Q54965	DPB-1 construct oligo	4.37e+02
	253	13	4.4	102	19	T19787	Human gene signature	4.37e+02
c	254	13	4.4	114	19	T23839	Human gene signature	4.37e+02
	255	13	4.4	118	19	T20349	Human gene signature	4.37e+02
	256	13	4.4	154	20	T21470	Human gene signature	4.37e+02
c	257	13	4.4	157	21	T10344	CAEV env gene fragmen	4.37e+02
c	258	13	4.4	165	6	Q36481	Mycoplasma primer/pro	4.37e+02
c	259	13	4.4	183	5	Q33806	Sequence upstream fro	4.37e+02
	260	13	4.4	194	21	T24733	Human gene signature	4.37e+02
	261	13	4.4	205	19	T20175	Human gene signature	4.37e+02
c	262	13	4.4	210	10	Q76486	Human genome fragment	4.37e+02
	263	13	4.4	216	2	Q13832	Part of grg-1/chymosi	4.37e+02
	264	13	4.4	235	10	Q76534	Human genome fragment	4.37e+02
	265	13	4.4	253	21	T24716	Human gene signature	4.37e+02
c	266	13	4.4	257	12	Q71375	DNA probe (Y2) used f	4.37e+02
c	267	13	4.4	258	19	T20083	Human gene signature	4.37e+02
	268	13	4.4	261	3	Q20726	Non-A non-B hepatitis	4.37e+02
c	269	13	4.4	270	10	Q77196	Human genome fragment	4.37e+02
c	270	13	4.4	273	6	Q36473	Mycoplasma primer/pro	4.37e+02
c	271	13	4.4	279	22	T25012	Human gene signature	4.37e+02
c	272	13	4.4	283	19	T20049	Human gene signature	4.37e+02
c	273	13	4.4	283	8	Q60960	Human brain Expressed	4.37e+02
	274	13	4.4	285	8	Q60928	Human brain Expressed	4.37e+02
	275	13	4.4	285	7	Q40454	E.coli malate dehydro	4.37e+02
	276	13	4.4	294	21	T24403	Human gene signature	4.37e+02
	277	13	4.4	294	10	Q77080	Human genome fragment	4.37e+02
c	278	13	4.4	294	20	T22499	Human gene signature	4.37e+02
c	279	13	4.4	301	22	T26029	Human gene signature	4.37e+02
c	280	13	4.4	301	19	T21136	Human gene signature	4.37e+02
	281	13	4.4	305	22	T26662	Human gene signature	4.37e+02
	282	13	4.4	309	13	Q73201	Korean-type Hepatitis	4.37e+02
	283	13	4.4	309	12	Q74718	KHCV E1G envelope pro	4.37e+02
	284	13	4.4	311	8	Q59792	Human brain Expressed	4.37e+02
c	285	13	4.4	320	8	Q59498	Human brain Expressed	4.37e+02
c	286	13	4.4	333	8	Q49724	DPB-1 locus synthetic	4.37e+02
	287	13	4.4	339	5	Q32455	HCV NS1(gp70) clone N	4.37e+02
	288	13	4.4	339	5	Q32454	HCV NS1(gp70) clone N	4.37e+02
	289	13	4.4	339	5	Q32456	HCV NS1(gp70) clone N	4.37e+02
c	290	13	4.4	343	40	V10262	Human COL4A5 gene 5'	4.37e+02
	291	13	4.4	351	14	Q79762	Hepatitis C virus J1	4.37e+02
	292	13	4.4	351	14	Q79751	Hepatitis C virus J1	4.37e+02
	293	13	4.4	351	14	Q79763	Hepatitis C virus J1	4.37e+02
c	294	13	4.4	353	18	T20977	Human gene signature	4.37e+02
	295	13	4.4	358	8	Q60102	Human brain Expressed	4.37e+02
c	296	13	4.4	359	21	T22621	Human gene signature	4.37e+02
	297	13	4.4	363	9	Q50972	Valpha TCR gene with	4.37e+02

c	298	13	4.4	368	37	V01471	Multiple HIV-epitope	4.37e+02
c	299	13	4.4	400	39	T88911	Arabidopsis thaliana	4.37e+02
	300	13	4.4	403	8	Q59260	Human brain Expressed	4.37e+02
	301	13	4.4	403	6	Q39848	Expressed Sequence Ta	4.37e+02
	302	13	4.4	405	6	Q39779	Expressed Sequence Ta	4.37e+02
	303	13	4.4	405	8	Q59191	Human brain Expressed	4.37e+02
c	304	13	4.4	405	3	Q21040	Region #2 specific to	4.37e+02
c	305	13	4.4	406	3	Q21039	Region #1 specific to	4.37e+02
	306	13	4.4	407	8	Q60970	Human brain Expressed	4.37e+02
	307	13	4.4	421	32	T49112	Partial DNA clone Aci	4.37e+02
c	308	13	4.4	436	10	Q56787	Gene tab3 DNA.	4.37e+02
c	309	13	4.4	471	6	Q39753	Expressed Sequence Ta	4.37e+02
c	310	13	4.4	471	8	Q59165	Human brain Expressed	4.37e+02
c	311	13	4.4	484	37	T91170	Secreted cytokine pro	4.37e+02
	312	13	4.4	495	40	V24735	H. pylori ORF hp2p106	4.37e+02
	313	13	4.4	523	7	Q40333	Sequence encoding gly	4.37e+02
	314	13	4.4	532	4	Q24835	Peroxidase isozyme ge	4.37e+02
c	315	13	4.4	534	10	Q56788	Gene tab5 DNA.	4.37e+02
c	316	13	4.4	549	30	T67475	H. pylori cytoplasmic	4.37e+02
	317	13	4.4	567	8	Q50050	Decorin (N-terminal t	4.37e+02
c	318	13	4.4	570	10	Q56789	Gene tab6 DNA.	4.37e+02
	319	13	4.4	576	14	Q83854	Hepatitis C virus env	4.37e+02
	320	13	4.4	576	14	Q83845	Hepatitis C virus env	4.37e+02
	321	13	4.4	576	21	T16567	Hepatitis C virus iso	4.37e+02
	322	13	4.4	576	21	T16562	Hepatitis C virus iso	4.37e+02
	323	13	4.4	576	21	T16571	Hepatitis C virus iso	4.37e+02
	324	13	4.4	576	14	Q83846	Hepatitis C virus env	4.37e+02
	325	13	4.4	576	14	Q83847	Hepatitis C virus env	4.37e+02
	326	13	4.4	576	14	Q83853	Hepatitis C virus env	4.37e+02
	327	13	4.4	576	14	Q83852	Hepatitis C virus env	4.37e+02
	328	13	4.4	576	21	T16581	Hepatitis C virus iso	4.37e+02
	329	13	4.4	576	21	T16580	Hepatitis C virus iso	4.37e+02
	330	13	4.4	576	14	Q83869	Hepatitis C virus env	4.37e+02
	331	13	4.4	576	21	T16560	Hepatitis C virus iso	4.37e+02
	332	13	4.4	576	21	T16561	Hepatitis C virus iso	4.37e+02
	333	13	4.4	576	14	Q83863	Hepatitis C virus env	4.37e+02
	334	13	4.4	576	21	T16578	Hepatitis C virus iso	4.37e+02
	335	13	4.4	576	21	T16583	Hepatitis C virus iso	4.37e+02
	336	13	4.4	576	14	Q83867	Hepatitis C virus env	4.37e+02
	337	13	4.4	576	14	Q83865	Hepatitis C virus env	4.37e+02
	338	13	4.4	576	21	T16579	Hepatitis C virus iso	4.37e+02
	339	13	4.4	576	21	T16568	Hepatitis C virus iso	4.37e+02
	340	13	4.4	576	14	Q83848	Hepatitis C virus env	4.37e+02
	341	13	4.4	576	21	T16566	Hepatitis C virus iso	4.37e+02
	342	13	4.4	576	14	Q83866	Hepatitis C virus env	4.37e+02
	343	13	4.4	576	14	Q83864	Hepatitis C virus env	4.37e+02
	344	13	4.4	576	21	T16559	Hepatitis C virus iso	4.37e+02
	345	13	4.4	576	21	T16565	Hepatitis C virus iso	4.37e+02
	346	13	4.4	576	14	Q83851	Hepatitis C virus env	4.37e+02
	347	13	4.4	576	21	T16576	Hepatitis C virus iso	4.37e+02
	348	13	4.4	576	14	Q83862	Hepatitis C virus env	4.37e+02

	349	13	4.4	576	21	T16577	Hepatitis C virus iso 4.37e+02
	350	13	4.4	576	14	Q83857	Hepatitis C virus env 4.37e+02
	351	13	4.4	583	37	T96577	Human bcl-w DNA. 4.37e+02
c	352	13	4.4	594	30	T67904	H. pylori cytoplasmic 4.37e+02
	353	13	4.4	606	18	T19008	Human gene signature 4.37e+02
	354	13	4.4	619	6	Q36046	Human sequence XXVII 4.37e+02
	355	13	4.4	636	18	T18245	HCV1 E1 gene. 4.37e+02
	356	13	4.4	642	21	T12704	HCV E1 construct HCCI 4.37e+02
	357	13	4.4	645	6	Q34948	Altered HBcAg coding 4.37e+02
	358	13	4.4	645	6	Q34945	Altered HBcAg coding 4.37e+02
	359	13	4.4	645	6	Q35102	Altered HBcAg coding 4.37e+02
	360	13	4.4	645	6	Q34996	Modified hepatitis B 4.37e+02
	361	13	4.4	645	6	Q34999	Modified hepatitis B 4.37e+02
	362	13	4.4	645	6	Q35105	Altered HBcAg coding 4.37e+02
	363	13	4.4	652	5	Q32469	gp70 clone N27N19-1. 4.37e+02
	364	13	4.4	656	2	Q06395	Exon 3 of ovine leuka 4.37e+02
c	365	13	4.4	661	21	T30568	Eukaryotic cell growt 4.37e+02
c	366	13	4.4	672	14	Q82874	Human STX sialyltrans 4.37e+02
	367	13	4.4	711	8	Q50051	Decorin (N-terminal t 4.37e+02
	368	13	4.4	717	1	N93192	Adenine-phosphoribosy 4.37e+02
	369	13	4.4	723	21	T12961	HCV E1 construct HCCI 4.37e+02
	370	13	4.4	742	3	Q20926	C10-E15 DNA fragment 4.37e+02
	371	13	4.4	747	9	Q50975	scValphaVbetapelBhis 4.37e+02
c	372	13	4.4	752	1	Q04691	Encodes Mammalian ami 4.37e+02
	373	13	4.4	753	9	Q50974	scValphaVbeta with (h 4.37e+02
	374	13	4.4	755	5	Q32449	HCV core-envelope clo 4.37e+02
	375	13	4.4	759	30	T68254	H. pylori cytoplasmic 4.37e+02
	376	13	4.4	772	2	N60203	Interferon-pseudo-ome 4.37e+02
	377	13	4.4	780	9	Q55089	fbp1 cDNA for reverse 4.37e+02
c	378	13	4.4	783	2	N93066	Clone encoding bovine 4.37e+02
	379	13	4.4	795	21	T12705	HCV E1 construct HCCI 4.37e+02
	380	13	4.4	849	8	Q50052	Decorin (N-terminal t 4.37e+02
	381	13	4.4	886	14	Q80013	Porcine beta-myosin H 4.37e+02
	382	13	4.4	888	2	N60551	Mouse kidney cell Ban 4.37e+02
c	383	13	4.4	916	14	Q82740	Mouse tissue inhibito 4.37e+02
	384	13	4.4	920	38	V06111	Viral infection gene 4.37e+02
	385	13	4.4	924	8	Q50053	Decorin (N-terminal t 4.37e+02
	386	13	4.4	932	3	Q20923	C10-E12 DNA fragment 4.37e+02
c	387	13	4.4	944	13	Q85037	DNA-ase-B gene for fu 4.37e+02
	388	13	4.4	945	21	T12970	HCV E1 construct HCCI 4.37e+02
c	389	13	4.4	946	24	T38230	Apoptosis-regulating 4.37e+02
	390	13	4.4	950	9	Q51001	LACI gene. 4.37e+02
c	391	13	4.4	960	3	Q22604	Gene encoding OMP of 4.37e+02
	392	13	4.4	961	21	T12971	HCV E1 construct HCCI 4.37e+02
	393	13	4.4	994	16	Q87015	Clone B1 encodes xylo 4.37e+02
	394	13	4.4	1002	8	Q50046	Mature decorin PT-65 4.37e+02
	395	13	4.4	1026	21	T12968	HCV E1 construct HCCI 4.37e+02
	396	13	4.4	1026	20	T10741	Human recombinant dec 4.37e+02
	397	13	4.4	1034	21	T12969	HCV E1 construct HCCI 4.37e+02
	398	13	4.4	1059	34	T85163	Macaque chemokine rec 4.37e+02
	399	13	4.4	1062	20	T09951	High-affinity melaton 4.37e+02

	400	13	4.4	1065	33	T86154	Human MIP-1alpha/RANT 4.37e+02
	401	13	4.4	1107	2	Q12240	Clone 136/155 encodin 4.37e+02
c	402	13	4.4	1120	2	N70643	Cycloheximide resista 4.37e+02
c	403	13	4.4	1125	17	T08145	Xylanase xynC gene (p 4.37e+02
	404	13	4.4	1128	31	T68936	Human p32 11-cis-reti 4.37e+02
	405	13	4.4	1164	2	Q12871	Polyhedrin-soluble LA 4.37e+02
	406	13	4.4	1167	2	N60289	Sequence encoding hum 4.37e+02
c	407	13	4.4	1167	13	Q89352	Acetobacter pasteuria 4.37e+02
	408	13	4.4	1173	2	Q11043	Encodes equine C-term 4.37e+02
	409	13	4.4	1173	2	Q11042	Probe to equine C-ter 4.37e+02
	410	13	4.4	1175	3	N40295	Sequence encoding a p 4.37e+02
	411	13	4.4	1175	3	Q20949	Prochymosin (prorenni 4.37e+02
	412	13	4.4	1207	7	Q40329	Sequence encoding gly 4.37e+02
	413	13	4.4	1207	7	Q40331	Sequence encoding gly 4.37e+02
	414	13	4.4	1207	7	Q40325	Sequence encoding gly 4.37e+02
	415	13	4.4	1207	7	Q40326	Sequence encoding gly 4.37e+02
	416	13	4.4	1207	7	Q40328	Sequence encoding gly 4.37e+02
	417	13	4.4	1207	7	Q40330	Sequence encoding gly 4.37e+02
	418	13	4.4	1210	6	Q32983	HCV E2/NS1 region. 4.37e+02
c	419	13	4.4	1218	4	Q24197	Transglutaminase (Act 4.37e+02
	420	13	4.4	1230	1	N80251	H-env probe for detec 4.37e+02
	421	13	4.4	1236	5	Q32441	gp70 clone N27MX24B-1 4.37e+02
	422	13	4.4	1240	18	T03006	Chymosin open reading 4.37e+02
	423	13	4.4	1244	18	T08142	Xylanase xynD gene. 4.37e+02
	424	13	4.4	1249	14	Q79772	Hepatitis C virus cor 4.37e+02
	425	13	4.4	1251	4	Q25610	HCV in expression vec 4.37e+02
	426	13	4.4	1251	4	Q26981	HCV gene 1. 4.37e+02
	427	13	4.4	1258	5	Q32452	HCV core-envelope clo 4.37e+02
	428	13	4.4	1258	5	Q32450	HCV core-envelope clo 4.37e+02
	429	13	4.4	1260	10	Q58004	NANBHV plasmid pIK4CE 4.37e+02
	430	13	4.4	1260	7	Q40237	NonA-nonB hepatitis v 4.37e+02
	431	13	4.4	1260	11	Q71677	Non-A non-B hepatitis 4.37e+02
	432	13	4.4	1263	10	Q56821	Non A Non B hepatitis 4.37e+02
c	433	13	4.4	1271	23	T34991	Vascular IBP-like gro 4.37e+02
c	434	13	4.4	1273	30	T64550	Activated platelet pr 4.37e+02
	435	13	4.4	1278	3	N40055	Sequence of prochymos 4.37e+02
	436	13	4.4	1289	16	Q87016	Clone B2 encodes func 4.37e+02
	437	13	4.4	1289	4	N30022	Sequence of veal chym 4.37e+02
	438	13	4.4	1291	1	N91157	Cloned sequence of (p 4.37e+02
c	439	13	4.4	1320	10	Q55424	CZP-3 Gene. 4.37e+02
c	440	13	4.4	1322	4	Q24201	Transglutaminase (Str 4.37e+02
c	441	13	4.4	1325	11	Q65609	Canine zona pellucida 4.37e+02
	442	13	4.4	1329	1	N92017	DNA encoding peptide 4.37e+02
	443	13	4.4	1338	4	Q27003	HK1. 4.37e+02
	444	13	4.4	1350	14	Q82828	VP1 protein gene. 4.37e+02
c	445	13	4.4	1350	14	Q82828	VP1 protein gene. 4.37e+02
c	446	13	4.4	1373	30	T64549	Human activated plate 4.37e+02
	447	13	4.4	1391	26	T28404	E. coli detection pro 4.37e+02
	448	13	4.4	1395	21	T12972	HCV E1 construct HCCL 4.37e+02
	449	13	4.4	1396	1	N80129	LinA' gene-contg sequ 4.37e+02
	450	13	4.4	1398	3	Q22140	Hepatitis C virus str 4.37e+02

	451	13	4.4	1401	6	Q38261	HCV CKS-3'ENV recombi 4.37e+02
	452	13	4.4	1401	6	Q38246	HCV CKS-3'ENV recombi 4.37e+02
	453	13	4.4	1401	6	Q38276	HCV CKS-3'ENV recombi 4.37e+02
	454	13	4.4	1402	33	T61549	Plant plastid phospho 4.37e+02
	455	13	4.4	1404	3	Q21688	pHCV-101 - recombinan 4.37e+02
	456	13	4.4	1404	8	Q45079	Human glucokinase gen 4.37e+02
	457	13	4.4	1450	40	V18820	Human hyaluronidase c 4.37e+02
	458	13	4.4	1460	3	N40180	Sequence of recombina 4.37e+02
	459	13	4.4	1460	4	N20043	Pre-prorennin-A gene 4.37e+02
	460	13	4.4	1477	21	T12967	HCV E2 construct HCCL 4.37e+02
	461	13	4.4	1495	11	Q62695	C-C chemokine recepto 4.37e+02
c	462	13	4.4	1497	22	T06750	Resynthesised jojoba 4.37e+02
	463	13	4.4	1500	35	T91970	DNA encoding partial 4.37e+02
	464	13	4.4	1511	4	Q23524	ORF1 encoding Abi 105 4.37e+02
c	465	13	4.4	1519	16	T11369	Alpha-Amylo-C cDNA. 4.37e+02
c	466	13	4.4	1538	3	Q21732	DNA encoding a RAP-2 4.37e+02
c	467	13	4.4	1540	39	V10136	Stealth virus nucleic 4.37e+02
c	468	13	4.4	1541	5	Q32671	Primer SK43' and SK44 4.37e+02
c	469	13	4.4	1551	21	T18378	3-hydroxybutyric acid 4.37e+02
	470	13	4.4	1554	5	Q32451	HCV core-envelope clo 4.37e+02
c	471	13	4.4	1578	39	V09401	Human ICE LAP-6 polyp 4.37e+02
	472	13	4.4	1593	1	Q04491	Sequence encoding hum 4.37e+02
	473	13	4.4	1598	7	Q40461	E.coli mdh promoter f 4.37e+02
c	474	13	4.4	1636	40	V10435	Human chitinase clone 4.37e+02
c	475	13	4.4	1637	38	T89180	Human chitotriosidase 4.37e+02
c	476	13	4.4	1643	27	T50833	Human chitinase cDNA 4.37e+02
c	477	13	4.4	1647	35	T91036	Human checkpoint gene 4.37e+02
c	478	13	4.4	1647	9	Q52272	Human RADcompB cDNA s 4.37e+02
	479	13	4.4	1649	7	Q41064	Clone derived when ch 4.37e+02
	480	13	4.4	1655	7	Q43720	Sequence which hybrid 4.37e+02
c	481	13	4.4	1656	40	V10436	Human chitinase clone 4.37e+02
	482	13	4.4	1667	4	Q33027	Exon 5 from the UGT1 4.37e+02
	483	13	4.4	1677	35	T91971	DNA encoding partial 4.37e+02
c	484	13	4.4	1686	10	Q56308	Murine 5HT5a serotoni 4.37e+02
	485	13	4.4	1687	35	T39798	Human clone 5 gene. 4.37e+02
	486	13	4.4	1690	14	Q81950	Prostaglandin-EP3-21 4.37e+02
c	487	13	4.4	1713	27	T50834	Human chitinase cDNA 4.37e+02
	488	13	4.4	1734	7	Q40439	Hepatitis C virus clo 4.37e+02
	489	13	4.4	1734	6	Q40430	Hepatitis C virus clo 4.37e+02
	490	13	4.4	1734	6	Q40433	Hepatitis C virus clo 4.37e+02
	491	13	4.4	1734	6	Q40432	Hepatitis C virus clo 4.37e+02
	492	13	4.4	1734	6	Q40428	Hepatitis C virus clo 4.37e+02
	493	13	4.4	1734	7	Q40438	Hepatitis C virus clo 4.37e+02
	494	13	4.4	1734	6	Q40429	Hepatitis C virus clo 4.37e+02
	495	13	4.4	1734	7	Q40437	Hepatitis C virus clo 4.37e+02
	496	13	4.4	1734	6	Q40431	Hepatitis C virus clo 4.37e+02
	497	13	4.4	1734	6	Q40425	Hepatitis C virus clo 4.37e+02
	498	13	4.4	1734	7	Q40436	Hepatitis C virus clo 4.37e+02
	499	13	4.4	1747	23	T35308	Rat M1PTP gene cDNA c 4.37e+02
	500	13	4.4	1765	14	Q79141	Hepatitis C virus gen 4.37e+02
	501	13	4.4	1765	14	Q79143	Hepatitis C virus gen 4.37e+02

c	502	13	4.4	1768	38	T89181	Human chitotriosidase 4.37e+02
c	503	13	4.4	1782	22	T18119	Porphyromonas gingiva 4.37e+02
c	504	13	4.4	1782	17	T04121	P. gingivalis cell su 4.37e+02
c	505	13	4.4	1794	23	T33873	Candida RAPT1 cDNA. 4.37e+02
c	506	13	4.4	1800	16	Q91769	Coding sequence of PE 4.37e+02
c	507	13	4.4	1801	31	T76552	DNA encoding hexose o 4.37e+02
c	508	13	4.4	1815	17	Q97809	Sequence encoding end 4.37e+02
c	509	13	4.4	1827	31	T63336	Human host cell prote 4.37e+02
	510	13	4.4	1851	6	Q38264	HCV CKS-ENV recombina 4.37e+02
	511	13	4.4	1851	6	Q38249	HCV CKS-ENV recombina 4.37e+02
	512	13	4.4	1851	6	Q38279	HCV CKS-ENV recombina 4.37e+02
	513	13	4.4	1854	3	Q21691	pHCV-104 - recombinan 4.37e+02
	514	13	4.4	1863	3	Q15362	Fragment of NANB hepa 4.37e+02
c	515	13	4.4	1864	15	Q78892	Aspergillus nidulans 4.37e+02
c	516	13	4.4	1864	2	N71405	Sequence of ANS-1 whi 4.37e+02
	517	13	4.4	1871	2	Q12869	Lymphocyte Activation 4.37e+02
	518	13	4.4	1872	34	T91477	Mycobacterium tubercu 4.37e+02
	519	13	4.4	1872	34	T91414	Mycobacterium tubercu 4.37e+02
	520	13	4.4	1876	12	Q71053	Rabbit C7-alpha-H cor 4.37e+02
	521	13	4.4	1880	4	Q24466	NANB hepatitis virus 4.37e+02
	522	13	4.4	1892	16	Q74063	The mouse beta-actin 4.37e+02
c	523	13	4.4	1915	38	V12198	Human neuronal nicoti 4.37e+02
c	524	13	4.4	1915	20	T48241	Neuronal nicotinic ac 4.37e+02
c	525	13	4.4	1929	35	T72642	Human testis-specific 4.37e+02
	526	13	4.4	1938	40	V24723	H. pylori ORF hplp138 4.37e+02
c	527	13	4.4	1952	11	Q66642	Mouse Activin recepto 4.37e+02
c	528	13	4.4	1952	16	T06032	Mouse ALK-6 cDNA. 4.37e+02
c	529	13	4.4	1953	29	T68316	cDNA encoding human h 4.37e+02
c	530	13	4.4	1953	22	T28769	Human homologue of ba 4.37e+02
c	531	13	4.4	1965	18	T13168	Human interleukin-1 t 4.37e+02
c	532	13	4.4	1975	1	Q05057	DraI/DraI fragment of 4.37e+02
	533	13	4.4	1981	30	T66241	Romaine lettuce viola 4.37e+02
	534	13	4.4	1983	11	Q66634	Human Activin recepto 4.37e+02
	535	13	4.4	1989	18	T12446	HCV1 E2/NS2 region DN 4.37e+02
	536	13	4.4	2006	12	Q50147	Phospholipase D-P gen 4.37e+02
c	537	13	4.4	2019	1	Q03730	Human carbastatin pol 4.37e+02
	538	13	4.4	2033	15	Q86788	Hepatitis C virus T7N 4.37e+02
	539	13	4.4	2033	11	Q64913	Hepatitis C Virus gen 4.37e+02
	540	13	4.4	2045	7	Q43374	Sequence of rat fibro 4.37e+02
	541	13	4.4	2047	1	Q03451	Recombinant DNA molec 4.37e+02
	542	13	4.4	2051	3	Q14944	Rhizopus PGK1 promote 4.37e+02
c	543	13	4.4	2059	2	N70271	DNA fragment for tran 4.37e+02
c	544	13	4.4	2070	14	Q82873	Human ST30 sialyltran 4.37e+02
c	545	13	4.4	2076	14	Q83531	Sequence encoding hon 4.37e+02
	546	13	4.4	2080	1	Q03708	Gene encoding murine 4.37e+02
	547	13	4.4	2082	9	Q52650	Sequence encoding mur 4.37e+02
c	548	13	4.4	2085	18	T10252	Human calpastatin cod 4.37e+02
	549	13	4.4	2086	21	T12973	HCV E1 construct HCCI 4.37e+02
c	550	13	4.4	2101	16	Q98750	DNA encoding human so 4.37e+02
c	551	13	4.4	2101	14	Q83528	P. falciparum EBL-e2 4.37e+02
c	552	13	4.4	2101	32	T72896	Plasmodium ebl-2 gene 4.37e+02

	553	13	4.4	2116	2	Q12242	Encodes PT-NANBH vira 4.37e+02
c	554	13	4.4	2125	16	Q91762	Fission yeast Rad51 g 4.37e+02
c	555	13	4.4	2148	2	N92751	DNA encoding human ca 4.37e+02
	556	13	4.4	2156	2	Q11562	Partial endothelial P 4.37e+02
	557	13	4.4	2156	34	T90384	Human MIP-1 alpha/RAN 4.37e+02
	558	13	4.4	2183	9	Q54769	T. parvi sporozoite 6 4.37e+02
c	559	13	4.4	2206	8	Q46161	DNA encoding tumour a 4.37e+02
c	560	13	4.4	2206	8	Q48170	Sequence encoding a 9 4.37e+02
c	561	13	4.4	2206	8	Q46888	Sequence encoding an 4.37e+02
	562	13	4.4	2214	33	T84275	Rat ICAM-4 cDNA RT-PC 4.37e+02
c	563	13	4.4	2226	2	N92429	Sequence encoding amy 4.37e+02
c	564	13	4.4	2264	17	T15542	Borrelia afzelii 66 k 4.37e+02
c	565	13	4.4	2285	7	Q40574	Gp85-97 clone 18 cDNA 4.37e+02
c	566	13	4.4	2291	30	T67198	Bovine brain Ca2+/cal 4.37e+02
c	567	13	4.4	2291	26	T51094	61 kD brain calcium/c 4.37e+02
c	568	13	4.4	2291	5	Q30163	pCAM-40 clone contg. 4.37e+02
c	569	13	4.4	2291	14	Q83957	Cyclic-GMP stimulated 4.37e+02
	570	13	4.4	2298	17	T10911	Sequence encoding VP1 4.37e+02
	571	13	4.4	2298	4	Q24516	Chicken anaemia virus 4.37e+02
c	572	13	4.4	2298	4	Q24516	Chicken anaemia virus 4.37e+02
c	573	13	4.4	2298	17	T10911	Sequence encoding VP1 4.37e+02
	574	13	4.4	2318	30	T59194	Chicken anaemia virus 4.37e+02
c	575	13	4.4	2318	30	T59194	Chicken anaemia virus 4.37e+02
	576	13	4.4	2319	17	T15871	Chicken anaemia virus 4.37e+02
	577	13	4.4	2319	3	Q22738	Chicken anaemia virus 4.37e+02
c	578	13	4.4	2319	17	T15871	Chicken anaemia virus 4.37e+02
c	579	13	4.4	2319	3	Q22738	Chicken anaemia virus 4.37e+02
c	580	13	4.4	2320	4	Q26694	pag C. 4.37e+02
c	581	13	4.4	2320	5	Q27739	pagC/AP fusion region 4.37e+02
c	582	13	4.4	2320	31	T70212	S. typhimurium pagC g 4.37e+02
c	583	13	4.4	2320	13	Q85099	PagC gene. 4.37e+02
	584	13	4.4	2337	39	T94212	Thermotoga maritima e 4.37e+02
	585	13	4.4	2351	4	Q27369	HUG-Br1. 4.37e+02
c	586	13	4.4	2352	33	T60798	ORFs 2-5 of PRRSV iso 4.37e+02
	587	13	4.4	2358	2	Q07000	Coding part of vector 4.37e+02
	588	13	4.4	2368	4	Q27370	HUG-Br2. 4.37e+02
	589	13	4.4	2433	21	T12974	HCV E1 construct HCCI 4.37e+02
	590	13	4.4	2440	2	Q06469	Sequence encoding inc 4.37e+02
	591	13	4.4	2451	37	V01876	Human NPIK gene. 4.37e+02
c	592	13	4.4	2457	22	T30736	Rat cryptdin 1 gene. 4.37e+02
c	593	13	4.4	2465	2	Q13136	Hyp operon. 4.37e+02
	594	13	4.4	2475	1	Q04780	Sequence encoding cAM 4.37e+02
	595	13	4.4	2478	1	N91246	Sequence of B. thurin 4.37e+02
	596	13	4.4	2484	3	Q15360	NP-YA. 4.37e+02
	597	13	4.4	2487	37	V01878	Human NPIK gene. 4.37e+02
c	598	13	4.4	2517	22	T36079	Interleukin binding f 4.37e+02
	599	13	4.4	2529	9	Q50499	Promoterless neo/aprt 4.37e+02
	600	13	4.4	2540	7	Q43888	NANB hepatitis virus 4.37e+02
	601	13	4.4	2540	5	Q29627	Hepatitis C virus HC- 4.37e+02
	602	13	4.4	2540	11	Q63752	NANBHV genomic fragme 4.37e+02
c	603	13	4.4	2556	38	V02872	Human hTAFII105 cDNA. 4.37e+02

c	604	13	4.4	2575	15	Q90331	Human cholesterol 7al 4.37e+02
c	605	13	4.4	2575	15	Q89558	Human CYP7 clone, lam 4.37e+02
c	606	13	4.4	2577	19	T18677	Human tastin cDNA. 4.37e+02
	607	13	4.4	2602	33	T86255	Mouse Rhotekin cDNA. 4.37e+02
	608	13	4.4	2629	1	N81338	Protein A gene with n 4.37e+02
	609	13	4.4	2629	4	Q24523	Vasoactive intestinal 4.37e+02
	610	13	4.4	2631	21	T13655	AcNPV ORF 50, residue 4.37e+02
c	611	13	4.4	2638	40	V20468	Human B-myb oncogene. 4.37e+02
c	612	13	4.4	2638	1	Q03741	Human myb related gen 4.37e+02
	613	13	4.4	2646	2	Q13349	DNA expressed during 4.37e+02
c	614	13	4.4	2676	13	Q76371	Mouse growth differen 4.37e+02
	615	13	4.4	2680	36	T92160	Human deleted in panc 4.37e+02
c	616	13	4.4	2729	35	T91062	Mouse gene fomy030 co 4.37e+02
c	617	13	4.4	2729	24	T41469	Mouse 030 gene (fomy0 4.37e+02
	618	13	4.4	2763	8	Q46037	Plasmid pSF1. 4.37e+02
	619	13	4.4	2763	8	Q48044	Lactobacillus lactis 4.37e+02
	620	13	4.4	2788	9	Q50342	Human haematopoietic 4.37e+02
	621	13	4.4	2817	39	V01590	Rat glycine transport 4.37e+02
c	622	13	4.4	2838	7	Q41087	Insulinoma-associated 4.37e+02
	623	13	4.4	2855	39	V00011	Human sex comb on mid 4.37e+02
	624	13	4.4	2855	39	T97297	Human sex comb on mid 4.37e+02
c	625	13	4.4	2880	2	N92068	Human muscarinic acet 4.37e+02
c	626	13	4.4	2880	1	Q02068	Human muscarinic acet 4.37e+02
c	627	13	4.4	2890	34	T89627	Genomic DNA encoding 4.37e+02
	628	13	4.4	2902	1	Q06633	Clone encoding rat ov 4.37e+02
c	629	13	4.4	2932	4	Q25388	TXA2 receptor gene. 4.37e+02
c	630	13	4.4	2939	22	T36080	Interleukin binding f 4.37e+02
	631	13	4.4	2975	11	Q65587	Heat resistant aldehy 4.37e+02
	632	13	4.4	2988	33	T84245	Rat ICAM-4 cDNA. 4.37e+02
c	633	13	4.4	2990	1	Q04123	Adenyl cyclase gene o 4.37e+02
c	634	13	4.4	2992	31	T71087	Yeast alanyl-tRNA syn 4.37e+02
	635	13	4.4	2996	2	Q12366	Gene encoding enzyme 4.37e+02
c	636	13	4.4	3031	17	T05112	Shc gene. 4.37e+02
c	637	13	4.4	3031	29	T60305	Human Shc1 locus. 4.37e+02
c	638	13	4.4	3039	16	Q91096	Human herpesvirus-6 M 4.37e+02
	639	13	4.4	3086	2	Q06470	Sequence encoding com 4.37e+02
c	640	13	4.4	3102	40	V17260	Human protein p164 co 4.37e+02
	641	13	4.4	3196	39	V01453	Streptomyces daunomyc 4.37e+02
	642	13	4.4	3196	39	V01447	Daunomycin C-14 hydro 4.37e+02
	643	13	4.4	3200	2	Q05638	Sequence which hybrdi 4.37e+02
	644	13	4.4	3226	2	Q11038	Encodes Rat C-termina 4.37e+02
c	645	13	4.4	3230	39	V09641	Human membrane protei 4.37e+02
	646	13	4.4	3233	12	Q72690	Retinoblastoma 94kD t 4.37e+02
	647	13	4.4	3255	39	T97299	Human sex comb on mid 4.37e+02
	648	13	4.4	3255	39	V00013	Human sex comb on mid 4.37e+02
	649	13	4.4	3290	2	Q10278	Type B alpha-amidatin 4.37e+02
c	650	13	4.4	3321	40	V04207	Human metabotropic gl 4.37e+02
c	651	13	4.4	3321	40	V04206	Human metabotropic gl 4.37e+02
	652	13	4.4	3324	37	V01879	Human NPIK gene. 4.37e+02
	653	13	4.4	3327	39	V00012	Human sex comb on mid 4.37e+02
	654	13	4.4	3327	39	T97298	Human sex comb on mid 4.37e+02

	655	13	4.4	3360	19	T03677	Hepatitis C genome fr 4.37e+02
c	656	13	4.4	3370	13	Q79534	Bovine tracheal antim 4.37e+02
	657	13	4.4	3401	11	Q64069	Non-A, non-B hepatiti 4.37e+02
	658	13	4.4	3401	20	T30387	5'UTR/CORE/ENV/NS1/NS 4.37e+02
	659	13	4.4	3415	22	T36481	Human integrin beta s 4.37e+02
	660	13	4.4	3458	3	Q14256	51.4 and 41.9 kD toxi 4.37e+02
	661	13	4.4	3458	8	Q48714	Insecticidal protein 4.37e+02
	662	13	4.4	3461	11	Q64068	Non-A, non-B hepatiti 4.37e+02
	663	13	4.4	3461	20	T30386	5'UTR/CORE/ENV/NS1/NS 4.37e+02
c	664	13	4.4	3479	38	V12284	Saccharomyces cerevis 4.37e+02
	665	13	4.4	3508	15	Q86693	Candida tropicalis pK 4.37e+02
c	666	13	4.4	3528	8	Q47292	Delta endotoxin gene. 4.37e+02
c	667	13	4.4	3531	13	Q72786	Bacillus thuringiensi 4.37e+02
c	668	13	4.4	3531	38	V00293	Bacillus thuringiensi 4.37e+02
c	669	13	4.4	3534	1	Q06561	Encodes full-length i 4.37e+02
c	670	13	4.4	3537	2	Q10179	Lepidopteran-active t 4.37e+02
c	671	13	4.4	3537	4	Q26929	DNA encoding the nove 4.37e+02
c	672	13	4.4	3537	1	Q06831	Sequence encoding B.t 4.37e+02
c	673	13	4.4	3565	14	Q84051	Sequence encoding mel 4.37e+02
c	674	13	4.4	3567	1	Q05797	Encodes B.t.entomocid 4.37e+02
	675	13	4.4	3586	2	Q05637	Sequence which hybrdi 4.37e+02
	676	13	4.4	3586	2	Q10277	Type A alpha-amidatin 4.37e+02
c	677	13	4.4	3596	9	Q52699	Corticosterone-bindin 4.37e+02
	678	13	4.4	3602	37	V01877	Human NPIK gene. 4.37e+02
c	679	13	4.4	3623	1	N81647	Sequence encoding mod 4.37e+02
c	680	13	4.4	3623	1	N81335	Sequence of gene enco 4.37e+02
c	681	13	4.4	3623	2	N90905	Sequence encoding ins 4.37e+02
	682	13	4.4	3633	37	T97611	Mouse E2A-binding pro 4.37e+02
c	683	13	4.4	3659	17	T01469	Plasmid pRW731.15 Pvu 4.37e+02
c	684	13	4.4	3659	18	T04699	pRW731.15 (contg.FPV 4.37e+02
c	685	13	4.4	3659	12	Q67858	Fowlpox PvuII-EcoRV f 4.37e+02
c	686	13	4.4	3659	18	T00518	TROVAC DNA fragment. 4.37e+02
c	687	13	4.4	3660	7	Q45687	Recombinant fowlpox v 4.37e+02
c	688	13	4.4	3661	5	Q29108	Fowlpox PvuII fragmen 4.37e+02
c	689	13	4.4	3661	6	Q36873	Fragment of TROVAC DN 4.37e+02
c	690	13	4.4	3706	17	T04688	Black widow spider de 4.37e+02
c	691	13	4.4	3717	10	Q56917	Glutamic acid recepto 4.37e+02
c	692	13	4.4	3720	13	Q80513	Genetic locus bcl-6. 4.37e+02
	693	13	4.4	3728	10	Q58708	Mouse OSF-5 coding se 4.37e+02
c	694	13	4.4	3786	5	Q29252	Human beta-8 intergin 4.37e+02
	695	13	4.4	3792	8	Q48463	Plasmid pG+host4 cont 4.37e+02
c	696	13	4.4	3807	21	T30862	Protein tyrosine kina 4.37e+02
c	697	13	4.4	3826	37	T86635	CryIA gene encoding a 4.37e+02
c	698	13	4.4	3826	11	Q66292	Crystal delta-endotox 4.37e+02
c	699	13	4.4	3826	25	T49478	Bacillus thuringiensi 4.37e+02
	700	13	4.4	3831	3	Q20262	ERK3. 4.37e+02
c	701	13	4.4	3833	40	V17116	Metabotropic glutamat 4.37e+02
c	702	13	4.4	3856	1	Q03097	Enod2b genomic clone. 4.37e+02
	703	13	4.4	3861	9	Q51543	Lipopolysaccharide in 4.37e+02
	704	13	4.4	3914	10	Q44393	Sequence of human OSF 4.37e+02
	705	13	4.4	3957	18	T14077	Human dihydropyrimidi 4.37e+02

c	706	13	4.4	4035	1	Q05872	Sequence encoding hum 4.37e+02
c	707	13	4.4	4035	2	Q06331	Human IL-2R beta chai 4.37e+02
c	708	13	4.4	4049	2	N70682	Sequence of cry-1-1 g 4.37e+02
	709	13	4.4	4086	9	Q51033	Human myoD gene and m 4.37e+02
	710	13	4.4	4100	16	Q74062	The rat beta-actin ge 4.37e+02
	711	13	4.4	4192	22	T16517	Collagen A1/decorin c 4.37e+02
	712	13	4.4	4197	15	Q99430	B. sphaericus SLP gen 4.37e+02
	713	13	4.4	4200	35	T78912	Spinocerebellar ataxi 4.37e+02
	714	13	4.4	4214	29	T59618	Connective tissue gro 4.37e+02
	715	13	4.4	4214	28	T58534	Human connective tiss 4.37e+02
c	716	13	4.4	4222	4	Q26932	Crystal protein gene 4.37e+02
c	717	13	4.4	4249	37	T91326	Arabidopsis thaliana 4.37e+02
	718	13	4.4	4291	18	T04702	ALVAC C3 flanking arm 4.37e+02
	719	13	4.4	4307	2	Q10988	Lambda clone 101 cont 4.37e+02
	720	13	4.4	4307	4	Q26078	Lambda clone 101. 4.37e+02
	721	13	4.4	4316	23	T13324	Kalinin/laminin 5 gam 4.37e+02
	722	13	4.4	4358	9	Q50500	aprt targetting vecto 4.37e+02
	723	13	4.4	4368	18	T03133	Human dihydropyrimidi 4.37e+02
	724	13	4.4	4369	18	T03143	Human dihydropyrimidi 4.37e+02
	725	13	4.4	4414	18	T03132	Bovine dihydropyrimid 4.37e+02
	726	13	4.4	4430	18	T03142	Bovine dihydropyrimid 4.37e+02
	727	13	4.4	4579	1	N81261	Probe for retinoblast 4.37e+02
c	728	13	4.4	4583	3	Q15174	Plasmid pAMVBT4. 4.37e+02
c	729	13	4.4	4583	29	T59737	Vector pAMVBTS encodi 4.37e+02
c	730	13	4.4	4583	1	N90183	DNA sequence of plant 4.37e+02
	731	13	4.4	4596	38	V03326	DNA sequence of NF-ka 4.37e+02
	732	13	4.4	4597	1	N81369	Human retinoblastoma 4.37e+02
	733	13	4.4	4597	12	Q70536	Human retinoblastoma 4.37e+02
c	734	13	4.4	4599	16	T01550	Rodent (hamster/rat) 4.37e+02
	735	13	4.4	4617	31	T72653	Rat tripeptidylpeptid 4.37e+02
c	736	13	4.4	4635	16	T01551	Rat sulphonylurea rec 4.37e+02
	737	13	4.4	4707	31	T70813	Deep Vent DNA polymer 4.37e+02
	738	13	4.4	4707	11	Q66294	Deep Vent DNA polymer 4.37e+02
c	739	13	4.4	4738	13	Q50997	Rat NMDA receptor sub 4.37e+02
c	740	13	4.4	4752	38	V04868	Nucleotide sequence o 4.37e+02
c	741	13	4.4	4779	25	T58751	Nucleolar/endosomal a 4.37e+02
c	742	13	4.4	4799	38	V04887	Nucleotide sequence o 4.37e+02
	743	13	4.4	4810	8	Q47194	Plasmid pHCV-168 codi 4.37e+02
	744	13	4.4	4810	15	Q97493	pHCV168 sequence. 4.37e+02
c	745	13	4.4	4815	38	V04886	Nucleotide sequence o 4.37e+02
	746	13	4.4	4857	30	T69962	PIPV nucleocapsid (N) 4.37e+02
c	747	13	4.4	4944	35	T91059	Human gene fohy030 co 4.37e+02
c	748	13	4.4	4944	24	T41470	Human 030 gene (fohy0 4.37e+02
c	749	13	4.4	4999	2	N60239	Sequence encoding ins 4.37e+02
c	750	13	4.4	5055	24	T41471	Human 030 gene (fohy0 4.37e+02
c	751	13	4.4	5055	35	T91060	Human gene fohy030 co 4.37e+02
	752	13	4.4	5056	1	N90489	cDNA of human retinob 4.37e+02
	753	13	4.4	5072	31	T71315	cDNA encoding C. eleg 4.37e+02
	754	13	4.4	5073	31	T71314	cDNA encoding C. eleg 4.37e+02
	755	13	4.4	5095	1	Q05590	Insert sequence encod 4.37e+02
	756	13	4.4	5100	15	Q90769	BAV3 E3 and fiber gen 4.37e+02

	757	13	4.4	5109 31	T72654	Rat brain homologue o	4.37e+02
c	758	13	4.4	5110 16	T01568	Rat sulphonylurea rec	4.37e+02
c	759	13	4.4	5113 26	T48479	Human tyrosine dephos	4.37e+02
	760	13	4.4	5153 18	T05116	STLVpan-p 5' region (	4.37e+02
	761	13	4.4	5183 39	V04063	Schizosaccharomyces p	4.37e+02
	762	13	4.4	5200 23	T13323	Kalinin/laminin 5 gam	4.37e+02
	763	13	4.4	5234 8	Q48464	Plasmid pG+host5 cont	4.37e+02
	764	13	4.4	5283 35	T91972	DNA encoding rat thro	4.37e+02
	765	13	4.4	5301 38	T93627	Human metastasis-asso	4.37e+02
	766	13	4.4	5363 9	Q50501	aprt/neo coding vecto	4.37e+02
	767	13	4.4	5394 17	T00872	Murine mC26 structura	4.37e+02
c	768	13	4.4	5460 20	T16508	Vector pAC3A1 contg.	4.37e+02
c	769	13	4.4	5679 38	V04867	Nucleotide sequence o	4.37e+02
	770	13	4.4	5775 2	N80920	Sequence encoding the	4.37e+02
	771	13	4.4	5981 8	Q46174	TSL1 gene.	4.37e+02
	772	13	4.4	6060 28	T48454	Human beta-globin gen	4.37e+02
	773	13	4.4	6142 27	T47878	Bacillus alkaline amy	4.37e+02
c	774	13	4.4	6151 13	Q75977	Insulin receptor subs	4.37e+02
	775	13	4.4	6464 8	Q48768	Arylamine N-acetyl-tr	4.37e+02
	776	13	4.4	6464 8	Q48770	Arylamine N-acetyl-tr	4.37e+02
	777	13	4.4	6464 8	Q48772	Arylamine N-acetyl-tr	4.37e+02
	778	13	4.4	6519 23	T37401	Human DOCK180 cDNA.	4.37e+02
c	779	13	4.4	6525 40	V16345	cDNA encoding full le	4.37e+02
c	780	13	4.4	6567 1	Q03324	Eimeria tenella genom	4.37e+02
	781	13	4.4	6722 8	Q48465	Plasmid pG+host6 cont	4.37e+02
	782	13	4.4	7091 33	T47530	H6 promoted cytomegal	4.37e+02
	783	13	4.4	7291 12	Q67860	ALVAC subgenomic frag	4.37e+02
c	784	13	4.4	7295 32	T72883	Plasmodium var-1 gene	4.37e+02
	785	13	4.4	7350 30	T69982	Canarypox virus DNA c	4.37e+02
	786	13	4.4	7351 3	Q22770	Canarypox virus BglII	4.37e+02
	787	13	4.4	7351 33	T47529	Canarypox virus DNA c	4.37e+02
	788	13	4.4	7474 31	T71320	Plasmid pTB73 encodin	4.37e+02
c	789	13	4.4	7655 15	Q86896	S. cerevisiae fks1 ge	4.37e+02
	790	13	4.4	7911 5	Q32436	HCV antigen clone T7N	4.37e+02
	791	13	4.4	8247 2	N70996	Sequence encoding N-t	4.37e+02
	792	13	4.4	8316 1	Q05955	Hepatitis C virus ope	4.37e+02
	793	13	4.4	8341 28	T09700	DNA encoding genes in	4.37e+02
	794	13	4.4	8363 16	Q92408	Human cyclin A gene.	4.37e+02
c	795	13	4.4	8433 14	Q78645	Marek's disease virus	4.37e+02
	796	13	4.4	8654 14	Q55138	Staphylococcus epider	4.37e+02
	797	13	4.4	8797 27	T50856	DNA cassette for amph	4.37e+02
	798	13	4.4	8841 14	Q89299	Tryptophan-2,3-dioxyg	4.37e+02
	799	13	4.4	8855 24	T29053	Phage T4 tail fibre g	4.37e+02
	800	13	4.4	9185 1	Q05956	Sense strand of the c	4.37e+02
	801	13	4.4	9185 2	Q10566	Hepatitis C virus str	4.37e+02
	802	13	4.4	9379 6	Q36209	Composite cDNA for He	4.37e+02
c	803	13	4.4	9391 6	Q38959	Hepatitis C virus gen	4.37e+02
c	804	13	4.4	9391 11	Q64175	NANBHV E1/E2 gene reg	4.37e+02
	805	13	4.4	9400 3	Q21744	Compiled HCV cDNA.	4.37e+02
	806	13	4.4	9401 18	T12710	Hepatitis C virus pol	4.37e+02
	807	13	4.4	9401 39	V09989	HCV polyprotein codin	4.37e+02

	808	13	4.4	9401 36	T99981	HCV polyprotein codin	4.37e+02
	809	13	4.4	9401 24	T41882	Hepatitis C virus pro	4.37e+02
	810	13	4.4	9402 7	Q41345	Human hepatitis C vir	4.37e+02
	811	13	4.4	9405 6	Q40426	Full-length Hepatitis	4.37e+02
	812	13	4.4	9413 14	Q80498	DNA encoding HCV prot	4.37e+02
	813	13	4.4	9413 15	Q81559	Hepatitis C virus hel	4.37e+02
	814	13	4.4	9413 22	T03960	Partial HCV non-struc	4.37e+02
	815	13	4.4	9416 3	Q22871	NANBV Hutch c59 isola	4.37e+02
	816	13	4.4	9436 11	Q63499	Blood transmiscible N	4.37e+02
	817	13	4.4	9472 6	Q33282	Korean hepatitis C vi	4.37e+02
c	818	13	4.4	9472 6	Q33282	Korean hepatitis C vi	4.37e+02
	819	13	4.4	9487 14	Q79140	Hepatitis C virus gen	4.37e+02
	820	13	4.4	9502 13	Q74770	Hepatitis C virus gen	4.37e+02
	821	13	4.4	9642 31	T71324	Plasmid pTB113 used f	4.37e+02
c	822	13	4.4	9714 39	T73135	Porcine transmissible	4.37e+02
c	823	13	4.4	10190 28	T33183	Fragment of human OCI	4.37e+02
	824	13	4.4	10443 31	T71319	Plasmid pTB72 encodin	4.37e+02
	825	13	4.4	11093 27	T50855	DNA cassette for amph	4.37e+02
	826	13	4.4	11288 15	Q90512	CEA clone HindIII-Sau	4.37e+02
c	827	13	4.4	12284 12	Q71308	Sequence encoding 55	4.37e+02
c	828	13	4.4	12284 1	Q06001	Sequence encoding pro	4.37e+02
c	829	13	4.4	12284 20	T29591	Hog cholera virus.	4.37e+02
	830	13	4.4	12720 10	Q64211	L.lactis branched ami	4.37e+02
	831	13	4.4	13414 31	T71131	Plasmid pCB50 encodin	4.37e+02
	832	13	4.4	13533 7	Q40862	VAC-beta (genomic - p	4.37e+02
	833	13	4.4	14255 7	Q43526	ALL-1 gene cDNA.	4.37e+02
	834	13	4.4	14255 40	V20477	Human ALL-1/AF-4 onco	4.37e+02
	835	13	4.4	14255 13	Q75181	ALL-1 (acute lymphocy	4.37e+02
	836	13	4.4	14556 15	Q90511	CEA genomic clone.	4.37e+02
	837	13	4.4	14928 2	Q11707	Immunoglobulin D-regi	4.37e+02
c	838	13	4.4	14985 40	V15693	Human Bin1 gene.	4.37e+02
c	839	13	4.4	14985 25	T43574	Human box-dependent m	4.37e+02
	840	13	4.4	15894 7	Q40480	Attenuated measles vi	4.37e+02
	841	13	4.4	18177 1	N90490	DNA of human retinobl	4.37e+02
	842	13	4.4	29392 40	V15422	Mouse poly Ig recepto	4.37e+02
	843	13	4.4	35100 40	V20441	Human c-fms oncogene.	4.37e+02
c	844	13	4.4	37895 24	T34137	Gene for biosynthetic	4.37e+02
c	845	13	4.4	37913 10	Q51128	Eicosapentaenoic acid	4.37e+02
	846	13	4.4	53526 40	T94101	Human PKD1 gene.	4.37e+02
	847	13	4.4	53577 40	T94108	Human PKD1 locus betw	4.37e+02
	848	13	4.4	53577 28	T18551	Human polycystic kidn	4.37e+02
	849	13	4.4	133894 17	T13635	AcNPV genomic DNA clo	4.37e+02
c	850	13	4.4	133894 17	T13635	AcNPV genomic DNA clo	4.37e+02
	851	13	4.4	134525 1	Q04525	Total base sequence o	4.37e+02
c	852	13	4.4	134525 1	Q04525	Total base sequence o	4.37e+02
	853	13	4.4	580073 27	T58840	Mycoplasma genitalium	4.37e+02
c	854	13	4.4	580073 27	T58840	Mycoplasma genitalium	4.37e+02
	855	12	4.0	25 17	Q93004	Pre-invasive human br	1.54e+03
c	856	12	4.0	38 39	V12472	Human oestrogen sulph	1.54e+03
c	857	12	4.0	38 33	T86158	Antisense primer ampl	1.54e+03
c	858	12	4.0	38 39	V15531	Primer for lymphoid-s	1.54e+03

c	859	12	4.0	38 30	T50929	Anchor primer used in	1.54e+03
c	860	12	4.0	38 22	T36731	Rat vasopressin V1b r	1.54e+03
c	861	12	4.0	39 39	V12723	Vkappa specific prime	1.54e+03
	862	12	4.0	167 21	T24514	Human gene signature	1.54e+03
c	863	12	4.0	208 38	V13472	Hepatitis G virus nuc	1.54e+03
c	864	12	4.0	225 14	Q81971	Human lysosomal membr	1.54e+03
	865	12	4.0	310 8	Q59647	Human brain Expressed	1.54e+03
	866	12	4.0	338 20	T22267	Human gene signature	1.54e+03
c	867	12	4.0	408 8	Q60016	Human brain Expressed	1.54e+03
c	868	12	4.0	416 4	Q27815	DNA encoding a portio	1.54e+03
c	869	12	4.0	423 33	T80654	Type II topoisomerase	1.54e+03
c	870	12	4.0	429 40	V09100	Nucleotide sequence o	1.54e+03
c	871	12	4.0	437 39	V05748	Nucleotide sequence o	1.54e+03
c	872	12	4.0	437 40	V17245	3' portion of clone O	1.54e+03
c	873	12	4.0	450 40	V10111	Human monoclonal anti	1.54e+03
c	874	12	4.0	458 16	Q90825	Wild type hMLH1 gene	1.54e+03
	875	12	4.0	480 34	T62968	pMON30379 encoding c-	1.54e+03
c	876	12	4.0	481 5	Q29873	Pheromone receptor cl	1.54e+03
	877	12	4.0	503 1	Q02418	Promoter of pCP-A1 ex	1.54e+03
	878	12	4.0	507 30	T67818	H. pylori cell envelo	1.54e+03
c	879	12	4.0	559 38	V01501	Fragment of human p53	1.54e+03
	880	12	4.0	576 21	T16564	Hepatitis C virus iso	1.54e+03
c	881	12	4.0	582 9	Q54717	Human NT-4 DNA isolat	1.54e+03
	882	12	4.0	597 22	T30453	Pertussis toxin B-oli	1.54e+03
c	883	12	4.0	638 40	V16668	Polynucleotide sequen	1.54e+03
	884	12	4.0	642 30	T68057	H. pylori cytoplasmic	1.54e+03
c	885	12	4.0	672 17	T11788	Mouse cysteine string	1.54e+03
c	886	12	4.0	686 23	T12957	AcaMAP45.	1.54e+03
c	887	12	4.0	708 29	T61242	Human anti-RSV monocl	1.54e+03
	888	12	4.0	741 23	T41907	Human thrombopoietin	1.54e+03
c	889	12	4.0	783 40	V03301	Ammonifex degensii KC	1.54e+03
	890	12	4.0	787 20	T14626	Rat R33 platelet deri	1.54e+03
c	891	12	4.0	812 39	V12610	Human DNA fragment vk	1.54e+03
	892	12	4.0	846 36	T85236	Human Fas ligand gene	1.54e+03
c	893	12	4.0	852 40	V24849	H. pylori ORF 04ap209	1.54e+03
c	894	12	4.0	894 40	T94103	Partial human PKD1 DN	1.54e+03
c	895	12	4.0	900 39	V12612	Human DNA fragment vk	1.54e+03
c	896	12	4.0	900 39	V12611	Human DNA fragment vk	1.54e+03
c	897	12	4.0	900 28	T37181	DNA fragment vk65.5,	1.54e+03
	898	12	4.0	954 7	Q42321	EAEC coagulation sequ	1.54e+03
c	899	12	4.0	958 39	T96064	Apoptosis associated	1.54e+03
c	900	12	4.0	999 8	Q48424	Human A2b adenosine r	1.54e+03
c	901	12	4.0	1100 40	V18890	Rat Hypertension rela	1.54e+03
	902	12	4.0	1101 1	N90131	DNA sequence of pBSF2	1.54e+03
	903	12	4.0	1139 1	N90255	Interleukin-6.	1.54e+03
c	904	12	4.0	1199 35	T39806	Mouse MLN50 gene.	1.54e+03
	905	12	4.0	1223 1	N82407	G1Tb clone contg. glu	1.54e+03
c	906	12	4.0	1241 39	V16184	Maize optimised DNA s	1.54e+03
c	907	12	4.0	1241 39	V16186	Maize optimised DNA s	1.54e+03
c	908	12	4.0	1241 5	Q30916	pSAB153 insert.	1.54e+03
	909	12	4.0	1267 17	Q99553	Thrombopoietin coding	1.54e+03

c	910	12	4.0	1302 38	T90222	Magnetospirullum sp.	1.54e+03
c	911	12	4.0	1313 27	T51049	Human adipocyte compl	1.54e+03
c	912	12	4.0	1320 40	V03316	Thermococcus celer ph	1.54e+03
c	913	12	4.0	1358 39	V16188	Maize optimised DNA s	1.54e+03
c	914	12	4.0	1365 38	V03980	P. furiosus ADPHK cod	1.54e+03
c	915	12	4.0	1389 39	V16175	Maize optimised DNA s	1.54e+03
c	916	12	4.0	1389 38	T74002	Maize optimised-Bacil	1.54e+03
c	917	12	4.0	1399 39	V16173	Maize optimised DNA s	1.54e+03
c	918	12	4.0	1399 20	T13953	Maize-optimized VIP2A	1.54e+03
	919	12	4.0	1405 23	T35761	Human non-conventiona	1.54e+03
c	920	12	4.0	1442 13	Q79524	Part of bovine bone m	1.54e+03
c	921	12	4.0	1447 38	V12758	Cytochrome P450 2C19	1.54e+03
c	922	12	4.0	1485 39	V10135	Stealth virus nucleic	1.54e+03
c	923	12	4.0	1512 40	T99593	Human TIE-2 receptor	1.54e+03
c	924	12	4.0	1521 38	V01503	Human p53 tumour supp	1.54e+03
c	925	12	4.0	1669 38	V12759	Wild type cytochrome	1.54e+03
	926	12	4.0	1694 36	T94635	TNF-R1-DD ligand prot	1.54e+03
c	927	12	4.0	1746 21	T03708	Cytochrome P450 2C19	1.54e+03
c	928	12	4.0	1858 10	Q56202	Partial coding sequen	1.54e+03
c	929	12	4.0	1891 40	V17551	LambdaMe2 containing	1.54e+03
	930	12	4.0	2004 22	T11599	Human alpha-1C adrene	1.54e+03
c	931	12	4.0	2034 38	V01497	Monkey p53 tumour sup	1.54e+03
c	932	12	4.0	2079 2	N80923	Sequence of thyroid h	1.54e+03
c	933	12	4.0	2096 39	V15795	cDNA for wild type hu	1.54e+03
c	934	12	4.0	2105 1	Q04685	SalI/SalI fragment of	1.54e+03
	935	12	4.0	2116 25	T18213	Major neutralising an	1.54e+03
	936	12	4.0	2146 25	T44320	Human TIE-2 ligand 1	1.54e+03
	937	12	4.0	2147 34	T59362	DNA encoding wild typ	1.54e+03
	938	12	4.0	2147 34	T70087	DNA encoding aspartok	1.54e+03
	939	12	4.0	2147 34	T70086	DNA encoding aspartok	1.54e+03
c	940	12	4.0	2156 38	V01498	Human p53 tumour supp	1.54e+03
c	941	12	4.0	2233 40	V10120	Human retinoid recept	1.54e+03
c	942	12	4.0	2277 14	Q89338	Mouse Stat1 cDNA.	1.54e+03
	943	12	4.0	2362 40	V15227	Heterodimeric human i	1.54e+03
c	944	12	4.0	2383 39	V09638	Human M6 cDNA.	1.54e+03
	945	12	4.0	2385 23	T33590	Signal transducer and	1.54e+03
	946	12	4.0	2412 36	T97869	cDNA encoding a chime	1.54e+03
	947	12	4.0	2415 24	T30268	Cotton fibre clone SK	1.54e+03
c	948	12	4.0	2448 38	V12201	Human neuronal nicoti	1.54e+03
	949	12	4.0	2451 39	V15809	Micrococcus luteus He	1.54e+03
c	950	12	4.0	2500 1	Q05784	Sequence encoding hum	1.54e+03
c	951	12	4.0	2530 33	T59545	Coding sequence for O	1.54e+03
c	952	12	4.0	2733 25	T31051	Human fibroblast grow	1.54e+03
	953	12	4.0	2735 2	Q11072	B.stearothermophilus	1.54e+03
	954	12	4.0	2750 25	T39335	Human adrenoleucodyst	1.54e+03
	955	12	4.0	2809 15	Q91287	ORC2 gene encoding a	1.54e+03
	956	12	4.0	2898 35	T68848	Photorhabdus luminesc	1.54e+03
c	957	12	4.0	2911 39	V05709	Human Th2/B19 cDNA.	1.54e+03
	958	12	4.0	2935 29	T59199	Ca-independent phosph	1.54e+03
c	959	12	4.0	2973 9	Q53501	Sequence of bean mali	1.54e+03
c	960	12	4.0	3065 39	V00014	Mouse sex comb on mid	1.54e+03

c	961	12	4.0	3065	39	V00010	Mouse sex comb on mid 1.54e+03
c	962	12	4.0	3225	40	V20467	Human c-myb oncogene. 1.54e+03
c	963	12	4.0	3300	40	V17259	Rat protein p138 codi 1.54e+03
c	964	12	4.0	3319	25	T42166	Leishmania chagasi K3 1.54e+03
	965	12	4.0	3328	24	T43277	DOA4 coding sequence. 1.54e+03
	966	12	4.0	3410	1	N80617	Genomic clone encodin 1.54e+03
	967	12	4.0	3471	33	T71761	PSKH-1 cDNA. 1.54e+03
c	968	12	4.0	3579	39	V15773	B. lactofermentum DNA 1.54e+03
c	969	12	4.0	3880	15	T03888	Human mGluR4 gene. 1.54e+03
c	970	12	4.0	3896	40	V03312	Isolate OC1/4V phosph 1.54e+03
c	971	12	4.0	4031	39	V16189	Maize optimised DNA s 1.54e+03
c	972	12	4.0	4164	40	V12216	Mouse retinoid metabo 1.54e+03
c	973	12	4.0	5125	5	Q29703	IRS-1. 1.54e+03
	974	12	4.0	5359	17	T12251	Cytochrome P450 isoen 1.54e+03
c	975	12	4.0	5362	39	T97296	Mouse additional sex 1.54e+03
c	976	12	4.0	5362	39	T97294	Mouse additional sex 1.54e+03
c	977	12	4.0	5775	40	V20437	Human c-K-ras oncogen 1.54e+03
c	978	12	4.0	6412	39	V06017	Human calcium protein 1.54e+03
c	979	12	4.0	6568	39	V15949	Nevoid basal cell car 1.54e+03
	980	12	4.0	6763	34	T79728	Beta-lactoglobulin cl 1.54e+03
c	981	12	4.0	6961	34	T84508	Rabbit LDL receptor a 1.54e+03
	982	12	4.0	7470	24	T39905	Maize acetyl CoA carb 1.54e+03
c	983	12	4.0	7493	40	V12115	Mus musculus factor V 1.54e+03
	984	12	4.0	7745	34	T79859	Genomic DNA encoding 1.54e+03
	985	12	4.0	8789	40	V09076	Mitosin nucleic acid 1.54e+03
	986	12	4.0	10163	33	T61085	Full-length HIV-2KR p 1.54e+03
c	987	12	4.0	14042	39	V05995	Human placenta calciu 1.54e+03
c	988	12	4.0	14044	39	V05997	Human parathyroid cal 1.54e+03
c	989	12	4.0	14060	40	T94102	Human PKD1 cDNA. 1.54e+03
c	990	12	4.0	14080	39	V05996	Human kidney calcium 1.54e+03
c	991	12	4.0	14086	39	V05994	Human calcium protein 1.54e+03
c	992	12	4.0	20303	31	T71699	Human deoxycytidylate 1.54e+03
	993	12	4.0	26700	40	V15946	Mouse syndecan gene s 1.54e+03
c	994	12	4.0	26700	40	V15946	Mouse syndecan gene s 1.54e+03
	995	12	4.0	26764	31	T71696	Human deoxycytidylate 1.54e+03
c	996	12	4.0	28994	39	V15826	Genomic DNA for inter 1.54e+03
c	997	12	4.0	29392	40	V15422	Mouse poly Ig recepto 1.54e+03
c	998	12	4.0	35100	40	V20441	Human c-fms oncogene. 1.54e+03
c	999	12	4.0	53526	40	T94101	Human PKD1 gene. 1.54e+03
c	1000	12	4.0	53577	40	T94108	Human PKD1 locus betw 1.54e+03

ALIGNMENTS

```
RESULT    1
ID   T23937 standard; cDNA to mRNA; 314 BP.
AC   T23937;
DT   23-AUG-1996  (first entry)
DE   Human gene signature HUMGS05879.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
```



```
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 1491; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   314 BP;    75 A;    71 C;    75 G;    77 T;

  Query Match          100.0%;  Score 298;  DB 20;  Length 314;
  Best Local Similarity 100.0%;  Pred. No. 1.65e-278;
  Matches  313;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 gatcctctctcccagggaatccgacacaggaggaaccccttctntggttganctggggcca 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      1 GATCCTCTCTCCCAGGGAATCCGACACAGGAGGAACCCCTTCTNTGGTTGANCTGGGCCA 60

Db     61 ggcctaagagtagcaggaactctaagaccacagagttttttatanaatgtataaaatgtatc 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     61 GGCCTAAGAGTAGCAGGAACTCTAAGACCACAGAGTTTTTTATAAATGTATAAATGTATC 120

Db    121 aagccaaatntgcagatgctaactggacatttctggggaactgggcaccaggngtgccttc 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    121 AAGCCAAATNTGCAGATGCTAACTGGACATTCTGGGGAACTGGGCACCAGGNGTGCCTTC 180

Db    181 atacactgtaccccagctctnttctaaaagagaagtnggtgggcacacttgaactgtttg 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    181 ATACACTGTACCCCAGCTCTNTTCTAAAAGAGAAGTNGGTGGGCACACTTGAACTGTTTG 240
```



```
Db    241 gtggnccnaaccacaggangctgcaattntntggnttagggtgatacttttggccctncn 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    241 GTGGNCCNAACCACAGGANGCTGCAATTTNTNTGGNTTAGGGTGATACTTTTGGCCCTNCN 300

Db    301 tgtggccccctntt 313
          ||||||||||||||
Qy    301 TGTGGCCCCTNTT 313


RESULT    2
ID   Q59241 standard; cDNA; 390 BP.
AC   Q59241;
DT   16-MAR-1994  (first entry)
DE   Human brain Expressed Sequence Tag EST00165.
KW   Gene transcription product; genetic markers; tagging; in vivo;
KW   transcription; mapping; locations; chromosomes; chromosomal; ss.
OS   Homo sapiens.
PN   WO9316178-A.
PD   19-AUG-1993.
PF   12-FEB-1993; U01294.
PR   12-FEB-1992; US-837195.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
PI   Adams MD, Moreno RF, Venter CJ;
DR   WPI; 93-272882/34.
PT   Enriched oligonucleotides and corresp. sequences - used as
PT   markers for human genes transcribed in-vivo, facilitate tagging
PT   of most human genes
PS   Example 1; Page 143; 500pp; English.
CC   The Expressed Sequence Tag was isolated from a human brain cDNA
CC   library as part of a large set of ESTs which can be used as markers
CC   for human genes transcribed in vivo. They can be used to facilitate
CC   tagging of most human genes, for mapping locations of expressed genes
CC   on chromosomes, for individual or forensic identification, for mapping
CC   locations of disease-associated genes, for identification of tissue
CC   type, and for prepn. of antisense sequences, probes and constructs.
CC   EST00165 has an "excellent" coding probability as evaluated using the
CC   coding-region prediction program CRM and hybridises to chromosome 15.
CC   See also Q59041-Q61440.
SQ   Sequence   390 BP;    72 A;    113 C;    140 G;    62 T;

  Query Match            6.0%;  Score 18;  DB 6;  Length 390;
  Best Local Similarity 100.0%;  Pred. No. 2.85e-01;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    151 ctggacattctggggaac 168
          ||||||||||||||||||
Qy    143 CTGGACATTCTGGGGAAC 160


RESULT    3
```



```
ID   Q39829 standard; DNA; 390 BP.
AC   Q39829;
DT   20-MAR-1995  (first entry)
DE   Expressed Sequence Tag human gene marker EST00165.
KW   expressed sequence tag; human genome project; chromosome;
KW   human gene sequencing; PCR mapping; somatic cell hybrids;
KW   sublocalisation; gene tagging; tissue typing.
OS   Synthetic.
PN   WO9300353-A.
PD   07-JAN-1993.
PF   19-JUN-1992; U05222.
PR   20-JUN-1991; US-716831.
PR   12-FEB-1992; US-837195.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
PI   Adams MD,  Venter JC;
DR   WPI; 93-036325/04.
PT   Particular expressed sequence tags from human CDNA - corresponds
PT   to transcription prods. of genes, useful for tagging genes, .
PT   mapping chromosomes and tissue typing
PS   Claim 3; Page 153; 199pp; English.
CC   This sequence represents an EST (expressed sequence tag) ESTs are markers
CC   for human genes actually transcribed in vivo. Unlike the random genomic
CC   DNA sequence tagged sites (STSs), ESTs point directly to expressed genes.
CC   The use of ESTs could facilitate the tagging of most expressed human
CC   genes within a few years at a fraction of the cost of complete genomic
CC   sequencing. Using PCR primers Q39419-Q39580 (sequences designed
CC   from the ESTs) sublocalisation of an EST can be achieved with panels of
CC   fragments from specific chromosomes or pools of large genomic clones in
CC   an analogous manner. This sequence represents EST00165.
SQ   Sequence   390 BP;    72 A;    113 C;    140 G;    62 T;

  Query Match            6.0%;  Score 18;  DB 6;  Length 390;
  Best Local Similarity 100.0%;  Pred. No. 2.85e-01;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    151 ctggacattctggggaac 168
          ||||||||||||||||||
Qy    143 CTGGACATTCTGGGGAAC 160


RESULT    4
ID   Q80319 standard; DNA; 1641 BP.
AC   Q80319;
DT   22-AUG-1995  (first entry)
DE   Human 38kDa FK-506 binding protein DNA.
KW   Human 38kDA FK-506 binding protein; allograft rejection;
KW   T-cell activation prevention; immunosuppressants; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             141..1208
FT                   /*tag= a
```

```
PN   GB2279349-A.
PD   04-JAN-1995.
PF   15-JUN-1994; 012000.
PR   25-JUN-1993; US-083944.
PA   (MERI ) MERCK & CO INC.
PI   Lam E, Wiederrecht GJ;
DR   WPI; 95-024909/04.
DR   P-PSDB; R64226.
PT   New human 38 kD FK-506 binding protein and DNA encoding it -
PT   useful for identifying cpds. able to block allo:graft rejection
PT   by preventing T-cell activation
PS   Claim 2; Pages 33-34; 44pp; English.
CC   Q80319 encodes R64226 the human 38kDa FK-506 binding protein. The
CC   protein can be used to identify cpds. with immunosuppressive
CC   activity, able to block allograft rejection by preventing T-cell
CC   activation. It can also be used, opt. when immobilised, to
CC   identify, quantify and purify FK-506 and FK-506 like cpds.
SQ   Sequence   1641 BP;     324 A;     550 C;     507 G;     260 T;

  Query Match           6.0%;  Score 18;  DB 13;  Length 1641;
  Best Local Similarity 100.0%;  Pred. No. 2.85e-01;
  Matches   18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    249 ctggacattctggggaac 266
          ||||||||||||||||||
Qy    143 CTGGACATTCTGGGGAAC 160


RESULT    5
ID   T68945 standard; cDNA; 1146 BP.
AC   T68945;
DT   18-AUG-1997  (first entry)
DE   Mouse p32 11-cis-retinol dehydrogenase cDNA clone.
KW   11-Cis retinol dehydrogenase; p32; 11-cis retinaldehyde;
KW   retinitis pigmentosa; psoriasis; acne; cancer; diagnosis; therapy;
KW   ss.
OS   Mus musculus.
FH   Key             Location/Qualifiers
FT   cds             53..1009
FT                   /*tag= a
PN   WO9719167-A1.
PD   29-MAY-1997.
PF   14-NOV-1996; U18295.
PR   22-NOV-1995; US-562114.
PR   11-OCT-1996; US-729594.
PA   (LUDW-) LUDWIG INST CANCER RES.
PI   Eriksson U,  Romert A,  Simon A;
DR   WPI; 97-298106/27.
DR   P-PSDB; W18337.
PT   New 11-cis retinol dehydrogenase gene  - used to develop products
PT   for the diagnosis and therapy of conditions such as retinitis
```



```
PT   pigmentosa, psoriasis and cancers
PS   Claim 16; Page 52; 68pp; English.
CC   A cDNA clone (T68945) was isolated from murine liver cDNA library
CC   using murine p32 PCR clone 200 (T68940) as probe; other liver
CC   clones (T68942-44) were isolated using p32 PCR clone 207 (T68939)
CC   as probe.  These 4 isolated clones code for p32 11-cis retinol
CC   dehydrogenase polypeptides (W18334-37, respectively).  p32 cDNAs
CC   (see also T68935-41) can be used to produce p32 polypeptides and to
CC   develop products for the diagnosis and therapy of conditions such
CC   as retinitis pigmentosa, psoriasis, acne and some cancers.
SQ   Sequence   1146 BP;     237 A;     342 C;     305 G;     262 T;

  Query Match           5.4%;  Score 16;  DB 31;  Length 1146;
  Best Local Similarity 100.0%;  Pred. No. 6.37e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    307 tccccagaatgtccag 322
          ||||||||||||||||
Cp    158 TCCCCAGAATGTCCAG 143


RESULT    6
ID   Q14638 standard; cDNA; 1649 BP.
AC   Q14638;
DT   30-JAN-1992  (first entry)
DE   Human temporal lobe cDNA insert from plasmid pPDE2RR.
KW   brain; pRATPDP; cAMP; phosphodiesterase; complemetation analysis; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             210..1018
FT                   /*tag= a
PN   WO9116457-A.
PD   31-OCT-1991.
PF   19-APR-1991; U02714.
PR   20-APR-1990; US-511715.
PA   (COLD-) COLD SPRING HARBOR.
PI   Wigler MH, Colicelli JJ;
DR   WPI; 91-339841/46.
DR   P-PSDB; R14849.
PT   Complementary screening for genes and prods. - e.g. RAS protein
PT   and cAMP, that modify, complement or suppress genetic defect and
PT   correct associated phenotypic alteration
PS   Claim 16; Page 119; 169pp; English.
CC   Plasmid pRATDPD was isolated from a rat brain cDNA library. It is
CC   thought to encode a cyclic nucleotide PDE. The RATDPD cDNA was used
CC   as a probe to isolate plasmid pPDE2RR by low stringency screens of
CC   a human temporal lobe cDNA library. The inventors have classified
CC   PDE2RR in PDE class IV4 along with TM72 (Q14631).
SQ   Sequence   1649 BP;     474 A;     378 C;     378 G;     419 T;

  Query Match           5.4%;  Score 16;  DB 3;  Length 1649;
```



```
  Best Local Similarity 100.0%;  Pred. No. 6.37e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    957 gattccctgggagaga 972
          ||||||||||||||||
Cp     21 GATTCCCTGGGAGAGA 6


RESULT    7
ID   T34379 standard; cDNA; 1649 BP.
AC   T34379;
DT   10-OCT-1996  (first entry)
DE   Plasmid pPDE2RR insert.
KW   Human; glioblastoma cell; plasmid; mammalian; complementation; pPDE2RR;
KW   probe; yeast; pPDET; pPDE10X inv; temporal lobe; cDNA library; pRATPDP;
KW   pTM72; pGB14; pGB18ARR; pTM3; pJC44x; pGB25; phosphodiesterase family IV;
KW   pPDE18; pPDE21; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             210..1019
FT                   /*tag= a
FT                   /product= cAMP phosphodiesterase
PN   US5527896-A.
PD   18-JUN-1996.
PF   20-APR-1990; 511715.
PR   20-APR-1990; US-511715.
PR   19-APR-1991; US-688352.
PA   (COLD-) COLD SPRING HARBOR LAB.
PI   Colicelli JJ,  Wigler MH;
DR   WPI; 96-299902/30.
DR   P-PSDB; W00095.
PT   DNA mols. isolated from human glioblastoma cells - encode
PT   RAS-related or cyclic nucleotide phosphodiesterase proteins
PS   Claim 5; Column 117-120; 101pp; English.
CC   The sequences given in T34377-84 represent plasmid fragments which
CC   were isolated by hybridisation with mammalian genes cloned by
CC   complementation.  These sequences were isolated using probes derived
CC   from the sequences given in T34366-76 which were cloned via
CC   complementation in yeast.  Plasmids pPDET, pPDE10X inv and pPDE2RR were
CC   isolated by low stringency hybridisation screens of a human temporal
CC   lobe cDNA library using the pRATPDP insert as a probe.  Comparison of
CC   the nucleotide sequences given in T34377-79 indicated that the inserts
CC   are representatives of the same genetic locus as the insert in pTM72.
CC   Plasmids pGB14 and pGB18ARR were obtained in the same manner.  DNA
CC   sequence analysis revealed that they are representatives of the same
CC   genetic locus as the inserts in pTM3 and pJC44x.  Plasmid pGB25 was also
CC   obtained at low stringency hybridisation using the pRATPDP insert as a
CC   probe.  Judged by its nucleotide and deduced amino acid sequence it
CC   represents a novel member of the phosphodiesterase family IV.  The cDNA
CC   insert of pBG25 was used as a probe to obtain pPDE18 and pPDE21.  The
CC   cDNA of pPDE18 represents the same locus as that of pGB25 and contains
```



```
CC   more sequence information than the pGB25 cDNA.  The pPDE21 insert
CC   represents a fourth member of phosphodiesterase family IV.  The
CC   assignment to family IV is based solely on sequence relationships.
SQ   Sequence   1649 BP;     474 A;     379 C;     375 G;     421 T;

  Query Match           5.4%;  Score 16;  DB 21;  Length 1649;
  Best Local Similarity 100.0%;  Pred. No. 6.37e+00;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    957 gattccctgggagaga 972
          ||||||||||||||||
Cp     21 GATTCCCTGGGAGAGA 6


RESULT    8
ID   Q03601 standard; DNA; 1891 BP.
AC   Q03601;
DT   02-AUG-1990  (first entry)
DE   Gene encoding yeast aminopeptidase I.
KW   Aminopeptidase; AP1; AP2; APX; vacuolar localisation; ds.
OS   Saccharomyces cerevisiai.
FH   Key             Location/Qualifiers
FT   cds             166..487
FT                   /*tag= a
FT   transit_peptide 166..300
FT                   /*tag= b
FT                   /label=Vacuolar localisation presequence.
FT   mat_peptide     301..487
FT                   /*tag= c
PN   EP-359164-A.
PD   21-MAR-1990.
PF   9-SEP-1989; 116734.
PR   13-SEP-1988; US-243733.
PR   14-DEC-1988; US-284244.
PA   (GEHO-) Gen Hospital Corp.
PI   Smith JA, Chang Y;
DR   WPI; 90-084991/12.
DR   P-PSDB; R05535.
PT   Substrate-specific amino:peptidase enzymes AP1, AP2 and APX -
PT   used to remove amino-terminal amino acid residues from
PT   peptide(s) and protein(s).
PS   Disclosure; 24pp; English.
CC   Aminopeptidases are useful in extending the stability, solubility and
CC   biological activity and half-life, or in decreasing the immunogenicity.
CC   The vacuolar location presequence encoded by the gene is useful ni
CC   directing linked polypeptides through the vacuolar membrane.
SQ   Sequence   1891 BP;     606 A;     387 C;     388 G;     510 T;

  Query Match           5.4%;  Score 16;  DB 1;  Length 1891;
  Best Local Similarity 94.1%;  Pred. No. 6.37e+00;
  Matches   16;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;
```

```
Db    207 tctgcagatgctaactg 223
          | ||||||||||||||||
Qy    129 TNTGCAGATGCTAACTG 145
```

```
RESULT    9
ID   T72943 standard; cDNA; 2058 BP.
AC   T72943;
DT   10-MAR-1998  (first entry)
DE   Phaffia carotenoid biosynthesis pathway enzyme gene PRcrtI.
KW   Glyceraldehyde-3-phosphate dehydrogenase; GAPDH; carotenoid;
KW   synthesis; promoter; recombinant DNA; astaxanthin; ribosomal protein;
KW   food colouring; ds.
OS   Phaffia rhodozyma.
FH   Key             Location/Qualifiers
FT   CDS             46..1794
FT                   /*tag=  a
FT                   /product= PRcrtI
PN   WO9723633-A1.
PD   03-JUL-1997.
PF   23-DEC-1996; E05887.
PR   11-APR-1996; EP-200943.
PR   22-DEC-1995; EP-203620.
PA   (KONN ) GIST-BROCADES BV.
PA   (OOIJ/) OOIJEN A J J.
PI   Verdoes JC,  Wery J;
DR   WPI; 97-351068/32.
DR   P-PSDB; W22499.
PT   Phaffia derived GAPDH and carotenoid synthesis genes and promoter
PT   fragment - used in the recombinant production of therapeutically
PT   useful proteins e.g. carotenoids for use in food colouring
PS   Claim 32; Page 56-58; 118pp; English.
CC   The present sequence represents a Phaffia derived carotenoid
CC   biosynthesis pathway enzyme gene which can be used in the novel
CC   recombinant DNA of the present invention. The recombinant DNA comprises
CC   a transcription promoter operably linked to a downstream sequence to be
CC   expressed, where the transcription promoter comprises a region found
CC   upstream of the open reading frame (ORF) of a highly expressed Phaffia
CC   gene (preferably GAPDH, a ribosomal protein gene, an enzyme involved in
CC   the biosynthesis pathway). The recombinant DNA can be used to transform
CC   hosts, preferably Phaffia. These transformed hosts are then used in the
CC   recombinant production of GAPDH or an enzyme involved in carotenoid
CC   synthesis, preferably astaxanthin. They may also be used to produce a
CC   pharmaceutical product. Purified carotenoids can be used as colourants
CC   in food and/or feed, and also in cosmetics.
SQ   Sequence   2058 BP;    498 A;    503 C;    485 G;    572 T;

  Query Match           5.4%;  Score 16;  DB 36;  Length 2058;
  Best Local Similarity  94.1%;  Pred. No. 6.37e+00;
  Matches    16;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;
```



```
Db    985 aagagaagttggtgggc 1001
          ||||||||| |||||||
Qy    208 AAGAGAAGTNGGTGGGC 224
```

```
RESULT   10
ID   Q71287 standard; cDNA to mRNA; 2226 BP.
AC   Q71287;
DT   12-MAY-1995  (first entry)
DE   Feline zona pellucida (FZP)-2 cDNA.
KW   Feline zona pellucida; FZP-2; contraceptive vaccine antigen; ds.
OS   Felis domesticus.
FH   Key             Location/Qualifiers
FT   cds             50..2200
FT                   /*tag=  a
FT                   /product= FZP-2
PN   J06217777-A.
PD   09-AUG-1994.
PF   29-JAN-1993; 013496.
PR   29-JAN-1993; JP-013496.
PA   (TOFU ) TONEN CORP.
DR   WPI; 94-305384/38.
DR   P-PSDB; R60532.
PT   New DNA sequence coding feline zona pellucida (FZP) 2 - used for
PT   prodn. of contraceptive vaccine antigen for cats
PS   Claim 1; Page 7-10; 10pp; Japanese.
CC   The feline zona pellucida (FZP)-2 coding sequence was obtained by
CC   PCR amplification of cDNA synthesised from mRNA isolated from cat
CC   ovaries. Subfragments of the FZP-2 protein can be expressed for use
CC   as antigens in contraceptive vaccines for cats.
SQ   Sequence   2226 BP;    614 A;    536 C;    490 G;    586 T;

  Query Match           5.4%;  Score 16;  DB 12;  Length 2226;
  Best Local Similarity 100.0%;  Pred. No. 6.37e+00;
  Matches    16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     34 ggtgatacttttggcc 49
          ||||||||||||||||
Qy    280 GGTGATACTTTTGGCC 295
```

```
RESULT   11
ID   T34375 standard; cDNA; 3186 BP.
AC   T34375;
DT   09-OCT-1996  (first entry)
DE   Plasmid pTM72 (ATCC 68602) insert.
KW   Human; glioblastoma cell; heat shock sensitivity; phosphodiesterase;
KW   deficient yeast strain 10DAB; pTM22; rat DPD phosphodiesterase; pde1-;
KW   bovine Ca2+/calmodulin dependent cAMP phosphodiesterase; heart; plasmid;
KW   RAS2(val19); pde2-; pTM3; pTM72; pRATDPD; pJC99; rolipram sensitive; ss.
```



```
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             139..2349
FT                   /*tag=  a
FT                   /product= cAMP phosphodiesterase
PN   US5527896-A.
PD   18-JUN-1996.
PF   20-APR-1990; 511715.
PR   20-APR-1990; US-511715.
PR   19-APR-1991; US-688352.
PA   (COLD-) COLD SPRING HARBOR LAB.
PI   Colicelli JJ,  Wigler MH;
DR   WPI; 96-299902/30.
DR   P-PSDB; W00093.
PT   DNA mols. isolated from human glioblastoma cells - encode
PT   RAS-related or cyclic nucleotide phosphodiesterase proteins
PS   Claim 4; Column 81-88; 101pp; English.
CC   The sequences given in T34374-76 represent plasmid fragments which
CC   contain human glioblastoma cell cDNA inserts which are capable of
CC   correcting the heat shock sensitivity of the phosphodiesterase
CC   deficient yeast strain 10DAB.  Several cDNA's were isolated and
CC   sequenced.  pTM22 encodes a novel human gene.  From computer analysis,
CC   pTM22 putatively encodes a protein homologous to various cAMP
CC   phosphodiesterases, such as the bovine Ca2+/calmodulin dependent cAMP
CC   phosphodiesterase and the rat DPD phosphodiesterase.  Sequences related
CC   to pTM22 were found to be expressed in human heart.  Plasmid pTM22 was
CC   unable to correct the heat shock sensitivity of RAS2(val19) yeast
CC   strains.  It thus appears that the pde1- and pde2- yeast strain 10DAB
CC   is more sensitive to phenotypic reversion by mammalian cAMP
CC   phosphodiesterase clones than is the RAS2(val19) yeast strain.  The
CC   inserts in the plasmids pTM3 and pTM72 were also characterised.  These
CC   two different cAMP phosphodiesterase cDNA's were found to be closely
CC   related to, but distinct from, the pRATDPD insert and the pJC99 insert.
CC   Biochemical analysis of cell lysates has established that the cDNA's of
CC   pTM3 and pTM72, pJC44x and pRATDPD encode rolipram sensitive cAMP
CC   phosphodiesterases.
SQ   Sequence   3186 BP;    927 A;    744 C;    734 G;    781 T;

  Query Match           5.4%;  Score 16;  DB 22;  Length 3186;
  Best Local Similarity 100.0%;  Pred. No. 6.37e+00;
  Matches    16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2287 gattccctgggagaga 2302
          ||||||||||||||||
Cp     21 GATTCCCTGGGAGAGA 6
```

```
RESULT   12
ID   Q14631 standard; DNA; 3186 BP.
AC   Q14631;
DT   30-JAN-1992  (first entry)
```



```
DE   Plasmid pTM72 insert encoding a human cAMP phosphodiesterase.
KW   Cyclic nucleotide; PDE; glioblastoma cell; rolipram-sensitive; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             139..2348
FT                   /*tag=  a
PN   WO9116457-A.
PD   31-OCT-1991.
PF   19-APR-1991; U02714.
PR   20-APR-1990; US-511715.
PA   (COLD-) COLD SPRING HARBOR.
PI   Wigler MH, Colicelli JJ;
DR   WPI; 91-339841/46.
DR   P-PSDB; R14843.
PT   Complementary screening for genes and prods. - e.g. RAS protein
PT   and cAMP, that modify, complement or suppress genetic defect and
PT   correct associated phenotypic alteration
PS   Claim 15; Page 91; 169pp; English.
CC   Plasmid pTM72 contains a human glioblastoma cDNA which encodes a
CC   rolipram-sensitive cAMP PDE. The sequence is very closely related
CC   to, but distinct from, the rat pRATDPD cDNA insert (see Q14624).
CC   N.B. This sequence is SEQ ID NO. 22 in the specification but is
CC   referred to as SEQ ID NO.23 in the text ,e.g. on page 32.
SQ   Sequence   3186 BP;    927 A;    743 C;    734 G;    782 T;

  Query Match           5.4%;  Score 16;  DB 3;  Length 3186;
  Best Local Similarity 100.0%;  Pred. No. 6.37e+00;
  Matches    16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2287 gattccctgggagaga 2302
          ||||||||||||||||
Cp     21 GATTCCCTGGGAGAGA 6
```

```
RESULT   13
ID   T99544 standard; cDNA; 3804 BP.
AC   T99544;
DT   21-MAY-1998  (first entry)
DE   Human haematopoietic stem and progenitor cell antigen AC133 cDNA.
KW   AC133; haematopoietic stem cell; progenitor cell; antigen;
KW   marker; human; bone marrow; gene therapy; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             38..2635
FT                   /*tag=  a
FT   sig_peptide     38..94
FT                   /*tag=  b
FT   mat_peptide     95..2632
FT                   /*tag=  c
PN   WO9741224-A1.
PD   06-NOV-1997.
```

PF 25-APR-1997; U06930.
PR 23-APR-1997; US-842382.
PR 26-APR-1996; US-639891.
PA (AMCE-) AMCELL CORP.
PI Buck D, Godfry WG, Miraglia S, Yin A;
DR WPI; 97-549728/50.
DR P-PSDB; W26769.
PT Antibody specific for haematopoietic stem/progenitor cell antigen,
PT AC133 - useful to purify such cells, e.g. for use as gene therapy
PT vehicles or for bone marrow reconstitution
PS Claim 21; Page 42-48; 78pp; English.
CC This claimed nucleic acid molecule includes a coding region for
CC a human haematopoietic progenitor cell antigen (see W26769),
CC designated AC133 antigen. It was isolated form a WERI-Rb-1
CC retinoblastoma cDNA library by PCR amplification using primers
CC based on isolated peptides of purified AC133. Expression of AC133
CC is highly tissue specific. It is detected on a subset of
CC haematopoietic progenitor cells derived from human bone marrow,
CC foetal bone marrow and liver, cord blood and adult peripheral
CC blood. The subset of cells recognised by AC133 is CD34bright and
CC contains substantially all of the granulocyte/monocyte colony
CC forming units present in the CD34+ population. The highly specific
CC distribution of AC133 makes it useful as a reagent for isolating
CC and characterising human haematopoietic progenitor and stem cells,
CC which can be used e.g. as gene therapy vehicles or for bone marrow
CC reconstitution. Exploiting AC133 as a marker ensures a high purity
CC population, reducing the risk of tumour cell contamination of donor
CC cells. AC133 polynucleotides can be used to identify related
CC sequences, or cells/tissues expressing AC133, and to produce
CC recombinant AC133.
SQ Sequence 3804 BP; 1128 A; 765 C; 813 G; 1098 T;

Query Match 5.4%; Score 16; DB 39; Length 3804;
Best Local Similarity 100.0%; Pred. No. 6.37e+00;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3093 accacagagtttttta 3108
||||||||||||||||
Qy 87 ACCACAGAGTTTTTTA 102

RESULT 14
ID Q71539 standard; cDNA; 3890 BP.
AC Q71539;
DT 28-APR-1995 (first entry)
DE Brain low Km, cAMP-specific phosphodiesterase cDNA.
KW Phosphodiesterase; PDE IV(B); brain; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 282..1973
FT /*tag= a

PN WO9420079-A.
PD 15-SEP-1994.
PF 10-MAR-1994; U02612.
PR 10-MAR-1993; US-029334.
PA (SMIK) SMITHKLINE BEECHAM CORP.
PI Livi GP, McLaughlin MM, Torphy TJ;
DR WPI; 94-332667/41.
DR P-PSDB; R60605.
PT Human low KM, cAMP-specific phospho-di:esterase and nucleic acids
PT encoding it - used to identify binding cpds. useful as
PT therapeutic agents and as probes to evaluate disease states.
PS Claim 3; Page 29-32; 54pp; English.
CC The cDNA encodes human low Km, cAMP-specific phosphodiesterase,
CC PDE IV(B). The enzyme and fusion proteins can be used to identify
CC ligands and drugs binding to it. These compounds are useful
CC antidepressant, antiasthmatic and antiinflammatory agents.
SQ Sequence 3890 BP; 1127 A; 846 C; 852 G; 1065 T;

Query Match 5.4%; Score 16; DB 12; Length 3890;
Best Local Similarity 100.0%; Pred. No. 6.37e+00;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1914 gattccctgggagaga 1929
||||||||||||||||
Cp 21 GATTCCCTGGGAGAGA 6

RESULT 15
ID Q76548 standard; DNA; 148 BP.
AC Q76548;
DT 23-SEP-1994 (first entry)
DE Human genome fragment. (Preferred)
KW Brain; placenta; bone marrow; genetic analysis; gene mapping;
KW detection; homology; human; adrenal tissue; ds.
OS Homo sapiens.
PN WO9401548-A.
PD 20-JAN-1994.
PF 13-JUL-1993; G01467.
PR 13-JUL-1992; GB-014857.
PA (MEDI-) MEDICAL RES COUNCIL.
PI Gross J, Hadfield KM, Howells D, Kelly M, Shaw D;
PI Sibson DR, Starkey M;
DR WPI; 94-035056/04.
PT New nucleic acid fragment encoding gene products - can be used
PT for genetic analysis and mapping
PS Claim 1; Page 91; 616pp; English.
CC Human nucleic acid fragments, isolated from brain adrenal tissue,
CC the placenta or bone marrow comprise any of: (A) a sequence
CC selected from (Q76401-Q77613), (B) an allelic variation of a
CC sequence as described in (A), or (C) a sequence complementary
CC to (A) or (B).

CC Preferred sequences exhibit no more than 90% homology to a human
CC sequence known per se.
SQ Sequence 148 BP; 47 A; 15 C; 24 G; 57 T;

Query Match 5.0%; Score 15; DB 10; Length 148;
Best Local Similarity 100.0%; Pred. No. 2.78e+01;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 113 tacatttataaaaaa 127
|||||||||||||||
Cp 110 TACATTTATAAAAAA 96

RESULT 16
ID Q76546 standard; DNA; 151 BP.
AC Q76546;
DT 23-SEP-1994 (first entry)
DE Human genome fragment. (Preferred)
KW Brain; placenta; bone marrow; genetic analysis; gene mapping;
KW detection; homology; human; adrenal tissue; ds.
OS Homo sapiens.
PN WO9401548-A.
PD 20-JAN-1994.
PF 13-JUL-1993; G01467.
PR 13-JUL-1992; GB-014857.
PA (MEDI-) MEDICAL RES COUNCIL.
PI Gross J, Hadfield KM, Howells D, Kelly M, Shaw D;
PI Sibson DR, Starkey M;
DR WPI; 94-035056/04.
PT New nucleic acid fragment encoding gene products - can be used
PT for genetic analysis and mapping
PS Claim 1; Page 90-91; 616pp; English.
CC Human nucleic acid fragments, isolated from brain adrenal tissue,
CC the placenta or bone marrow comprise any of: (A) a sequence
CC selected from (Q76401-Q77613), (B) an allelic variation of a
CC sequence as described in (A), or (C) a sequence complementary
CC to (A) or (B).
CC Preferred sequences exhibit no more than 90% homology to a human
CC sequence known per se.
SQ Sequence 151 BP; 49 A; 14 C; 25 G; 58 T;

Query Match 5.0%; Score 15; DB 10; Length 151;
Best Local Similarity 100.0%; Pred. No. 2.78e+01;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 115 tacatttataaaaaa 129
|||||||||||||||
Cp 110 TACATTTATAAAAAA 96

RESULT 17

ID Q76549 standard; DNA; 204 BP.
AC Q76549;
DT 23-SEP-1994 (first entry)
DE Human genome fragment. (Preferred)
KW Brain; placenta; bone marrow; genetic analysis; gene mapping;
KW detection; homology; human; adrenal tissue; ds.
OS Homo sapiens.
PN WO9401548-A.
PD 20-JAN-1994.
PF 13-JUL-1993; G01467.
PR 13-JUL-1992; GB-014857.
PA (MEDI-) MEDICAL RES COUNCIL.
PI Gross J, Hadfield KM, Howells D, Kelly M, Shaw D;
PI Sibson DR, Starkey M;
DR WPI; 94-035056/04.
PT New nucleic acid fragment encoding gene products - can be used
PT for genetic analysis and mapping
PS Claim 1; Page 92; 616pp; English.
CC Human nucleic acid fragments, isolated from brain adrenal tissue,
CC the placenta or bone marrow comprise any of: (A) a sequence
CC selected from (Q76401-Q77613), (B) an allelic variation of a
CC sequence as described in (A), or (C) a sequence complementary
CC to (A) or (B).
CC Preferred sequences exhibit no more than 90% homology to a human
CC sequence known per se.
SQ Sequence 204 BP; 58 A; 18 C; 41 G; 81 T;

Query Match 5.0%; Score 15; DB 10; Length 204;
Best Local Similarity 100.0%; Pred. No. 2.78e+01;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 115 tacatttataaaaaa 129
|||||||||||||||
Cp 110 TACATTTATAAAAAA 96

RESULT 18
ID T21944 standard; cDNA to mRNA; 212 BP.
AC T21944;
DT 14-AUG-1996 (first entry)
DE Human gene signature HUMGS03487.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN WO9514772-A1.
PD 01-JUN-1995.
PF 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATS/) MATSUBARA K.
PA (OKUB/) OKUBO K.

```
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 1007; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   212 BP;    76 A;    26 C;    29 G;    74 T;

  Query Match           5.0%;  Score 15;  DB 19;  Length 212;
  Best Local Similarity  93.8%;  Pred. No. 2.78e+01;
  Matches    15;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db     16 ttagcatctgcaaatt 31
          ||||||||||||| |||
Cp    142 TTAGCATCTGCANATT 127


RESULT   19
ID   T19597 standard; cDNA to mRNA; 252 BP.
AC   T19597;
DT   28-JUN-1996  (first entry)
DE   Human gene signature HUMGS00664.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
```



```
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 427; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   252 BP;    82 A;    29 C;    38 G;    99 T;

  Query Match           5.0%;  Score 15;  DB 18;  Length 252;
  Best Local Similarity 100.0%;  Pred. No. 2.78e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     22 tttataaatgtataa 36
          ||||||||||||||| 
Qy     99 TTTATAAATGTATAA 113


RESULT   20
ID   Q61191 standard; DNA; 342 BP.
AC   Q61191;
DT   16-MAR-1994  (first entry)
DE   Human brain Expressed Sequence Tag EST01229.
KW   Gene transcription product; genetic markers; tagging; in vivo;
KW   transcriptiton; mapping; locations; chromosomes; chromosomal; ss.
OS   Homo sapiens.
PN   WO9316178-A.
PD   19-AUG-1993.
PF   12-FEB-1993; U01294.
PR   12-FEB-1992; US-837195.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
PI   Adams MD, Moreno RF, Venter CJ;
DR   WPI; 93-272882/34.
PT   Enriched oligonucleotides and corresp. sequences - used as
PT   markers for human genes transcribed in-vivo, facilitate tagging
PT   of most human genes
PS   Example 4; Page 452; 500pp; English.
CC   The Expressed Sequence Tag was isolated from a human brain cDNA
CC   library as part of a large set of ESTs which can be used as markers
```



```
CC   for human genes transcribed in vivo. They can be used to facilitate
CC   tagging of most human genes, for mapping locations of expressed genes
CC   on chromosomes, for individual or forensic identification, for mapping
CC   locations of disease-associated genes, for identification of tissue
CC   type, and for prepn. of antisense sequences, probes and constructs.
CC   EST01229 has a "poor" coding probability as evaluated using the
CC   coding-region prediction program CRM. See also Q59041-Q61440.
SQ   Sequence   342 BP;    82 A;    84 C;    101 G;    73 T;

  Query Match           5.0%;  Score 15;  DB 8;  Length 342;
  Best Local Similarity  93.8%;  Pred. No. 2.78e+01;
  Matches    15;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db    320 ctgggcaccaggggtg 335
          ||||||||||||| |||
Qy    160 CTGGGCACCAGGNGTG 175


RESULT   21
ID   Q60937 standard; DNA; 370 BP.
AC   Q60937;
DT   16-MAR-1994  (first entry)
DE   Human brain Expressed Sequence Tag EST01020.
KW   Gene transcription product; genetic markers; tagging; in vivo;
KW   transcription; mapping; locations; chromosomes; chromosomal; ss.
OS   Homo sapiens.
PN   WO9316178-A.
PD   19-AUG-1993.
PF   12-FEB-1993; U01294.
PR   12-FEB-1992; US-837195.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
PI   Adams MD, Moreno RF, Venter CJ;
DR   WPI; 93-272882/34.
PT   Enriched oligonucleotides and corresp. sequences - used as
PT   markers for human genes transcribed in-vivo, facilitate tagging
PT   of most human genes
PS   Example 4; Page 413-414; 500pp; English.
CC   The Expressed Sequence Tag was isolated from a human brain cDNA
CC   library as part of a large set of ESTs which can be used as markers
CC   for human genes transcribed in vivo. They can be used to facilitate
CC   tagging of most human genes, for mapping locations of expressed genes
CC   on chromosomes, for individual or forensic identification, for mapping
CC   locations of disease-associated genes, for identification of tissue
CC   type, and for prepn. of antisense sequences, probes and constructs.
CC   EST01020 has a "poor" coding probability as evaluated using the
CC   coding-region prediction program CRM. See also Q59041-Q61440.
SQ   Sequence   370 BP;    79 A;    88 C;    123 G;    75 T;

  Query Match           5.0%;  Score 15;  DB 8;  Length 370;
  Best Local Similarity 100.0%;  Pred. No. 2.78e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```



```
Db    166 cctggtgcccagttc 180
          |||||||||||||||
Cp    171 CCTGGTGCCCAGTTC 157


RESULT   22
ID   T12305 standard; DNA; 534 BP.
AC   T12305;
DT   16-JUL-1996  (first entry)
DE   Gyrase A probe.
KW   Gyrase A; amplification; detection; point mutation; quinolone;
KW   antibiotic; resistance; resistant; primer; probe;
KW   Staphylococcus aureus; ss.
OS   Synthetic.
PN   EP-688873-A2.
PD   27-DEC-1995.
PF   09-JUN-1995; 108890.
PR   23-JUN-1994; DE-421901.
PA   (FARB ) BAYER AG.
PI   Bremm K,  Endermann R,  Springer W;
DR   WPI; 96-041584/05.
PT   Amplification primers and probes for rapid detection of
PT   Staphylococcus  - resistant to quinolone antibiotics in clinical
PT   samples
PS   Claim 5; Page 16; 22pp; German.
CC   The gyrase A sequences (nucleotides 2394-2658) of ca. 50
CC   quinolone resistant S. aureus isolates were analysed.
CC   Seven variant sequences were found (T12306-T12312). 4 types of
CC   mutations were identified: TCA to TTA; ATT to ATC; GAA to AAA
CC   and TCA to GCA.  The 5th mutation (GAC to GAT) does not result
CC   in an amino acid substitution. Amplification primers
CC   (T12300-T12303) and probes (T12304 and T12305) have been produced
CC   for the rapid detection of the above mutations.
CC   Primer 1 and 4 detect the C/T mutation at position 2533.
CC   Primer 2 and 4 detect the T/G mutation at position 2532.
CC   Primer 3 and 4 detect the G/A mutation at position 2544.
SQ   Sequence   534 BP;    157 A;    94 C;    123 G;    160 T;

  Query Match           5.0%;  Score 15;  DB 19;  Length 534;
  Best Local Similarity 100.0%;  Pred. No. 2.78e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    161 cacacttgaactgtt 175
          |||||||||||||||
Qy    224 CACACTTGAACTGTT 238


RESULT   23
ID   Q23090 standard; DNA; 732 BP.
AC   Q23090;
```

```
DT   17-AUG-1992  (first entry)
DE   Antigen ac-1b gene.
KW   Oocysts; MAb 12-09; sporozoite; ss.
OS   Eimeria acervulina.
PN   WO9204460-A.
PD   19-MAR-1992.
PF   05-SEP-1991; U06430.
PR   12-SEP-1990; US-581693.
PA   (GENE-) GENEX CORP.
PI   Jacobson JW, Strausberg RL, Wilson SD, Pope SH, Strausberg SL;
PI   Ruff MD, Augustine PC, Danforth HD;
DR   WPI; 92-114365/14.
DR   P-PSDB; R22390.
PT   Vaccine against avian coccidiosis - comprising recombinant
PT   Eimeria antigen ac-1b or ac-6b gene, or microorganisms expressing
PT   them
PS   Claim 1; Page 39 + Fig 1; 56pp; English.
CC   To identify antigens of E. acervulina, expression libraries were
CC   prepd. in lambda vector, lambda gt11, using cDNA prepd. from polyA
CC   mRNA isolated from E. acervulina oocysts.  The cDNA expression
CC   library was screened with monoclonal antibody (MAb) 12-09 which was
CC   raised against the sporozoite stage of E. acervulina.  The library
CC   to be screened was plated on a host that allows lysis and plaque
CC   formation.  During induction of the antigens encoded by the phage,
CC   the plaques were identified by screening the filters with MAb 12-09.
CC   The cDNA inserts from the MAb 12-09 positive phage were cloned into
CC   bacteriophage M13 and subjected to sequence analysis.  Following
CC   sequence analysis, E. acervulina antigen ac-1b was identified.
SQ   Sequence   732 BP;    212 A;    180 C;    221 G;    119 T;

  Query Match             5.0%;  Score 15;  DB 4;  Length 732;
  Best Local Similarity 100.0%;  Pred. No. 2.78e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    102 gaatccgacacagga 116
          ||||||||||||||||
Qy     17 GAATCCGACACAGGA 31


RESULT   24
ID   Q40029 standard; DNA; 1131 BP.
AC   Q40029;
DT   13-AUG-1993  (first entry)
DE   Sequence encoding apo B-100.
KW   Lipoprotein; apoprotein; B-100; A-I; LDL; HDL; assay; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             1..1131
FT                   /*tag=  a
PN   WO9307165-A.
PD   15-APR-1993.
```



```
PF   09-OCT-1992; U08634.
PR   09-OCT-1991; US-774633.
PR   28-JUN-1992; US-901706.
PR   08-OCT-1992; US-555555.
PA   (SCRI ) SCRIPPS RES INST.
PI   Curtiss LK,  Koduri KR,  Smith RS,  Witztum JL,  Young SG;
DR   WPI; 93-134370/16.
DR   P-PSDB; R34031.
PT   Polypeptide mimic of native apo B-100 and native apo A-I - useful
PT   in assays for LDL and HDL in plasma samples
PS   Claim 10; Pages 103-104; 137pp; English.
CC   The inventors claim a portion of the polypeptide contg. apo B-100
CC   that immunoreacts with antibodies secreted by the hybridoma MB47
CC   having ATCC Accession No. 8746.  Polypeptides specifically claimed
CC   include residues 217-297, 216-310, 216-331, 216-352, 216-377, 1-377,
CC   205-297, 173-297, 140-297.  DNA sequences encoding the polypeptides
CC   are also claimed.  Also claimed are a fusion polypolypeptide that
CC   contains: (a) a first amino acid residue sequence up to 250 residues
CC   in length that includes residues 120-135 of apo A-I, (b) a second
CC   amino acid residue sequence up to 375 residues in length that
CC   includes residues 217-297 of apo B-100 and DNA encoding it.
SQ   Sequence   1131 BP;    325 A;    277 C;    210 G;    319 T;

  Query Match             5.0%;  Score 15;  DB 7;  Length 1131;
  Best Local Similarity  93.8%;  Pred. No. 2.78e+01;
  Matches   15;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db    152 gtgtgccttcatacac 167
          | ||||||||||||||
Qy    171 GNGTGCCTTCATACAC 186


RESULT   25
ID   Q89633 standard; cDNA; 1131 BP.
AC   Q89633;
DT   31-OCT-1995  (first entry)
DE   Human apo B-100 encoding polypeptide s9.
KW   Apo B-100; hybridoma MB47; LDL cholesterol; low density lipoprotein;
KW   lipid; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             1..1131
FT                   /*tag=  a
PN   US5408038-A.
PD   18-APR-1995.
PF   09-OCT-1991; 774633.
PR   09-OCT-1991; US-774633.
PR   18-JUN-1992; US-901706.
PR   08-OCT-1992; US-959946.
PA   (SCRI ) SCRIPPS RES INST.
PI   Curtiss LK,  Koduri KR,  Smith RS,  Witztum JL,  Young SG;
```



```
DR   WPI; 95-161146/21.
DR   P-PSDB; R72704.
PT   New apo:lipoprotein B-100 peptide(s) and fusion peptide(s) - used
PT   in assay systems for detecting LDL and HDL cholesterol levels in
PT   body fluids.
PS   Disclosure; Fig 1; 41pp; English.
CC   Q89633 and R72704 depict the AA sequence of human apo B-100 and its
CC   corresp. cDNA, from AA residue 3214 through residue 3590 according
CC   to the numbering scheme of Ludwig et al., DNA, 6:363 (1987). Ludwig
CC   et al. reported the full cDNA sequence for the human apo B-100 gene,
CC   consisting of 29 exons. Q89633 corresp. to nts 9642-10770. The
CC   apo B-100 cDNA segment in Q89633 encodes an apo B-100 polypeptide
CC   called s9 . Other truncated AA sequences and cDNA sequences of apo B-
CC   100 identified in the disclosure are:
CC   No.     N terminus      C terminus
CC   AA      nt      AA      nt      AA
CC   s1      3429    10287   3510    10530
CC   s2      3418    10254   3510    10530
CC   s3      3386    10158   3510    10530
CC   s4      3353    10059   3510    10530
CC   s5      3429    10287   3523    10569
CC   s6      3429    10287   3544    10632
CC   s7      3429    10287   3565    10695
CC   s8      3429    10287   3590    10770
CC   s9      3214    9642    3590    10770
CC   Forward (F) and reverse (R) primers that can be used to generate a
CC   cDNA segment coding for the above fragments are given in Q86935-
CC   Q86944.
SQ   Sequence   1131 BP;    325 A;    277 C;    210 G;    319 T;

  Query Match             5.0%;  Score 15;  DB 14;  Length 1131;
  Best Local Similarity  93.8%;  Pred. No. 2.78e+01;
  Matches   15;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db    152 gtgtgccttcatacac 167
          | ||||||||||||||
Qy    171 GNGTGCCTTCATACAC 186


RESULT   26
ID   N91115 standard; DNA; 1471 BP.
AC   N91115;
DT   22-JUN-1990  (first entry).
DE   Portion of clone P4A encoding flagellin gene of Bacillus subtillis.
KW   Flagellin; fusion protein; heterologous protein; ds.
OS   Bacillus subtilis.
FH   Key             Location/Qualifiers
FT   cds             176..1088
FT                   /*tag=  a
FT   misc_difference 1090..1090
FT                   /*tag=  b
```



```
FT                   /label= T base should be G
PN   US4801536-A.
PD   31-JAN-1989.
PF   2-JUN-1987; 57881.
PR   11-OCT-1985; US-786749.
PR   10-OCT-1986; WO-U02168.
PR   02-JUN-1987; US-57881.
PA   (GENE-) Genetics Inst Inc.
PI   Stahl ML, LaVallie ER;
DR   WPI; 89-053638/07.
DR   P-PSDB; P94623.
PT   Prodn. of heterologous protein in bacterial host cells -
PT   exported into the culture medium as fusion protein with
PT   flagellin protein native to host.
PS   Disclosure; 4pp; English.
CC   By ligating the flagellin coding sequence to a protein required to be
CC   secreted into the medium, the product can be collected. If a linker
CC   encoding selectively cleavable peptide is used in ligation, the flagellin
CC   portion may be cleaved from the product.
SQ   Sequence   1471 BP;    492 A;    292 C;    284 G;    403 T;

  Query Match             5.0%;  Score 15;  DB 1;  Length 1471;
  Best Local Similarity 100.0%;  Pred. No. 2.78e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1202 tttttataaatgtat 1216
          ||||||||||||||||
Qy     97 TTTTTATAAATGTAT 111


RESULT   27
ID   N70984 standard; DNA; 1472 BP.
AC   N70984;
DT   18-APR-1991  (first entry)
DE   Sequence of flagellin gene of B. subtilis in clone p4A.
KW   Heterologous protein prodn; flagellin; ss.
OS   B. subtilis.
FH   Key             Location/Qualifiers
FT   cds             178..1092
FT                   /*tag=  a
PN   WO8702385-A.
PD   23-APR-1987.
PF   10-OCT-1986; U02168.
PR   11-OCT-1985; US-786749.
PR   02-JUN-1987; US-057881.
PA   (GENE-) GENETICS INST.
PI   Stahl ML, Lavallie E;
DR   WPI; 87-122742/17.
DR   P-PSDB; P70637.
PT   Heterologous protein prodn. in bacterial host cell - using
PT   nucleotide sequence encoding at least N-terminal portion of
```

```
PT   flagellin protein
PS   Disclosure; Table 1, p10A; 42pp; English.
CC   The DNA sequence encoding the N-terminal portion of flagellin, e.g.
CC   a portion of the hag gene of B. subtilis, is operatively linked to
CC   an expression control sequence. The flagellin gene is linked via its
CC   3' terminus to the 5' terminus of a heterologous protein via a
CC   linking nucleotide sequence which encodes a selectably cleavable
CC   polypeptide such that the exported protein contains a selectably
CC   cleavable site. Proteins which may be produced include hormones.
SQ   Sequence   1472 BP;    485 A;    296 C;    287 G;    404 T;

  Query Match           5.0%;  Score 15;  DB 2;  Length 1472;
  Best Local Similarity 100.0%;  Pred. No. 2.78e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1202 tttttataaatgtat 1216
        ||||||||||||||||
Qy     97 TTTTTATAAATGTAT 111


RESULT   28
ID   N40290 standard; cDNA; 1578 BP.
AC   N40290;
DT   06-NOV-1992  (first entry)
DE   Sequence of the alpha-fetoprotein cDNA
KW   Characterisation method; identification; chromatographic pattern;
KW   restriction endonuclease cleavage site; hybridisation; ds.
OS   Homo sapiens.
PN   EP-120658-A.
PD   03-OCT-1984.
PF   16-MAR-1984; 301800.
PR   21-MAR-1983; US-477431.
PA   (WEBS/) WEBSTER J A.
PI   Webster JA;
DR   WPI; 84-245180/40.
PT   Characterisation of unknown organism - by determn. of
PT   evolutionarily conserved sequences in genetic material relative
PT   to other organisms
PS   Example; Page 97-99; 130pp; English.
CC   The inventors claim a method for the characterisation of an unknown
CC   organism by determn. of evolutionarily conserved sequences in
CC   genetic material relative to other organisms.  The method permits
CC   rapid, reliable and accurate characterisation and identification of
CC   organisms, including prokaryotic and eukaryotic organisms, such as
CC   bacteria, plants and animals. For example, the alpha-fetoprotein
CC   (AFP) gene was used as a probe to distinguish humans from rats
CC   on the basis of RFLP following digestion with endonuclease MnlI.
CC   It was calculated that fragments of 24, 34, 104 and 108 bps
CC   describe human DNA, whilst fragments of 24, 59, 90 and 145
CC   bps describe rat DNA. Whilst both sequences contain the 24bp
CC   fragment, it is the set of fragments (taxonomic characters)
```



```
CC   that is of significance.
SQ   Sequence   1578 BP;    472 A;    368 C;    374 G;    364 T;

  Query Match           5.0%;  Score 15;  DB 4;  Length 1578;
  Best Local Similarity 100.0%;  Pred. No. 2.78e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    363 aggaggaacccctto 377
          |||||||||||||||
Qy     28 AGGAGGAACCCCTTC 42


RESULT   29
ID   T72149 standard; cDNA; 2155 BP.
AC   T72149;
DT   22-AUG-1997  (first entry)
DE   Murine mortalin cDNA sequence.
KW   Murine; mortalin; mouse embryonic fibroblast; MEF; heat shock protein;
KW   complementation group; immortalised cell; tumour cell; microtumour;
KW   cancer; ss.
OS   Mus musculus.
FH   Key             Location/Qualifiers
FT   cds             90..2129
FT                   /*tag=  a
FT                   /product= Mortalin
PN   US5627039-A.
PD   06-MAY-1997.
PF   18-MAR-1994; 214583.
PR   18-MAR-1994; US-214583.
PA   (BAYU ) BAYLOR COLLEGE MEDICINE.
PI   Pereira-Smith OM,  Wadhwa R;
DR   WPI; 97-271308/24.
DR   P-PSDB; W19911.
PT   Determination of the complementation group of tumour cells - by
PT   determining intracellular mortalin distribution
PS   Example 1; Fig 1; 18pp; English.
CC   This sequence encodes murine mortalin protein.  Mortalin is found in
CC   mouse embryonic fibroblasts.  It has significant homology to heat shock
CC   proteins but is not heat shock inducible.  The amino and carboxy termini
CC   of mortalin differ significantly from other heat shock family members.
CC   The invention concerns a method for determining the complementation group
CC   of an immortalised cell.  The method comprises: (a) determining the
CC   intracellular distribution of a mortalin in the cell; and (b) correlating
CC   the determined distribution with the mortalin distribution in cells of
CC   complementation group A, B, C or D.  The capacity to readily determine
CC   the complementation group of a tumour cell would provide a valuable
CC   means for assessing the severity of the disease and determining the
CC   aggressiveness needed to treat the cancer. A sensitive means for
CC   discerning tumour cells in microtumours, or for determining if a
CC   tumour contains different classes of tumour cells would assist in
CC   the diagnosis and treatment of cancer.
```



```
SQ   Sequence   2155 BP;    609 A;    450 C;    581 G;    515 T;

  Query Match           5.0%;  Score 15;  DB 31;  Length 2155;
  Best Local Similarity 93.8%;  Pred. No. 2.78e+01;
  Matches   15;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db    348 acccctttctgtggttg 363
          ||||||||| ||||||
Qy     35 ACCCCTTCTNTGGTTG 50


RESULT   30
ID   Q05688 standard; DNA; 2599 BP.
AC   Q05688;
DT   03-JAN-1991  (first entry)
DE   Islets of Langerhans cell clone ICA525 (ATCC 40704).
KW   Pancreatic Islet Cell Antigen (ICA) 525; ATCC 40704;
KW   insulin-dependent (Type I) diabetes mellitus; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             300..2438
FT                   /*tag=  a
PN   EP-383129-A.
PD   22-AUG-1990.
PF   03-FEB-1990; 102179.
PR   17-FEB-1989; US-312543.
PR   04-DEC-1989; US-441703.
PA   (MOLE-) MOLECULAR DIAGNOSTI.
PI   Rabin D;.
DR   WPI; 90-255450/34.
DR   P-PSDB; R06471.
PT   Pancreatic islet cell antigens obtd. by mol. cloning - used in
PT   research, diagnosis and treatment of insulin-dependent (Type I)
PT   diabetes mellitus.
PS   Disclosure; Fig 18; 45pp; English.
CC   A cDNA library was constructed from RNA extracted from purified
CC   human islets of Langerhans cells. The library was screened with
CC   sera from autoimmune patients with Type I diabetes mellitus.
CC   ICA 525 is one of nine clones found to react predominantly with
CC   diabetic sera. The protein expressed by the clone was analysed by
CC   expressing the clones in E.coli hosts. The protein or peptide
CC   fragments of it can be used as immunoassay reagents in the
CC   presymptomatic diagnosis of insulin-dependent diabetes mellitus
CC   (IDDM).
CC   They may also be used to identify or mark T- or B-cells involved in
CC   IDDM or to bind or block these same cells for therapy.
CC   See also Q05681-Q05687.
SQ   Sequence   2599 BP;    849 A;    539 C;    605 G;    606 T;

  Query Match           5.0%;  Score 15;  DB 1;  Length 2599;
  Best Local Similarity 100.0%;  Pred. No. 2.78e+01;
```



```
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2396 gtgggcacacttgaa 2410
          ||||||||||||||
Qy    219 GTGGGCACACTTGAA 233


RESULT   31
ID   V02185 standard; cDNA; 4588 BP.
AC   V02185;
DT   12-MAY-1998  (first entry)
DE   Murine type C lectin encoding cDNA - E-selectin homologue.
KW   Murine; type C lectin; E-selectin; competitive inhibitor;
KW   mouse; molecular marker; ss.
OS   Mus sp.
FH   Key             Location/Qualifiers
FT   CDS             105..4544
FT                   /*tag=  a
FT                   /product= "type C lectin"
FT                   /transl_except= (pos:300..302,aa:Val)
PN   WO9740154-A1.
PD   30-OCT-1997.
PF   17-APR-1997; U06347.
PR   24-APR-1996; US-637021.
PA   (GETH ) GENENTECH INC.
PI   Lasky LA, Wu K.
DR   WPI; 97-535838/49.
DR   P-PSDB; W44118.
PT   Human and mouse type C lectin(s) - useful as competitive inhibitor
PT   of lectin activity and as molecular markers for tissues that express
PT   them
PS   Example 1; Page 39-42; 97pp; English.
CC   The present sequence encodes a novel type C lectin. Type C lectins
CC   can be used to identify and purify their native ligands and compete
CC   with them for their binding, especially useful as competitive
CC   inhibitors of the biological activity of native type C lectins.
CC   This makes type C lectins useful as molecular markers for tissues
CC   in which they are expressed. The nucleic acid, encoding type C lectin,
CC   can be used to identify other type C lectin coding sequences.
SQ   Sequence   4588 BP;    987 A;    1361 C;    1373 G;    867 T;

  Query Match           5.0%;  Score 15;  DB 38;  Length 4588;
  Best Local Similarity 93.8%;  Pred. No. 2.78e+01;
  Matches   15;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db   3270 aactgggcaccaggag 3285
          |||||||||||||| |
Qy    158 AACTGGGCACCAGGNG 173


RESULT   32
```

```
ID   Q31189 standard; cDNA; 5499 BP.
AC   Q31189;
DT   23-MAR-1993  (first entry)
DE   Alpha 6B integrin subunit cDNA.
KW   Human; alpha 6A; alpha 6B; integrin; cell surface receptor; adhesion;
KW   extracellular matrix; cytoskeleton; heterodimer; laminin receptor; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   variation       3260..3261
FT                   /*tag=  a
FT                   /label= "Position of 130 bp deletion cf alpha 6A
FT                   subunit"
PN   WO9219647-A.
PD   12-NOV-1992.
PF   27-APR-1992; U03527.
PR   03-MAY-1991; US-695564.
PA   (SCRI ) SCRIPPS RES INST.
PI   Quaranta V, Tamura RN;
DR   WPI; 92-398799/48.
DR   P-PSDB; R28822.
PT   Integrin alpha sub-unit cytoplasmic domain polypeptide(s) - used
PT   for prodn. of antibodies and in detection of integrin sub-units
PT   in body samples
PS   Disclosure; Page 83-87; 115pp; English.
CC   The sequences given in Q31188-89 encode the human alpha 6A and 6B
CC   integrin subunits.  Integrins are a family of cell surface receptors
CC   which serve cellular adhesion functions.  These receptors form a link
CC   between the extracellular matrix and the cytoskeleton through their
CC   binding to various extracellular components.  Each integrin receptor
CC   is a heterodimer comprised of an alpha and a beta subunit.  Each alpha
CC   subunit tends to associate with only one type of beta subunit but
CC   there are several exceptions to this rule.  These integrins correspond
CC   to the laminin receptor. The cytoplasmic domain of the 6A and 6B
CC   integrins differs from previously isolated alpha 6 integrins.
SQ   Sequence   5499 BP;    1624 A;    1073 C;    1258 G;    1544 T;

  Query Match             5.0%;  Score 15;  DB 5;  Length 5499;
  Best Local Similarity 100.0%;  Pred. No. 2.78e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1701 gtgggcacacttgaa 1715
          |||||||||||||||
Qy    219 GTGGGCACACTTGAA 233


RESULT   33
ID   Q65673 standard; DNA; 5529 BP.
AC   Q65673;
DT   27-JAN-1995  (first entry)
DE   Sequence encoding alpha subunit of integrin cell surface receptor.
KW   Integrin; alpha; beta; subunit; glycoprotein; heterodimer;
```



```
KW   transmembrane; extracellular matrix; cell signalling; cytoskeleton;
KW   behaviour; signal transduction; receptor; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             147..3368
FT                   /*tag=  a
FT                   /product= Alpha6 subunit of integrin.
FT   misc_binding    230..238
FT                   /*tag=  b
FT                   /note= "Sequence weakly homologous with cation
FT                   binding site of other integrin alpha
FT                   subunits."
FT   misc_binding    324..332
FT                   /*tag=  c
FT                   /label= Putative cation binding site.
FT   misc_binding    386..394
FT                   /*tag=  d
FT                   /label= Putative caion binding site.
FT   misc_binding    441..449
FT                   /*tag=  e
FT                   /label= Putative cation binding site.
FT   misc_feature    899..903
FT                   /*tag=  f
FT                   /note= "This sequence closely resembles the
FT                   cleavage site of other integrin O chains."
FT   misc_feature    1040..1044
FT                   /*tag=  g
FT                   /note= "This sequence is conserved in all but the
FT                   Drosophila PS2 alpha subunit"
PN   US5320942-A.
PD   14-JUN-1994.
PF   19-FEB-1987; 016552.
PR   19-FEB-1987; US-016552.
PR   04-JAN-1989; US-293384.
PR   01-OCT-1990; US-591105.
PA   (KAJI/) KAJIJI S.
PA   (QUAR/) QUARANTA V.
PI   Kajiji S,  Quaranta V;
DR   WPI; 94-191533/23.
DR   P-PSDB; R55272.
PT   Diagnosing presence of abnormal epithelial tissue in vitro -
PT   utilises monoclonal antibodies to alpha6 beta4 cell surface protein
PS   Example 5; Figure 6; 34pp; English.
CC   Integrins are heterodimers comrised of alpha and beta subunits, that
CC   are non-covalently associated transmembrane glycoproteins.  11 alpha
CC   chains and 6 beta chains have been recognised in man.  Each alpha
CC   subunit tends to associate with only one type of beta subunit but
CC   there are several exceptions.  Integrins mediate (in part) the
CC   interaction of cells with the extracellular matrix, forming a link
CC   between the extracellular matrix and the cytoskeleton.  They may
CC   transmit signals from the extracellular to the intracellular
```



```
CC   environment, affecting cell behaviour.  This sequence encodes the
CC   alpha6 subunit of an alpha6 beta4 integrin.
SQ   Sequence   5529 BP;    1641 A;    1077 C;    1268 G;    1543 T;

  Query Match             5.0%;  Score 15;  DB 11;  Length 5529;
  Best Local Similarity 100.0%;  Pred. No. 2.78e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1701 gtgggcacacttgaa 1715
          |||||||||||||||
Qy    219 GTGGGCACACTTGAA 233


RESULT   34
ID   Q31188 standard; cDNA; 5629 BP.
AC   Q31188;
DT   23-MAR-1993  (first entry)
DE   Alpha 6A integrin subunit cDNA.
KW   Human; alpha 6A; alpha 6B; integrin; cell surface receptor; adhesion;
KW   extracellular matrix; cytoskeleton; heterodimer; laminin receptor; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   misc_feature    147..149
FT                   /*tag=  a
FT                   /note= "Transcription initiator"
FT   mat_peptide     216..3365
FT                   /*tag=  b
FT   misc_feature    3264..3278
FT                   /*tag=  c
FT                   /note= "Cytoplasmic sequence of GFFKR"
PN   WO9219647-A.
PD   12-NOV-1992.
PF   27-APR-1992; U03527.
PR   03-MAY-1991; US-695564.
PA   (SCRI ) SCRIPPS RES INST.
PI   Quaranta V, Tamura RN;
DR   WPI; 92-398799/48.
DR   P-PSDB; R28821.
PT   Integrin alpha sub-unit cytoplasmic domain polypeptide(s) - used
PT   for prodn. of antibodies and in detection of integrin sub-units
PT   in body samples
PS   Disclosure; Page 74-77; 115pp; English.
CC   The sequences given in Q31188-89 encode the human alpha 6A and 6B
CC   integrin subunits.  Integrins are a family of cell surface receptors
CC   which serve cellular adhesion functions.  These receptors form a link
CC   between the extracellular matrix and the cytoskeleton through their
CC   binding to various extracellular components.  Each integrin receptor
CC   is a heterodimer comprised of an alpha and a beta subunit.  Each alpha
CC   subunit tends to associate with only one type of beta subunit but
CC   there are several exceptions to this rule.  These integrins correspond
CC   to the laminin receptor. The cytoplasmic domain of the 6A and 6B
```



```
CC   integrins differs from previously isolated alpha 6 integrins.
SQ   Sequence   5629 BP;    1664 A;    1099 C;    1281 G;    1584 T;

  Query Match             5.0%;  Score 15;  DB 5;  Length 5629;
  Best Local Similarity 100.0%;  Pred. No. 2.78e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1701 gtgggcacacttgaa 1715
          |||||||||||||||
Qy    219 GTGGGCACACTTGAA 233


RESULT   35
ID   N50530 standard; DNA; 5760 BP.
AC   N50530;
DT   30-SEP-1991  (first entry)
DE   Sequence encoding the P195 protein of Plasmodium falciparum.
KW   Malaria vaccine; epitope; antigen; immunogen; ss.
OS   Plasmodium falciparum.
FH   Key             Location/Qualifiers
FT   cds             216..5179
FT                   /*tag=  a
PN   EP-154454-A.
PD   11-SEP-1985.
PF   21-FEB-1985; 301173.
PR   22-FEB-1984; GB-004692.
PR   26-SEP-1984; GB-024340.
PR   21-FEB-1985; GB-004429.
PA   (WELL ) WELLCOME FOUNDATION LTD.
PI   Holder A, Sandhu J, Odink K, Lockyer M, Riveros-Moreno V;
DR   WPI; 85-224845/37.
DR   P-PSDB; P50777.
PT   Cloned DNA sequence encoding plasmodium falciparum protein -
PT   useful for expressing the protein for use in vaccines against
PT   malaria
PS   Claim 2; Fig 1; 51pp; English.
CC   The sequence encoding the P195  protein of Plasmodium falciparum
CC   (N50530) and a peptide comprising at least one of its epitopes
CC   (see P50777) are claimed. Also claimed is a vaccine for inducing
CC   immunity to malaria comprising the novel peptide or P195 or a
CC   peptide comprising at least one epitope when derived from the new
CC   DNA sequence, together with a carrier.
SQ   Sequence   5760 BP;    2565 A;    630 C;    725 G;    1840 T;

  Query Match             5.0%;  Score 15;  DB 3;  Length 5760;
  Best Local Similarity 100.0%;  Pred. No. 2.78e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   5399 atttatacatttata 5413
          |||||||||||||||
Cp    115 ATTTATACATTTATA 101
```

```
RESULT   36
ID   Q95896 standard; DNA; 18 BP.
AC   Q95896;
DT   21-FEB-1996  (first entry)
DE   Primer B (Group 12, set B) for marker D16S415, chromosome 16.
KW   primer; polymerase chain reaction; PCR; linkage study; locus;
KW   microsatellite marker sequence; automated genotyping; allele;
KW   polymorphism; detection; Homo sapiens; ss.
OS   Synthetic.
PN   WO9515400-A1.
PD   08-JUN-1995.
PF   05-DEC-1994; U13945.
PR   03-DEC-1993; US-160837.
PA   (UYJO ) UNIV JOHNS HOPKINS.
PI   Levitt RC;
DR   WPI; 95-215278/28.
PT   Kit for automated genotyping contg. pairs of PCR primers - designed
PT   to amplify polymorphic nucleotide repeat sequences, arranged in sets
PT   each with a characteristic fluorescence label, useful e.g. in
PT   detection of disease related genetic rearrangement
PS   Disclosure; Fig 7L-3; 104pp; English.
CC   The method aims to provide a collection of highly reproducible
CC   microsatellite marker sequences (MMS) at approx. 10-50 cM intervals
CC   throughout the human genome which can be detectably labelled. The
CC   MMS are polymorphic, simple sequence repeats and can be used in
CC   automated genotyping. esp. fluorescence-based. The primers correspond
CC   to the unique DNA sequence surrounding each marker, and PCR is used to
CC   detect each polymorphism. When the MMS show considerable polymorphism
CC   (ie. a difference in the number of repeats) between individuals, the
CC   markers can be particularly informative. The MMS can be ideal for
CC   linkage studies. Kits comprise at least 4 groups, of at least 3 sets,
CC   each comprising labelled primers for PCR amplification of the DNA.
CC   Group 12 primer pairs are shown in Q95883-914. The published size range
CC   of the D16S451 allele is 208-234 bp, and the degree of heterozygosity
CC   in the population is about 72%.
SQ   Sequence   18 BP;    4 A;    8 C;    3 G;    3 T;

  Query Match           4.7%;  Score 14;  DB 16;  Length 18;
  Best Local Similarity 100.0%;  Pred. No. 1.14e+02;
  Matches   14;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      5 cactgtaccccagc 18
          ||||||||||||||
Qy    184 CACTGTACCCCAGC 197


RESULT   37
ID   V00376 standard; DNA; 59 BP.
AC   V00376;
```



```
DT   23-APR-1998  (first entry)
DE   Insecticidal gene sequence modification oligonucleotide BTK94.
KW   Insecticidal protein; Bacillus thuringiensis; monocotyledonous plant;
KW   structural gene; maize; CryI(b); CryIIB; ss.
OS   Synthetic.
OS   Bacillus thuringiensis.
PN   US5689052-A.
PD   18-NOV-1997.
PF   19-SEP-1995; 530492.
PR   22-DEC-1993; US-172333.
PR   19-SEP-1995; US-530492.
PA   (MONS ) MONSANTO CO.
PI   Brown SM, Dean DA, Fromm ME, Sanders PR;
DR   WPI; 98-008070/01.
PT   Genes encoding insecticidal proteins of Bacillus thuringiensis -
PT   modified to enhance expression in monocotyledonous plants
PS   Example 1; Column 17; 86pp; English.
CC   The present sequence represents an oligonucleotide used in the present
CC   invention describing new structural genes capable of being expressed in
CC   a monocotyledonous plant. The new genes comprise modified nucleotide
CC   sequences which encode insecticidal proteins of Bacillus thuringiensis.
CC   The genes have been modified to reduce the usage of codons that are
CC   rare or semi-rare in monocotyledon DNA, thereby increasing
CC   transformation efficiency and/or increasing accumulation of the
CC   insecticidal protein in monocotyledon tissues.
SQ   Sequence   59 BP;    9 A;    25 C;    15 G;    10 T;

  Query Match           4.7%;  Score 14;  DB 38;  Length 59;
  Best Local Similarity 100.0%;  Pred. No. 1.14e+02;
  Matches   14;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     43 caagtgtgcccacc 56
          ||||||||||||||
Cp    231 CAAGTGTGCCCACC 218


RESULT   38
ID   T21539 standard; cDNA to mRNA; 109 BP.
AC   T21539;
DT   29-AUG-1996  (first entry)
DE   Human gene signature HUMGS02913.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
```



```
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 909; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   109 BP;    35 A;    25 C;    13 G;    35 T;

  Query Match           4.7%;  Score 14;  DB 20;  Length 109;
  Best Local Similarity 100.0%;  Pred. No. 1.14e+02;
  Matches   14;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     20 ttagagttcctgct 33
          ||||||||||||||
Cp     85 TTAGAGTTCCTGCT 72


RESULT   39
ID   T21176 standard; cDNA to mRNA; 187 BP.
AC   T21176;
DT   01-AUG-1996  (first entry)
DE   Human gene signature HUMGS02486.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
```



```
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 821; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   187 BP;    54 A;    47 C;    48 G;    37 T;

  Query Match           4.7%;  Score 14;  DB 19;  Length 187;
  Best Local Similarity 87.5%;  Pred. No. 1.14e+02;
  Matches   14;  Conservative    0;  Mismatches   2;  Indels   0;  Gaps   0;

Db    143 ccctccatgtggcccc 158
          |||| | |||||||||
Qy    294 CCCTNCNTGTGGCCCC 309


RESULT   40
ID   T24941 standard; cDNA to mRNA; 195 BP.
AC   T24941;
DT   05-NOV-1996  (first entry)
DE   Human gene signature HUMGS07054.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
```

```
PS   Claim 1; Page 1732; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   195 BP;    36 A;    65 C;    46 G;    43 T;

  Query Match           4.7%;  Score 14;  DB 22;  Length 195;
  Best Local Similarity 100.0%;  Pred. No. 1.14e+02;
  Matches   14;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     82 cctctctcccaggg 95
          ||||||||||||||
Qy      4 CCTCTCTCCCAGGG 17


RESULT   41
ID   T24104 standard; cDNA to mRNA; 210 BP.
AC   T24104;
DT   06-SEP-1996  (first entry)
DE   Human gene signature HUMGS06094.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 1531; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
```



```
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   210 BP;    45 A;    46 C;    48 G;    60 T;

  Query Match           4.7%;  Score 14;  DB 20;  Length 210;
  Best Local Similarity 93.3%;  Pred. No. 1.14e+02;
  Matches   14;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db     23 tcccagggaatccna 37
          ||||||||||||| |
Qy     10 TCCCAGGGAATCCGA 24


RESULT   42
ID   T23555 standard; cDNA to mRNA; 250 BP.
AC   T23555;
DT   02-SEP-1996  (first entry)
DE   Human gene signature HUMGS05407.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 1397; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
```



```
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   250 BP;    35 A;    77 C;    66 G;    59 T;

  Query Match           4.7%;  Score 14;  DB 20;  Length 250;
  Best Local Similarity 93.3%;  Pred. No. 1.14e+02;
  Matches   14;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db    128 tgcagcctcctgtgg 142
          |||||| ||||||||
Cp    265 TGCAGCNTCCTGTGG 251


RESULT   43
ID   T24718 standard; cDNA to mRNA; 284 BP.
AC   T24718;
DT   09-OCT-1996  (first entry)
DE   Human gene signature HUMGS06789.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 1678; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
```



```
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   284 BP;    80 A;    54 C;    76 G;    59 T;

  Query Match           4.7%;  Score 14;  DB 21;  Length 284;
  Best Local Similarity 93.3%;  Pred. No. 1.14e+02;
  Matches   14;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db    115 gggcaccaggtgtgc 129
          ||||||||| |||| |
Qy    162 GGGCACCAGGNGTGC 176


RESULT   44
ID   Q95197 standard; DNA; 301 BP.
AC   Q95197;
DT   08-FEB-1996  (first entry)
DE   Simple tandem repeat (STR) corresponding to wg1g9.
KW   Simple tandem repeat; STR; wg1g9; treatment; genetic; diagnosis;
KW   characterisation; mapping; linkage studies; analysis; alleles; ss.
OS   Synthetic.
PN   WO9517522-A2.
PD   29-JUN-1995.
PF   21-DEC-1994; G02789.
PR   21-DEC-1993; GB-026052.
PA   (UYLE-) UNIV LEICESTER.
PI   Armour J,  Jeffreys AJ;
DR   WPI; 95-240682/31.
PT   Identifying simple tandem repeat loci in DNA - by screening DNA
PT   library to enrich for fragments contg. the repeats before cloning
PT   and rescreening, also simple tandem repeats for treatment or
PT   diagnosis
PS   Claim 26; Page 24; 51pp; English.
CC   Q95197 is a simple tandem repeat (STR) corresponding to wg1g9. The
CC   STR can be used for treatment and diagnosis in human and veterinary
CC   medicine, partic. for genetic characterisation, mapping, linkage
CC   studies and analysis/diagnosis of acquired disease alleles.
SQ   Sequence   301 BP;    117 A;    55 C;    104 G;    25 T;

  Query Match           4.7%;  Score 14;  DB 16;  Length 301;
  Best Local Similarity 100.0%;  Pred. No. 1.14e+02;
  Matches   14;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      9 cactgtaccccagc 22
          ||||||||||||||
Qy    184 CACTGTACCCCAGC 197
```

```
RESULT   45
ID   Q59747 standard; cDNA; 302 BP.
AC   Q59747;
DT   16-MAR-1994  (first entry)
DE   Human brain Expressed Sequence Tag EST01523.
KW   Gene transcription product; genetic markers; tagging; in vivo;
KW   transcription; mapping; locations; chromosomes; chromosomal; ss.
OS   Homo sapiens.
PN   WO9316178-A.
PD   19-AUG-1993.
PF   12-FEB-1993; U01294.
PR   12-FEB-1992; US-837195.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
PI   Adams MD, Moreno RF, Venter CJ;
DR   WPI; 93-272882/34.
PT   Enriched oligonucleotides and corresp. sequences - used as
PT   markers for human genes transcribed in-vivo, facilitate tagging
PT   of most human genes
PS   Example 4; Page 225; 500pp; English.
CC   The Expressed Sequence Tag was isolated from a human brain cDNA
CC   library as part of a large set of ESTs which can be used as markers
CC   for human genes transcribed in vivo. They can be used to facilitate
CC   tagging of most human genes, for mapping locations of expressed genes
CC   on chromosomes, for individual or forensic identification, for mapping
CC   locations of disease-associated genes, for identification of tissue
CC   type, and for prepn. of antisense sequences, probes and constructs.
CC   EST01523 has a "poor" coding probability as evaluated using the
CC   coding-region prediction program CRM. See also Q59041-Q61440.
SQ   Sequence   302 BP;    113 A;    46 C;    58 G;    83 T;

  Query Match             4.7%;  Score 14;  DB 8;  Length 302;
  Best Local Similarity  93.3%;  Pred. No. 1.14e+02;
  Matches    14;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Db    229 tgaactgggccaggc 243
          ||| ||||||||||||
Qy     49 TGANCTGGGCCAGGC 63


RESULT   46
ID   Q39815 standard; DNA; 316 BP.
AC   Q39815;
DT   20-MAY-1993  (first entry)
DE   Expressed Sequence Tag human gene marker EST00154.
KW   expressed sequence tag; human genome project; chromosome;
KW   human gene sequencing; PCR mapping; somatic cell hybrids;
KW   sublocalisation; gene tagging; tissue typing.
OS   Synthetic.
PN   WO9300353-A.
```



```
PD   07-JAN-1993.
PF   19-JUN-1992; U05222.
PR   20-JUN-1991; US-716831.
PR   12-FEB-1992; US-837195.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
PI   Adams MD,  Venter JC;
DR   WPI; 93-036325/04.
PT   Particular expressed sequence tags from human CDNA - corresponds
PT   to transcription prods. of genes, useful for tagging genes,
PT   mapping chromosomes and tissue typing
PS   Claim 3; Page 149; 199pp; English.
CC   This sequence represents an EST (expressed sequence tag) ESTs are markers
CC   for human genes actually transcribed in vivo. Unlike the random genomic
CC   DNA sequence tagged sites (STSs), ESTs point directly to expressed genes.
CC   The use of ESTs could facilitate the tagging of most expressed human
CC   genes within a few years at a fraction of the cost of complete genomic
CC   sequencing. Using PCR primers Q39419-Q39580 (sequences designed
CC   from the ESTs) sublocalisation of an EST can be achieved with panels of
CC   fragments from specific chromosomes or pools of large genomic clones in
CC   an analogous manner. This sequence represents EST00154.
SQ   Sequence   316 BP;    92 A;    63 C;    90 G;    70 T;

  Query Match             4.7%;  Score 14;  DB 6;  Length 316;
  Best Local Similarity 100.0%;  Pred. No. 1.14e+02;
  Matches    14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Db      3 atttatacatttat 16
          ||||||||||||||
Cp    115 ATTTATACATTTAT 102


RESULT   47
ID   Q59227 standard; cDNA; 316 BP.
AC   Q59227;
DT   16-MAR-1994  (first entry)
DE   Human brain Expressed Sequence Tag EST00154.
KW   Gene transcription product; genetic markers; tagging; in vivo;
KW   transcription; mapping; locations; chromosomes; chromosomal; ss.
OS   Homo sapiens.
PN   WO9316178-A.
PD   19-AUG-1993.
PF   12-FEB-1993; U01294.
PR   12-FEB-1992; US-837195.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
PI   Adams MD, Moreno RF, Venter CJ;
DR   WPI; 93-272882/34.
PT   Enriched oligonucleotides and corresp. sequences - used as
PT   markers for human genes transcribed in-vivo, facilitate tagging
PT   of most human genes
PS   Example 1; Page 141; 500pp; English.
CC   The Expressed Sequence Tag was isolated from a human brain cDNA
```



```
CC   library as part of a large set of ESTs which can be used as markers
CC   for human genes transcribed in vivo. They can be used to facilitate
CC   tagging of most human genes, for mapping locations of expressed genes
CC   on chromosomes, for individual or forensic identification, for mapping
CC   locations of disease-associated genes, for identification of tissue
CC   type, and for prepn. of antisense sequences, probes and constructs.
CC   EST00154 has a "poor" coding probability as evaluated using the
CC   coding-region prediction program CRM.
CC   See also Q59041-Q61440.
SQ   Sequence   316 BP;    89 A;    63 C;    90 G;    70 T;

  Query Match             4.7%;  Score 14;  DB 8;  Length 316;
  Best Local Similarity 100.0%;  Pred. No. 1.14e+02;
  Matches    14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Db      3 atttatacatttat 16
          ||||||||||||||
Cp    115 ATTTATACATTTAT 102


RESULT   48
ID   Q60627 standard; cDNA; 343 BP.
AC   Q60627;
DT   16-MAR-1994  (first entry)
DE   Human brain Expressed Sequence Tag EST02640.
KW   Gene transcription product; genetic markers; tagging; in vivo;
KW   transcription; mapping; locations; chromosomes; chromosomal; ss.
OS   Homo sapiens.
PN   WO9316178-A.
PD   19-AUG-1993.
PF   12-FEB-1993; U01294.
PR   12-FEB-1992; US-837195.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
PI   Adams MD, Moreno RF, Venter CJ;
DR   WPI; 93-272882/34.
PT   Enriched oligonucleotides and corresp. sequences - used as
PT   markers for human genes transcribed in-vivo, facilitate tagging
PT   of most human genes
PS   Example 4; Page 364; 500pp; English.
CC   The Expressed Sequence Tag was isolated from a human brain cDNA
CC   library as part of a large set of ESTs which can be used as markers
CC   for human genes transcribed in vivo. They can be used to facilitate
CC   tagging of most human genes, for mapping locations of expressed genes
CC   on chromosomes, for individual or forensic identification, for mapping
CC   locations of disease-associated genes, for identification of tissue
CC   type, and for prepn. of antisense sequences, probes and constructs.
CC   EST02640 has a "marginal" coding probability as evaluated using the
CC   coding-region prediction program CRM. See also Q59041-Q61440.
SQ   Sequence   343 BP;    113 A;    67 C;    67 G;    94 T;

  Query Match             4.7%;  Score 14;  DB 8;  Length 343;
```



```
  Best Local Similarity 100.0%;  Pred. No. 1.14e+02;
  Matches    14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Db    206 aaatgtataaatgt 219
          ||||||||||||||
Qy    104 AAATGTATAAATGT 117


RESULT   49
ID   Q28186 standard; cDNA to mRNA; 345 BP.
AC   Q28186;
DT   11-FEB-1993  (first entry)
DE   Human T lymphocyte receptor V-region "IGR b 16".
KW   TCR; V beta 13 subfamily; variable region; anchored PCR;
KW   polymerase chain reaction; T cell receptor; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   intron          1..55
FT                   /*tag= a
FT                   /note= "within leader region"
FT   exon            56..345
FT                   /*tag= b
FT                   /note= "reading frame starts at nucleotide 58"
PN   WO9213950-A.
PD   20-AUG-1992.
PF   12-FEB-1992; F00130.
PR   12-FEB-1991; FR-001613.
PR   12-APR-1991; FR-004523.
PA   (ROUS ) ROUSSEL-UCLAF.
PI   Ferradini L, Hercend T, Roman-Roman S, Triebel F;
DR   WPI; 92-300036/36.
DR   P-PSDB; R26973.
PT   Variable regions of b-chain of T-lymphocyte receptors and their
PT   DNA - useful as immuno:modulant(s) and for diagnosing immune
PT   disorders
PS   Claim 1; Page 50; 75pp; French.
CC   RNA was isolated from peripheral lymphocytes and converted to cDNA
CC   using a C-beta-specific primer. The cDNA was amplified by anchored
CC   PCR using C-beta and polyC primers, then amplified again using a
CC   different C-beta specific primer. The amplified product was SacII-
CC   restricted, inserted into Bluescript SK+ vector and used to transform
CC   E.coli XL-1blue. Transformants were screened with a C-beta specific
CC   probe and DNA from positive clones was sequenced in the C-beta
CC   region. The sequence designated "IGR b 16" is a consensus sequence;
CC   a C has been observed in stead of a T in position 231 and an A has
CC   been observed at position 259, in stead of a G. IGR b16 has a
CC   homology of 94% with clone HBVP34, a previously identified member
CC   of the V beta 13 subfamily.
CC   See Q28173-Q28228.
SQ   Sequence   345 BP;    83 A;    96 C;    83 G;    83 T;
```

```
  Query Match           4.7%;  Score 14;  DB 4;  Length 345;
  Best Local Similarity  87.5%;  Pred. No. 1.14e+02;
  Matches   14;  Conservative    0;  Mismatches   2;  Indels   0;  Gaps   0;

Db    263 ccagatcaaccacaga 278
          |||| ||||||| |||
Cp     56 CCAGNTCAACCANAGA 41


RESULT   50
ID   Q89257 standard; DNA; 396 BP.
AC   Q89257;
DT   17-OCT-1995  (first entry)
DE   BMP-5 encoding DNA.
KW   BMP-5; bone morphogenetic protein; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             1..396
FT                   /*tag= a
PN   US5399677-A.
PD   21-MAR-1995.
PF   07-DEC-1993; 163877.
PR   07-DEC-1993; US-163877.
PA   (GEMY ) GENETICS INST INC.
PI   McCoy J,  Wolfman NM;
DR   WPI; 95-130739/17.
DR   P-PSDB; R71972.
PT   Mutant bone morphogenetic proteins (BMPs) and corresp. DNA -
PT   useful for prodn. of correctly folded, biologically active BMPs
PT   in bacteria
PS   Disclosure; Column 19-22; 16pp; English.
CC   The DNA sequences given in Q89255-59 encode the bone morphogenetic
CC   proteins BMP-2, BMP-4, BMP-5, BMP-6 and BMP-7 (R71970-74),
CC   respectively.  The DNA is used to prepare mutant forms of
CC   recombinant BMPs, as well as BMP homodimers or heterodimers, that
CC   are expressed in bacteria, pref. Escherichia coli G1724 (ATCC
CC   55155).
SQ   Sequence   396 BP;    118 A;    92 C;    79 G;    107 T;

  Query Match           4.7%;  Score 14;  DB 14;  Length 396;
  Best Local Similarity 100.0%;  Pred. No. 1.14e+02;
  Matches   14;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     41 ccagaatgtccagt 54
          ||||||||||||||
Cp    155 CCAGAATGTCCAGT 142


Search completed: Tue Dec 22 22:37:40 1998
Job time : 168 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Tue Dec 22 19:36:14 1998; MasPar time 424.18 Seconds
1249.645 Million cell updates/sec

Tabular output not generated.

Title: >US-08-530-112A-563
Description: (1-296) from US08530112A.seq
Perfect Score: 296
N.A. Sequence: 1 GATCTGCCATAAGAAAATCT.....TTTTGGTTGTTAAACTGAAA 296
Comp: CTAGACGGTATTCTTTTAGA.....AAAACCAACAATTTGACTTT

Scoring table: TABLE jmetric
Gap 60

Nmatch STD : Dbase 0; Query 0

Searched: 2275026 seqs, 895388244 bases x 2

Post-processing: Minimum Match 0%
Listing first 1000 summaries

Database: embl-est55
1:em_est1 2:em_gss1 3:em_gss2 4:em_gss3
Database: genbank-est107
5:gb_est1 6:gb_est10 7:gb_est11 8:gb_est12 9:gb_est13
10:gb_est14 11:gb_est15 12:gb_est16 13:gb_est17
14:gb_est18 15:gb_est19 16:gb_est2 17:gb_est20
18:gb_est21 19:gb_est3 20:gb_est4 21:gb_est5 22:gb_est6
23:gb_est7 24:gb_est8 25:gb_est9 26:gb_gss1 27:gb_gss2
28:gb_gss3 29:gb_gss4

Statistics: Mean 8.481; Variance 1.540; scale 5.506

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
	1	296	100.0	296	7	HUMGS00628	Human HL60 3'directed	0.00e+00
c	2	296	100.0	413	15	AA906845	ol20a09.s1 Soares_NFL_	0.00e+00
c	3	296	100.0	447	14	AA836401	od37d12.s1 NCI_CGAP_GC	0.00e+00
c	4	290	98.0	342	12	AA748329	oa54e10.s1 NCI_CGAP_GC	0.00e+00
c	5	216	73.0	403	12	AA808708	oa47d05.s1 NCI_CGAP_GC	0.00e+00
	6	20	6.8	116	5	T81784	yd30d06.s1 Homo sapien	1.36e-05
	7	20	6.8	163	8	AA547753	EST188662 B. glabrata	1.36e-05
c	8	20	6.8	589	23	AA186021	mt35d04.r1 Soares mous	1.36e-05
	9	20	6.8	625	26	B53879	CIT-HSP-2013H3.TF CIT-	1.36e-05
	10	19	6.4	181	6	W80705	zh51a06.s1 Soares feta	3.32e-04
	11	19	6.4	254	11	AA678036	z125d03.s1 Soares feta	3.32e-04
	12	19	6.4	273	20	R81084	EST22b023 WATM1 Homo s	3.32e-04
	13	19	6.4	289	11	AA064804	zm50a08.s1 Stratagene	3.32e-04
	14	19	6.4	293	16	R69529	yj82c11.s1 Homo sapien	3.32e-04
	15	19	6.4	294	8	AA262768	zs24d12.s1 NCI_CGAP_GC	3.32e-04
	16	19	6.4	307	13	AA845909	ak86b11.s1 Barstead sp	3.32e-04
	17	19	6.4	318	14	AA887742	ng97f10.s1 NCI_CGAP_Co	3.32e-04
	18	19	6.4	321	27	AQ042161	CIT-HSP-2330A19.TR CIT	3.32e-04
	19	19	6.4	334	13	AA634259	ac75g01.s1 Stratagene	3.32e-04
	20	19	6.4	334	5	T09855	0428m7 Plasmodium falc	3.32e-04
c	21	19	6.4	339	11	AA083747	zm62b12.r1 Stratagene	3.32e-04
c	22	19	6.4	344	14	AA960086	vw59f12.s1 Soares mous	3.32e-04
	23	19	6.4	350	11	AA112884	zm62b12.s1 Stratagene	3.32e-04
	24	19	6.4	353	14	AA991247	os40h07.s1 NCI_CGAP_Br	3.32e-04
	25	19	6.4	359	19	H67871	yu67g08.s1 Homo sapien	3.32e-04
	26	19	6.4	365	28	AQ121777	HS_3079_B2_C06_MR CIT	3.32e-04
	27	19	6.4	377	6	AA399557	zt88d05.s1 Soares test	3.32e-04
	28	19	6.4	388	18	AI150715	qb77d10.x1 Soares_feta	3.32e-04
	29	19	6.4	390	17	AI070901	UI-R-C2-mt-b-06-0-UI.s	3.32e-04
	30	19	6.4	395	7	C25635	Dictyostelium discoide	3.32e-04
c	31	19	6.4	415	13	AA171915	zo95d07.r1 Stratagene	3.32e-04
	32	19	6.4	417	20	W42584	zc23a05.s1 Soares sene	3.32e-04
	33	19	6.4	418	10	AA649293	ns31g01.s1 NCI_CGAP_GC	3.32e-04
	34	19	6.4	421	15	AI058966	UI-R-C1-lr-h-08-0-UI.s	3.32e-04
	35	19	6.4	445	5	R10366	yf37a04.s1 Homo sapien	3.32e-04
	36	19	6.4	449	18	AI150662	qc06c11.x1 Soares_feta	3.32e-04
	37	19	6.4	450	16	H22455	yn69e02.s1 Homo sapien	3.32e-04
c	38	19	6.4	452	23	AA154911	ms02d03.r1 Stratagene	3.32e-04
c	39	19	6.4	500	16	R59260	yg97e02.r1 Homo sapien	3.32e-04
c	40	19	6.4	504	29	AQ253742	HS_3059_B2_A04_MF CIT	3.32e-04
	41	19	6.4	509	18	AI144725	UI-R-BT0-po-c-01-0-UI.	3.32e-04
	42	19	6.4	520	14	AA876015	ob93d12.s1 NCI_CGAP_GC	3.32e-04

	43	19	6.4	625	6	AA398463	zt62e11.s1 Soares test	3.32e-04
c	44	19	6.4	687	27	AG012998	Homo sapiens genomic D	3.32e-04
c	45	19	6.4	687	26	AG005353	Homo sapiens genomic D	3.32e-04
	46	19	6.4	706	26	AG005363	Homo sapiens genomic D	3.32e-04
	47	19	6.4	706	27	AG013008	Homo sapiens genomic D	3.32e-04
c	48	19	6.4	931	26	B08181	F15B10-Sp6 IGF Arabido	3.32e-04
	49	18	6.1	152	22	AA102128	zl80a03.s1 Stratagene	7.21e-03
c	50	18	6.1	173	13	AA283357	TH222 HTCDL1 Homo sapi	7.21e-03
c	51	18	6.1	180	13	AA283359	TH271 HTCDL1 Homo sapi	7.21e-03
c	52	18	6.1	202	7	HUM254C08B	Human aorta cDNA 5'-en	7.21e-03
	53	18	6.1	233	17	AI087567	SWAMCAC20B10SK Brugia	7.21e-03
	54	18	6.1	251	17	AI096851	qb40b05.x1 NCI_CGAP_Br	7.21e-03
c	55	18	6.1	251	21	AA055081	zf19d02.s1 Soares feta	7.21e-03
	56	18	6.1	259	10	AA617994	np88g10.s1 NCI_CGAP_Th	7.21e-03
c	57	18	6.1	273	7	HUM322B08B	Human aorta cDNA 5'-en	7.21e-03
c	58	18	6.1	290	13	AA283364	TH557 HTCDL1 Homo sapi	7.21e-03
c	59	18	6.1	293	23	T17247	NIB580 Normalized infa	7.21e-03
c	60	18	6.1	295	7	HUM349D03B	Human aorta cDNA 5'-en	7.21e-03
	61	18	6.1	300	6	C08231	C.elegans cDNA clone y	7.21e-03
	62	18	6.1	304	14	AA912075	oi57c10.s1 NCI_CGAP_HN	7.21e-03
c	63	18	6.1	319	12	AA714118	nw05d03.s1 NCI_CGAP_SS	7.21e-03
c	64	18	6.1	320	14	AA888556	nw81d01.s1 NCI_CGAP_Pr	7.21e-03
c	65	18	6.1	320	7	HUM213D03B	Human aorta cDNA 5'-en	7.21e-03
c	66	18	6.1	326	5	R15832	ya46f07.s2 Homo sapien	7.21e-03
c	67	18	6.1	330	7	HUM258B02B	Human aorta cDNA 5'-en	7.21e-03
c	68	18	6.1	339	6	AA044887	zk72f12.r1 Soares preg	7.21e-03
c	69	18	6.1	344	7	HUM340F01B	Human aorta cDNA 5'-en	7.21e-03
	70	18	6.1	350	13	AA846663	aj98e05.s1 Soares para	7.21e-03
c	71	18	6.1	353	16	T35708	EST89898 Homo sapiens	7.21e-03
	72	18	6.1	359	16	H43309	yp09a05.s1 Homo sapien	7.21e-03
	73	18	6.1	364	5	R35652	yh90c07.s1 Homo sapien	7.21e-03
c	74	18	6.1	364	7	HUM254A05B	Human aorta cDNA 5'-en	7.21e-03
c	75	18	6.1	367	27	B66458	CIT-HSP-2011P21.TF3 CI	7.21e-03
c	76	18	6.1	368	15	AA972425	op78c03.s1 Soares_NFL_	7.21e-03
	77	18	6.1	370	16	H14631	yl26e03.s1 Homo sapien	7.21e-03
c	78	18	6.1	378	27	AQ055822	CIT-HSP-2341F23.TR CIT	7.21e-03
	79	18	6.1	382	10	AA664009	ac03f06.s1 Stratagene	7.21e-03
	80	18	6.1	386	5	R41393	yf94a09.s1 Homo sapien	7.21e-03
c	81	18	6.1	390	22	W99329	ze03c04.s1 Soares feta	7.21e-03
c	82	18	6.1	395	24	AA307138	EST178263 Colon carcin	7.21e-03
c	83	18	6.1	401	11	AA716590	zg66d08.s1 Soares feta	7.21e-03
c	84	18	6.1	406	28	AQ190798	HS_3220_A1_G06_MR CIT	7.21e-03
	85	18	6.1	407	28	AQ178045	HS_2216_B2_E12_MF CIT	7.21e-03
	86	18	6.1	422	5	R43863	yg26a06.s1 Homo sapien	7.21e-03
c	87	18	6.1	422	14	AA956558	UI-R-E1-fk-a-02-0-UI.s	7.21e-03
c	88	18	6.1	422	28	AQ215613	HS_2257_B1_E11_T7 CIT	7.21e-03
c	89	18	6.1	422	28	AQ132469	HS_3003_A2_F01_MR CIT	7.21e-03
	90	18	6.1	423	19	N22960	yx73b11.s1 Homo sapien	7.21e-03
	91	18	6.1	427	15	AI010543	EST204994 Normalized r	7.21e-03
	92	18	6.1	433	18	AI082430	ox73b06.x1 Soares_NhHM	7.21e-03
	93	18	6.1	455	28	AQ155037	HS_3039_B1_D03_MF CIT	7.21e-03

c	94	18	6.1	455	27	AQ003870	CIT-HSP-2289H12.TF CIT	7.21e-03
	95	18	6.1	460	11	AA704925	zj82c12.s1 Soares feta	7.21e-03
	96	18	6.1	461	18	AI052484	os19f10.x1 Soares_feta	7.21e-03
c	97	18	6.1	465	16	H43347	yp09a05.r1 Homo sapien	7.21e-03
	98	18	6.1	466	12	AA781279	zi49g03.s1 Soares para	7.21e-03
	99	18	6.1	470	13	AA857869	oh99c02.s1 NCI_CGAP_HN	7.21e-03
	100	18	6.1	475	8	AA515632	ng70c09.s1 NCI_CGAP_Li	7.21e-03
	101	18	6.1	479	14	AA857652	of65e07.s1 NCI_CGAP_Co	7.21e-03
c	102	18	6.1	480	28	AQ223574	HS_2011_B1_C11_T7 CIT	7.21e-03
	103	18	6.1	481	18	AI077620	oy26e08.s1 Soares_sene	7.21e-03
	104	18	6.1	486	20	N90729	za90d12.s1 Soares feta	7.21e-03
	105	18	6.1	486	18	AI185090	qe46h12.x1 Soares_feta	7.21e-03
	106	18	6.1	493	27	AQ078379	CIT-HSP-2361K10.TF CIT	7.21e-03
c	107	18	6.1	495	13	AA224018	zr11e01.r1 Stratagene	7.21e-03
c	108	18	6.1	496	14	C92357	Dictyostelium discoide	7.21e-03
	109	18	6.1	497	18	AI094759	qa13b03.x1 NCI_CGAP_Br	7.21e-03
	110	18	6.1	498	17	AI001157	ot40b04.s1 Soares_test	7.21e-03
	111	18	6.1	502	13	AA164663	zo98h10.s1 Stratagene	7.21e-03
c	112	18	6.1	504	5	R41639	yf86e10.s1 Homo sapien	7.21e-03
c	113	18	6.1	512	10	AA618697	SWOvL3CA731SK Onchocer	7.21e-03
	114	18	6.1	513	8	AA279302	zs84c05.s1 NCI_CGAP_GC	7.21e-03
c	115	18	6.1	519	11	AA669264	ac17g12.s1 Stratagene	7.21e-03
	116	18	6.1	527	12	AA781559	zi56a05.s1 Soares para	7.21e-03
	117	18	6.1	549	12	AA771727	zi32d04.s1 Soares para	7.21e-03
	118	18	6.1	549	19	N39029	yv22d07.s1 Homo sapien	7.21e-03
	119	18	6.1	563	12	AA771708	zi32b04.s1 Soares para	7.21e-03
c	120	18	6.1	574	20	N78410	yz75h02.r1 Homo sapien	7.21e-03
	121	18	6.1	589	10	AA662425	nu97g08.s1 NCI_CGAP_Pr	7.21e-03
	122	18	6.1	628	13	AA872200	ob78d12.s1 NCI_CGAP_GC	7.21e-03
c	123	18	6.1	660	20	W22082	61D11 Human retina cDN	7.21e-03
c	124	18	6.1	670	27	AQ083860	CpG0669A CpIOWAgDNA1 C	7.21e-03
	125	18	6.1	675	26	AG007228	Homo sapiens genomic D	7.21e-03
	126	18	6.1	677	26	AG007227	Homo sapiens genomic D	7.21e-03
	127	18	6.1	840	11	AA659242	nu12h08.s1 NCI_CGAP_Pr	7.21e-03
	128	17	5.7	91	17	AD013220	Schizosaccharomyces po	1.37e-01
	129	17	5.7	98	5	T48108	yb25a03.s1 Homo sapien	1.37e-01
	130	17	5.7	123	22	AA205169	zq78c02.s1 Stratagene	1.37e-01
c	131	17	5.7	183	13	H31979	EST106577 Rat PC-12 ce	1.37e-01
	132	17	5.7	201	12	AA788983	ah29c03.s1 Soares para	1.37e-01
	133	17	5.7	201	7	HUM074D04B	Human fetal brain cDNA	1.37e-01
	134	17	5.7	214	13	AA873092	ob36b12.s1 NCI_CGAP_Ki	1.37e-01
	135	17	5.7	223	20	HSC0NE052	H. sapiens partial cDN	1.37e-01
	136	17	5.7	230	11	AA719553	zh36e08.s1 Soares pine	1.37e-01
	137	17	5.7	240	17	AI051484	ox43a03.s1 Soares_tota	1.37e-01
c	138	17	5.7	247	6	AA035487	zk27c05.s1 Soares preg	1.37e-01
	139	17	5.7	255	19	HSC2UG032	H. sapiens partial cDN	1.37e-01
	140	17	5.7	274	18	AI175523	EST219077 Normalized r	1.37e-01
	141	17	5.7	283	12	AA729859	nx39h12.s1 NCI_CGAP_GC	1.37e-01
	142	17	5.7	287	8	AA480876	aa28c02.s1 NCI_CGAP_GC	1.37e-01
	143	17	5.7	289	12	AA814359	nz07h01.s1 NCI_CGAP_GC	1.37e-01
	144	17	5.7	295	5	R44153	yg33c12.s1 Homo sapien	1.37e-01

	145	17	5.7	298	27	AQ070296	HS_3040_A2_E07_MR CIT 1.37e-01
	146	17	5.7	300	15	AA916266	ol63b02.s1 NCI_CGAP_Ki 1.37e-01
c	147	17	5.7	300	9	C32876	C.elegans cDNA clone y 1.37e-01
c	148	17	5.7	300	9	C32343	C.elegans cDNA clone y 1.37e-01
c	149	17	5.7	300	9	C35851	C.elegans cDNA clone y 1.37e-01
c	150	17	5.7	300	9	C36704	C.elegans cDNA clone y 1.37e-01
c	151	17	5.7	300	9	C35971	C.elegans cDNA clone y 1.37e-01
	152	17	5.7	315	18	AI179331	EST223027 Normalized r 1.37e-01
	153	17	5.7	321	19	N29197	yx42g03.r1 Homo sapien 1.37e-01
	154	17	5.7	329	22	AA014206	mh18e09.r1 Soares mous 1.37e-01
c	155	17	5.7	329	22	AA014206	mh18e09.r1 Soares mous 1.37e-01
	156	17	5.7	337	17	AI018116	ot91f03.s1 Soares_tota 1.37e-01
	157	17	5.7	338	10	AA645761	vn10f02.r1 Stratagene 1.37e-01
	158	17	5.7	339	10	AA665706	ag53h03.s1 Gessler Wil 1.37e-01
	159	17	5.7	340	24	AA329706	EST33376 Embryo, 12 we 1.37e-01
c	160	17	5.7	346	18	AI159144	vz83g11.r1 Soares mous 1.37e-01
	161	17	5.7	348	27	AQ084065	HS_2230_B2_H09_MF CIT 1.37e-01
c	162	17	5.7	348	6	AA397314	mr38f11.r1 Life Tech m 1.37e-01
c	163	17	5.7	358	11	AA688784	vr07f03.r1 Knowles Sol 1.37e-01
c	164	17	5.7	358	19	H70691	yu19g04.r1 Homo sapien 1.37e-01
c	165	17	5.7	358	16	H44204	yn58a12.s1 Homo sapien 1.37e-01
	166	17	5.7	360	6	C09823	C.elegans cDNA clone y 1.37e-01
c	167	17	5.7	361	26	CI1B9	Ciona intestinalis gen 1.37e-01
	168	17	5.7	367	27	B95857	CIT-HSP-2170B2.TR CIT- 1.37e-01
c	169	17	5.7	368	18	AI145513	UI-R-BT0-py-b-03-0-UI. 1.37e-01
	170	17	5.7	373	12	AA738953	vv62e04.r1 Soares 2NbM 1.37e-01
c	171	17	5.7	373	12	AA738953	vv62e04.r1 Soares 2NbM 1.37e-01
	172	17	5.7	379	28	AQ102252	HS_3029_A2_D07_MF CIT 1.37e-01
	173	17	5.7	379	18	AI161070	qc77g07.x1 Soares_plac 1.37e-01
	174	17	5.7	379	14	AA835953	oc79g02.s1 NCI_CGAP_GC 1.37e-01
	175	17	5.7	379	15	AA973293	on94c01.s1 Soares_NFL_ 1.37e-01
	176	17	5.7	379	27	AQ080052	CIT-HSP-2356K14.TR CIT 1.37e-01
	177	17	5.7	381	28	AQ191714	HS_3226_B1_H11_MR CIT 1.37e-01
	178	17	5.7	385	15	AU007984	Schizosaccharomyces po 1.37e-01
	179	17	5.7	386	28	AQ101917	HS_3017_B1_D01_T7 CIT 1.37e-01
c	180	17	5.7	387	10	AA293922	SW0v3MCA1482SK Onchoce 1.37e-01
	181	17	5.7	388	16	R68379	yi05c11.s1 Homo sapien 1.37e-01
c	182	17	5.7	388	27	B93625	CIT-HSP-264N19.TF CIT 1.37e-01
	183	17	5.7	389	13	AA826953	ob52f04.s1 NCI_CGAP_GC 1.37e-01
	184	17	5.7	395	15	AI030625	UI-R-C0-jh-f-10-0-UI.s 1.37e-01
	185	17	5.7	395	15	AA894779	oj55d01.s1 NCI_CGAP_Ki 1.37e-01
	186	17	5.7	395	12	AA766622	oa38a08.s1 NCI_CGAP_GC 1.37e-01
	187	17	5.7	401	14	AA907132	od44b09.s1 NCI_CGAP_GC 1.37e-01
c	188	17	5.7	401	18	AI182625	uc28b08.r1 Soares mous 1.37e-01
	189	17	5.7	402	13	AA598460	ae48g12.s1 Stratagene 1.37e-01
	190	17	5.7	405	21	W67571	zd41e01.s1 Soares feta 1.37e-01
	191	17	5.7	407	20	N89957	xb36b06.s1 Soares para 1.37e-01
c	192	17	5.7	407	23	AA184373	mt34h03.r1 Soares mous 1.37e-01
c	193	17	5.7	411	23	AA124373	mq76h11.r1 Stratagene 1.37e-01
	194	17	5.7	412	15	AI049852	an21f02.x1 Gessler Wil 1.37e-01
	195	17	5.7	413	6	AA035344	zk24d11.s1 Soares preg 1.37e-01

c	196	17	5.7	415	12	AA793642	vp51a04.s1 Knowles Sol 1.37e-01
	197	17	5.7	421	6	AA388951	vb25e10.r1 Soares mous 1.37e-01
	198	17	5.7	422	14	AA897258	am07c06.s1 Soares_NFL_ 1.37e-01
	199	17	5.7	424	11	AA710950	vt93f12.r1 Soares mous 1.37e-01
	200	17	5.7	425	6	AA147616	zl52c06.s1 Soares preg 1.37e-01
	201	17	5.7	425	17	AI039229	ox32d12.s1 Soares_tota 1.37e-01
c	202	17	5.7	426	7	AA462741	vf91b04.r1 Soares mous 1.37e-01
	203	17	5.7	432	17	AA722332	UI-R-C2-ng-d-06-0-UI.s 1.37e-01
c	204	17	5.7	434	19	H93708	yv16c12.s1 Homo sapien 1.37e-01
	205	17	5.7	439	14	AA910993	ok57e06.s1 NCI_CGAP_Le 1.37e-01
	206	17	5.7	443	21	W56775	zd14h07.s1 Soares feta 1.37e-01
c	207	17	5.7	443	27	AQ041796	CIT-HSP-2336L9.TF CIT- 1.37e-01
c	208	17	5.7	444	26	B72511	RPCI11-9D15.TV RPCI11 1.37e-01
c	209	17	5.7	445	29	AQ244552	HS_2053_B1_C11_T7 CIT 1.37e-01
	210	17	5.7	446	16	R49637	yg61g07.s1 Homo sapien 1.37e-01
c	211	17	5.7	446	11	AA070257	zm69b06.r1 Stratagene 1.37e-01
c	212	17	5.7	447	14	AA942537	LD26810.5prime LD Dros 1.37e-01
	213	17	5.7	450	19	H98639	yx12e04.s1 Homo sapien 1.37e-01
	214	17	5.7	451	18	AI052615	ox10c01.x1 Soares_feta 1.37e-01
c	215	17	5.7	452	12	AA793641	vp51a03.s1 Knowles Sol 1.37e-01
c	216	17	5.7	452	19	N21045	yx47b09.s1 Homo sapien 1.37e-01
	217	17	5.7	454	6	AA127392	zn81b03.s1 Stratagene 1.37e-01
c	218	17	5.7	454	7	AA499605	vg04g09.r1 Soares mous 1.37e-01
	219	17	5.7	455	17	AA409232	EST03724 Mouse 7.5 dpc 1.37e-01
c	220	17	5.7	456	23	AA163655	mr19e09.r1 Soares mous 1.37e-01
c	221	17	5.7	460	23	AA170452	ms67a11.r1 Soares mous 1.37e-01
	222	17	5.7	467	6	AA029295	zk10c02.s1 Soares preg 1.37e-01
	223	17	5.7	467	22	N62099	ys61d01.s1 Soares mult 1.37e-01
c	224	17	5.7	467	10	AA427043	ve78e08.r1 Soares mous 1.37e-01
c	225	17	5.7	467	10	AA427082	ve78g08.r1 Soares mous 1.37e-01
	226	17	5.7	468	11	AA704343	zf85c02.s1 Soares pine 1.37e-01
c	227	17	5.7	468	28	AQ222542	HS_3252_B1_F03_T7 CIT 1.37e-01
	228	17	5.7	472	28	AQ126388	HS_3037_B1_A02_MF CIT 1.37e-01
	229	17	5.7	473	17	AI024581	ov68b09.x1 Soares_test 1.37e-01
	230	17	5.7	473	13	AA827075	ob56c05.s1 NCI_CGAP_GC 1.37e-01
c	231	17	5.7	478	23	AA184711	mt94a04.r1 Soares mous 1.37e-01
c	232	17	5.7	480	15	AU010618	Schizosaccharomyces po 1.37e-01
	233	17	5.7	481	27	AQ037624	CIT-HSP-2333A18.TR CIT 1.37e-01
c	234	17	5.7	484	28	AQ154532	HS_3014_B1_H06_T7 CIT 1.37e-01
c	235	17	5.7	484	17	AU013824	Schizosaccharomyces po 1.37e-01
c	236	17	5.7	486	8	AA542446	fa07a06.s1 Zebrafish I 1.37e-01
	237	17	5.7	489	13	AA223282	zr08c01.s1 Stratagene 1.37e-01
	238	17	5.7	490	20	W90213	zh77h03.r1 Soares feta 1.37e-01
	239	17	5.7	490	12	AA779756	af44d12.s1 Soares tota 1.37e-01
	240	17	5.7	490	28	AQ192276	HS_2242_B2_G03_MR CIT 1.37e-01
c	241	17	5.7	496	23	AA177881	mt07f05.r1 Soares mous 1.37e-01
	242	17	5.7	497	14	C92517	Dictyostelium discoide 1.37e-01
c	243	17	5.7	498	14	AA851151	EST193919 Normalized r 1.37e-01
	244	17	5.7	503	26	B72904	RPCI11-10J23.TP RPCI11 1.37e-01
	245	17	5.7	505	16	H05262	yl90e06.s1 Homo sapien 1.37e-01
	246	17	5.7	505	18	AI171539	EST217504 Normalized r 1.37e-01

c	247	17	5.7	506	24	AA266656	mx90e03.r1 Soares mous 1.37e-01
c	248	17	5.7	507	14	AA915364	vz29b10.r1 Soares 2NbM 1.37e-01
	249	17	5.7	508	14	AA900460	UI-R-E0-bw-b-07-0-UI.s 1.37e-01
	250	17	5.7	508	22	AA187641	zp73f08.s1 Stratagene 1.37e-01
	251	17	5.7	510	5	W68964	yg57f08.s1 Homo sapien 1.37e-01
	252	17	5.7	510	14	AA925918	UI-R-A1-es-d-04-0-UI.s 1.37e-01
c	253	17	5.7	512	23	AA125566	mq77h11.r1 Stratagene 1.37e-01
	254	17	5.7	513	18	AI169093	EST214924 Normalized r 1.37e-01
c	255	17	5.7	515	22	N53845	yv71f12.s1 Soares feta 1.37e-01
c	256	17	5.7	516	28	AQ164673	HS_3006_B2_B04_T7 CIT 1.37e-01
c	257	17	5.7	516	27	AQ065160	HS_2181_B2_H12_MR CIT 1.37e-01
c	258	17	5.7	517	11	AA702237	zl84f08.s1 Soares feta 1.37e-01
	259	17	5.7	519	29	AQ248272	HS_2063_A1_D11_MR CIT 1.37e-01
c	260	17	5.7	519	13	AA833140	ud02c10.r1 Soares mous 1.37e-01
	261	17	5.7	521	12	AA819766	UI-R-A0-av-h-07-0-UI.s 1.37e-01
c	262	17	5.7	524	17	AI121518	uc35e10.r1 Soares mous 1.37e-01
	263	17	5.7	528	19	N30183	yw82e11.s1 Homo sapien 1.37e-01
	264	17	5.7	528	28	AQ190679	HS_3220_A1_G06_T7 CIT 1.37e-01
c	265	17	5.7	533	19	N42098	yy05c07.r1 Homo sapien 1.37e-01
c	266	17	5.7	533	10	AA611552	vo85c04.r1 Barstead mo 1.37e-01
	267	17	5.7	534	18	AI101231	EST210520 Normalized r 1.37e-01
	268	17	5.7	537	20	W07290	za94h10.r1 Soares feta 1.37e-01
c	269	17	5.7	538	23	AA230617	mw03f11.r1 Soares mous 1.37e-01
c	270	17	5.7	548	17	AI115692	uc34c02.r1 Soares mous 1.37e-01
	271	17	5.7	551	26	FR0014184	F.rubripes GSS sequenc 1.37e-01
	272	17	5.7	551	20	W19540	zb32b10.r1 Soares para 1.37e-01
c	273	17	5.7	555	29	AQ237281	RPCI11-66J19.TJ RPCI11 1.37e-01
	274	17	5.7	562	26	B47833	RPCI11-2M5.TP RPCI11 H 1.37e-01
	275	17	5.7	563	9	AA521141	aa73a12.s1 NCI_CGAP_GC 1.37e-01
c	276	17	5.7	568	12	AA441358	LD16261.5prime LD Dros 1.37e-01
	277	17	5.7	572	18	AI102053	EST211342 Normalized r 1.37e-01
c	278	17	5.7	572	18	AI102053	EST211342 Normalized r 1.37e-01
c	279	17	5.7	578	22	AA124313	mn46f06.r1 Beddington 1.37e-01
	280	17	5.7	584	19	N64730	za22f08.s1 Homo sapien 1.37e-01
c	281	17	5.7	584	26	AQ019294	CIT-HSP-2305H11.TF CIT 1.37e-01
	282	17	5.7	587	18	AI102750	EST212039 Normalized r 1.37e-01
	283	17	5.7	597	19	N64689	yz87h07.s1 Homo sapien 1.37e-01
c	284	17	5.7	599	23	W36242	253 Mouse VM cDNA libr 1.37e-01
	285	17	5.7	604	18	AI169293	EST215128 Normalized r 1.37e-01
c	286	17	5.7	607	12	AA540682	LD20452.5prime LD Dros 1.37e-01
	287	17	5.7	613	16	R60048	yh12f01.s1 Homo sapien 1.37e-01
c	288	17	5.7	619	29	AQ233862	HS_2057_A1_E02_T7 CIT 1.37e-01
c	289	17	5.7	619	26	FR0004624	F.rubripes GSS sequenc 1.37e-01
	290	17	5.7	637	26	B16541	342J1.TPB CIT978SKA1 H 1.37e-01
c	291	17	5.7	655	26	B02675	cSRL-158E7-u cSRL flow 1.37e-01
	292	17	5.7	659	26	B21631	F1D9-T7 IGF Arabidopsi 1.37e-01
c	293	17	5.7	690	28	AG013236	Homo sapiens genomic D 1.37e-01
c	294	17	5.7	690	26	AG005591	Homo sapiens genomic D 1.37e-01
c	295	17	5.7	699	26	AG005589	Homo sapiens genomic D 1.37e-01
c	296	17	5.7	699	28	AG013234	Homo sapiens genomic D 1.37e-01
	297	17	5.7	709	28	AG013216	Homo sapiens genomic D 1.37e-01

	298	17	5.7	709	26	AG005571	Homo sapiens genomic D 1.37e-01
	299	17	5.7	711	14	C92431	Dictyostelium discoide 1.37e-01
	300	17	5.7	714	8	AA550495	1650m3 gmbPfHB3.1, G. 1.37e-01
	301	17	5.7	716	26	AG005570	Homo sapiens genomic D 1.37e-01
	302	17	5.7	716	28	AG013215	Homo sapiens genomic D 1.37e-01
c	303	17	5.7	746	14	AA990855	LD23130.5prime LD Dros 1.37e-01
	304	17	5.7	751	28	AQ110849	CIT-HSP-2378K4.TF CIT- 1.37e-01
	305	17	5.7	796	27	AQ050694	nbxb0004cB05f CUGI Ric 1.37e-01
	306	17	5.7	831	27	AQ072948	EP(2)2162 Drosophila m 1.37e-01
c	307	17	5.7	849	23	AA170154	ms88a10.r1 Soares mous 1.37e-01
c	308	17	5.7	948	26	B08123	F4C6-T7 IGF Arabidopsi 1.37e-01
	309	17	5.7	1045	26	B13010	T27N22-Sp6 TAMU Arabid 1.37e-01
c	310	17	5.7	1250	26	B08983	F10B13-Sp6 IGF Arabido 1.37e-01
c	311	16	5.4	70	7	AA509208	MBAPCX8F01T3 Brugia ma 2.23e+00
c	312	16	5.4	115	10	AA662307	nu97c09.s1 NCI_CGAP_Pr 2.23e+00
c	313	16	5.4	126	11	AA721215	ne68d04.s1 NCI_CGAP_GC 2.23e+00
	314	16	5.4	134	21	R85824	yq22c09.s1 Soares reti 2.23e+00
c	315	16	5.4	134	10	AA417867	zv01e07.r1 NCI_CGAP_GC 2.23e+00
	316	16	5.4	136	26	AQ012512	CIT-HSP-2298M16.TR CIT 2.23e+00
c	317	16	5.4	166	6	AA383666	EST97102 Testis I Homo 2.23e+00
	318	16	5.4	174	10	AA593177	nm97a08.s1 NCI_CGAP_Co 2.23e+00
c	319	16	5.4	183	5	R19879	yg38c11.r1 Homo sapien 2.23e+00
c	320	16	5.4	189	7	AA455520	zx77g10.s1 Soares ovar 2.23e+00
c	321	16	5.4	193	18	AI075707	ox24f07.x1 Soares_tota 2.23e+00
c	322	16	5.4	193	5	R14454	yf82a12.r1 Homo sapien 2.23e+00
c	323	16	5.4	199	19	H59436	yz37c02.s1 Homo sapien 2.23e+00
	324	16	5.4	203	8	AA515099	nf63d02.s1 NCI_CGAP_Co 2.23e+00
	325	16	5.4	204	20	N95076	zb33e08.s1 Homo sapien 2.23e+00
c	326	16	5.4	205	25	AA350972	EST58529 Infant brain 2.23e+00
c	327	16	5.4	205	20	HSC05G062	H. sapiens partial cDN 2.23e+00
	328	16	5.4	208	26	HSU77505	Human genomic clone P3 2.23e+00
	329	16	5.4	209	10	AA526272	zu88c01.s1 Soares test 2.23e+00
c	330	16	5.4	209	20	HSC0SD082	H. sapiens partial cDN 2.23e+00
c	331	16	5.4	214	20	HSCEVH072	H. sapiens partial cDN 2.23e+00
	332	16	5.4	218	9	AA515079	nf63b02.s1 NCI_CGAP_Co 2.23e+00
c	333	16	5.4	219	20	N80595	zb02f08.s1 Homo sapien 2.23e+00
c	334	16	5.4	222	27	AQ007049	CIT-HSP-2286B21.TF CIT 2.23e+00
	335	16	5.4	223	15	AA989408	oa23c08.s1 NCI_CGAP_Ki 2.23e+00
	336	16	5.4	227	27	AQ041979	CIT-HSP-2329H22.TR CIT 2.23e+00
c	337	16	5.4	227	22	AA094536	cp0629.seq.F Fetal hea 2.23e+00
	338	16	5.4	228	8	AA515281	nf64c02.s1 NCI_CGAP_Co 2.23e+00
	339	16	5.4	234	20	N76375	yz77h11.r1 Homo sapien 2.23e+00
	340	16	5.4	235	26	AQ012544	CIT-HSP-2299E11.TR CIT 2.23e+00
c	341	16	5.4	235	22	W84920	SWAMCA1071SK Brugia ma 2.23e+00
c	342	16	5.4	236	12	AA751258	ISSU0069 Rice Immature 2.23e+00
c	343	16	5.4	239	19	HSC2WE041	H. sapiens partial cDN 2.23e+00
	344	16	5.4	240	18	AI057362	oz06g03.x1 Soares_feta 2.23e+00
c	345	16	5.4	242	8	AA491789	ne96g01.s1 NCI_CGAP_Ki 2.23e+00
c	346	16	5.4	242	16	T33719	EST58876 Homo sapiens 2.23e+00
	347	16	5.4	243	10	AA430210	zw59e04.s1 Soares tota 2.23e+00
c	348	16	5.4	243	10	AA651664	ns89d10.s1 NCI_CGAP_Pr 2.23e+00

c	349	16	5.4	249	17	AI071647	UI-R-C2-mw-h-09-0-UI.s 2.23e+00
c	350	16	5.4	251	9	AA241713	MBL2SJ8C7T3 Brugia mal 2.23e+00
	351	16	5.4	254	24	AA313013	EST183807 Bone VII Hom 2.23e+00
c	352	16	5.4	256	5	R33488	yh83d04.s1 Homo sapien 2.23e+00
c	353	16	5.4	257	18	AI155437	ud85e07.r1 Soares mous 2.23e+00
c	354	16	5.4	257	22	R47581	SWAMCA103SK Brugia mal 2.23e+00
c	355	16	5.4	259	19	N29344	yw85c09.s1 Homo sapien 2.23e+00
	356	16	5.4	261	15	AA995781	os29e01.s1 NCI_CGAP_Ki 2.23e+00
c	357	16	5.4	262	14	AA888406	nv74b08.s1 NCI_CGAP_Pr 2.23e+00
c	358	16	5.4	263	16	R40705	yf73g10.s1 Homo sapien 2.23e+00
	359	16	5.4	264	6	C01486	HUMGS0008480, Human Ge 2.23e+00
c	360	16	5.4	264	25	AA372340	EST84488 Colon adenoca 2.23e+00
c	361	16	5.4	265	19	H89736	yu82e08.s1 Homo sapien 2.23e+00
	362	16	5.4	266	15	AA998234	UI-R-C0-ie-b-05-0-UI.s 2.23e+00
	363	16	5.4	266	20	MMU53536	Mus musculus chromosom 2.23e+00
c	364	16	5.4	266	14	AA888168	oj14c04.s1 NCI_CGAP_Ki 2.23e+00
c	365	16	5.4	266	15	AA998234	UI-R-C0-ie-b-05-0-UI.s 2.23e+00
c	366	16	5.4	266	17	AA988236	os08g03.s1 NCI_CGAP_Lu 2.23e+00
c	367	16	5.4	267	5	T52609	yb21d08.r1 Homo sapien 2.23e+00
	368	16	5.4	268	27	AQ0009151	CIT-HSP-2115J20.TF CIT 2.23e+00
c	369	16	5.4	268	19	HSC2BD042	H. sapiens partial cDN 2.23e+00
c	370	16	5.4	268	5	R39953	yf52g12.s1 Homo sapien 2.23e+00
c	371	16	5.4	268	6	AA115052	z106b10.r1 Soares preg 2.23e+00
	372	16	5.4	269	7	AA510652	vh54e02.r1 Soares mous 2.23e+00
	373	16	5.4	269	9	AA570129	ne46g08.s1 NCI_CGAP_Co 2.23e+00
c	374	16	5.4	269	27	AQ0070429	HS_3035_B1_A11_MR CIT 2.23e+00
c	375	16	5.4	270	13	C88353	Mus musculus fertilize 2.23e+00
	376	16	5.4	272	21	AA060198	mj67h11.r1 Soares mous 2.23e+00
	377	16	5.4	273	20	HSC0RB032	H. sapiens partial cDN 2.23e+00
c	378	16	5.4	273	8	AA414724	vc69e03.s1 Knowles Sol 2.23e+00
c	379	16	5.4	275	9	AA534374	nf79f12.s1 NCI_CGAP_Co 2.23e+00
	380	16	5.4	276	18	AI074763	oy25g04.s1 Soares_sene 2.23e+00
c	381	16	5.4	277	27	AQ0072206	CIT-HSP-2294K24.TF CIT 2.23e+00
c	382	16	5.4	279	19	HSC1EC042	H. sapiens partial cDN 2.23e+00
c	383	16	5.4	279	21	W81564	zd88a01.s1 Soares feta 2.23e+00
	384	16	5.4	280	27	AQ0059380	CIT-HSP-2346E18.TR CIT 2.23e+00
c	385	16	5.4	281	19	HSC1PE092	H. sapiens partial cDN 2.23e+00
c	386	16	5.4	281	21	T16503	NIB1403 Normalized inf 2.23e+00
	387	16	5.4	285	24	AA319571	EST21812 Adrenal gland 2.23e+00
c	388	16	5.4	286	6	AA406380	zv10d06.s1 Soares NhHM 2.23e+00
	389	16	5.4	287	10	AA655196	vv13b07.r1 Stratagene 2.23e+00
c	390	16	5.4	287	5	T70566	yd15d09.s1 Homo sapien 2.23e+00
c	391	16	5.4	287	8	AA552746	nk56c12.s1 NCI_CGAP_Pr 2.23e+00
	392	16	5.4	288	15	AI047695	uh82b06.r1 Soares mous 2.23e+00
c	393	16	5.4	289	23	T17418	NIB882 Normalized infa 2.23e+00
	394	16	5.4	290	5	T37630	EST102812 Saccharomyce 2.23e+00
	395	16	5.4	290	16	H27998	yl59g01.s1 Homo sapien 2.23e+00
c	396	16	5.4	290	5	T37630	EST102812 Saccharomyce 2.23e+00
c	397	16	5.4	290	21	T19301	e04006s Testis 2 Homo 2.23e+00
c	398	16	5.4	290	27	AQ0070314	HS_3040_A2_G03_MR CIT 2.23e+00
c	399	16	5.4	292	8	AA474840	vd53c11.s1 Knowles Sol 2.23e+00

c	400	16	5.4	294	8	AA286812	zs54a07.s1 NCI_CGAP_GC 2.23e+00
c	401	16	5.4	295	17	AI016533	ot31h07.s1 Soares_test 2.23e+00
c	402	16	5.4	295	7	HUM090B05A	Human fetal brain cDNA 2.23e+00
	403	16	5.4	296	7	HUM090B05B	Human fetal brain cDNA 2.23e+00
c	404	16	5.4	296	14	AA964428	UI-R-E1-gp-a-09-0-UI.s 2.23e+00
c	405	16	5.4	296	19	HSC2BD052	H. sapiens partial cDN 2.23e+00
c	406	16	5.4	299	25	AA359192	EST68367 Fetal lung II 2.23e+00
c	407	16	5.4	299	14	AA933235	SWBmL3SA243T3 Brugia m 2.23e+00
	408	16	5.4	300	22	N97709	1618C3 czapPFDd2.1, De 2.23e+00
	409	16	5.4	301	5	R46337	yj52b08.s1 Homo sapien 2.23e+00
	410	16	5.4	302	27	AQ0068913	HS_2254_B2_G06_MF CIT 2.23e+00
c	411	16	5.4	302	7	AA459677	zx49c11.s1 Soares test 2.23e+00
c	412	16	5.4	303	5	R50799	yg70h01.s1 Homo sapien 2.23e+00
	413	16	5.4	304	19	N44785	yy24b03.r1 Homo sapien 2.23e+00
c	414	16	5.4	304	17	AI120322	ub69c06.r1 Soares mous 2.23e+00
c	415	16	5.4	305	7	AA497227	fa03f06.s1 Zebrafish I 2.23e+00
c	416	16	5.4	305	7	HUM074D04A	Human fetal brain cDNA 2.23e+00
c	417	16	5.4	305	9	AA241575	MBL2SJ2G9T3 Brugia mal 2.23e+00
c	418	16	5.4	306	27	B80350	CIT-HSP-2045H11.TR CIT 2.23e+00
	419	16	5.4	309	10	AA419937	vf54a03.r1 Soares mous 2.23e+00
	420	16	5.4	310	8	AA508886	ng86f12.s1 NCI_CGAP_Pr 2.23e+00
	421	16	5.4	311	12	AA765291	oa01c07.s1 NCI_CGAP_GC 2.23e+00
c	422	16	5.4	311	16	HSC17H092	H. sapiens partial cDN 2.23e+00
	423	16	5.4	313	17	AA994739	ou18f08.s1 Soares_NFL_ 2.23e+00
	424	16	5.4	319	5	R19277	yf47c04.s1 Homo sapien 2.23e+00
c	425	16	5.4	319	16	H21836	yn78b10.s1 Homo sapien 2.23e+00
	426	16	5.4	321	5	R10374	yf37b05.s1 Homo sapien 2.23e+00
	427	16	5.4	321	19	H82709	yv81a03.s1 Homo sapien 2.23e+00
c	428	16	5.4	321	18	AI184367	qc64c04.x1 Soares_plac 2.23e+00
c	429	16	5.4	322	18	AI141161	qe19f11.x1 Soares_feta 2.23e+00
c	430	16	5.4	323	16	T33133	EST56786 Homo sapiens 2.23e+00
c	431	16	5.4	323	9	AA275507	MBL2SK1F07T3 Brugia ma 2.23e+00
c	432	16	5.4	325	19	HSC2PE042	H. sapiens partial cDN 2.23e+00
c	433	16	5.4	326	17	AI111097	MF001B01.NK3 G. manson 2.23e+00
c	434	16	5.4	327	14	C90671	Dictyostelium discoide 2.23e+00
c	435	16	5.4	327	20	HSC3LB032	H. sapiens partial cDN 2.23e+00
c	436	16	5.4	328	19	H75105	230 Pinus taeda cDNA c 2.23e+00
	437	16	5.4	331	7	CELK133GZR	C.elegans cDNA clone y 2.23e+00
	438	16	5.4	331	6	AA432170	zw71f05.s1 Soares test 2.23e+00
c	439	16	5.4	331	22	AA054927	SWMFCA800SK Brugia mal 2.23e+00
c	440	16	5.4	331	24	AA269020	SWMFCA1816SK Brugia ma 2.23e+00
	441	16	5.4	332	27	AQ0068911	HS_2254_B2_G04_MF CIT 2.23e+00
c	442	16	5.4	332	27	B58477	CIT-HSP-2012P16.TR CIT 2.23e+00
	443	16	5.4	335	10	AA628066	ns67b08.s1 NCI_CGAP_Pr 2.23e+00
	444	16	5.4	335	16	T30358	EST15361 Homo sapiens 2.23e+00
c	445	16	5.4	337	13	AA836432	od21b06.s1 NCI_CGAP_GC 2.23e+00
c	446	16	5.4	338	27	AQ069803	HS_2252_B1_H02_MF CIT 2.23e+00
	447	16	5.4	339	7	AA510684	vh54h01.r1 Soares mous 2.23e+00
	448	16	5.4	340	25	AA324214	EST56409 Infant brain 2.23e+00
	449	16	5.4	341	16	H41918	yo60e02.s1 Homo sapien 2.23e+00
c	450	16	5.4	341	23	AA235572	zt31h07.s1 Soares ovar 2.23e+00

c	451	16	5.4	341	28	AQ128032	HS_3087_A2_D11_MR CIT 2.23e+00
c	452	16	5.4	342	27	AQ0070557	HS_2268_B2_B09_MR CIT 2.23e+00
c	453	16	5.4	344	10	AA664376	ac77b07.s1 Stratagene 2.23e+00
c	454	16	5.4	345	14	AA960113	ub54e03.s1 Soares mous 2.23e+00
	455	16	5.4	347	10	AA625974	zu92f09.s1 Soares test 2.23e+00
	456	16	5.4	347	10	AA631899	np69d09.s1 NCI_CGAP_Br 2.23e+00
c	457	16	5.4	347	7	HUM359G08B	Human aorta cDNA 5'-en 2.23e+00
	458	16	5.4	348	8	AA253243	zr53h06.s1 Soares NhHM 2.23e+00
	459	16	5.4	348	6	AA397314	mr39f11.r1 Life Tech m 2.23e+00
	460	16	5.4	348	14	AA911838	oh09b07.s1 NCI_CGAP_Co 2.23e+00
c	461	16	5.4	348	13	AA841562	MB3D6AA9D05T3 Brugia m 2.23e+00
c	462	16	5.4	348	16	H13682	yj09b08.s1 Homo sapien 2.23e+00
c	463	16	5.4	349	27	B71207	CIT-HSP-2028O12.TF8 CI 2.23e+00
c	464	16	5.4	350	17	AA961875	or44b11.s1 NCI_CGAP_GC 2.23e+00
c	465	16	5.4	353	27	GGA320101	Gallus gallus anonymou 2.23e+00
	466	16	5.4	354	21	W81614	zd88a05.r1 Soares feta 2.23e+00
c	467	16	5.4	355	14	AA915898	oh86b10.s1 NCI_CGAP_Co 2.23e+00
	468	16	5.4	359	6	AA432450	vd92h11.r1 Soares mous 2.23e+00
	469	16	5.4	360	10	W10482	ma53e03.r1 Soares mous 2.23e+00
c	470	16	5.4	360	23	AA237497	mx10a12.r1 Soares mous 2.23e+00
c	471	16	5.4	360	9	C70780	C.elegans cDNA clone y 2.23e+00
c	472	16	5.4	360	9	C69329	C.elegans cDNA clone y 2.23e+00
c	473	16	5.4	361	28	AQ107700	HS_3109_A1_D05_T7 CIT 2.23e+00
c	474	16	5.4	362	22	AA022933	se72g03.s1 Soares feta 2.23e+00
c	475	16	5.4	362	17	AI041286	ov66c11.x1 Soares_test 2.23e+00
c	476	16	5.4	362	15	AA988590	or79g03.s1 NCI_CGAP_Lu 2.23e+00
	477	16	5.4	363	5	R05525	ye92g04.s1 Homo sapien 2.23e+00
	478	16	5.4	364	7	HUMGS01746	Human HL60 3'directed 2.23e+00
c	479	16	5.4	364	26	AQ015851	CIT-HSP-2314C12.TF CIT 2.23e+00
	480	16	5.4	365	21	W42689	zc24a03.s1 Soares sene 2.23e+00
c	481	16	5.4	367	28	AQ095249	HS_3027_A1_C03_MF CIT 2.23e+00
	482	16	5.4	368	28	AQ121119	HS_3075_A2_E04_MR CIT 2.23e+00
c	483	16	5.4	368	18	C99777	Arabidopsis thaliana Y 2.23e+00
c	484	16	5.4	368	17	AI071129	UI-R-C2-mt-c-08-0-UI.s 2.23e+00
c	485	16	5.4	368	17	AI015326	ot35e05.s1 Soares_test 2.23e+00
	486	16	5.4	369	17	AI081969	ov21e07.x1 NCI_CGAP_Br 2.23e+00
c	487	16	5.4	369	14	AA903935	oe73d11.s1 NCI_CGAP_Lu 2.23e+00
c	488	16	5.4	369	16	HSC17A022	H. sapiens partial cDN 2.23e+00
c	489	16	5.4	369	27	B86135	RPCI11-22C12.TP RPCI11 2.23e+00
c	490	16	5.4	370	13	AA840640	CLF9 Leaf cDNA library 2.23e+00
c	491	16	5.4	371	23	AA226903	zr19d09.s1 Stratagene 2.23e+00
c	492	16	5.4	371	28	AQ185449	HS_3065_B1_C06_MR CIT 2.23e+00
c	493	16	5.4	371	8	AA255878	zs19c01.s1 NCI_CGAP_GC 2.23e+00
c	494	16	5.4	372	15	AI047747	uh83a07.r1 Soares mous 2.23e+00
c	495	16	5.4	373	19	N23722	yw46g02.s1 Homo sapien 2.23e+00
c	496	16	5.4	373	12	AA813757	ai69g05.s1 Soares test 2.23e+00
	497	16	5.4	375	5	T38739	EST104288 Saccharomyce 2.23e+00
c	498	16	5.4	375	5	T38739	EST104288 Saccharomyce 2.23e+00
	499	16	5.4	377	12	AA440091	LD15172.5prime LD Dros 2.23e+00
	500	16	5.4	377	18	AI184548	qf29h12.x1 Soares_test 2.23e+00
c	501	16	5.4	378	6	AA406338	zv10f06.s1 Soares NhHM 2.23e+00

c	502	16	5.4	378	15	AA918331	on42e02.s1 NCI_CGAP_Co 2.23e+00
c	503	16	5.4	379	8	AA234339	zr72d12.s1 Soares NhHM 2.23e+00
c	504	16	5.4	379	13	AA598854	ae37a02.s1 Gessler Wil 2.23e+00
	505	16	5.4	381	11	H37052	15181 Lambda-PRL2 Arab 2.23e+00
c	506	16	5.4	381	8	AA482480	zv29a08.s1 Soares ovar 2.23e+00
	507	16	5.4	382	17	AI121587	uc43f03.r1 Soares mous 2.23e+00
c	508	16	5.4	382	22	N97725	1154C3 czapPFDd2.1, De 2.23e+00
c	509	16	5.4	382	28	AQ167240	HS_3171_A1_G04_MR CIT 2.23e+00
c	510	16	5.4	382	13	AA620899	af95h12.s1 Soares test 2.23e+00
c	511	16	5.4	382	18	AI128752	qa63c06.s1 Soares_feta 2.23e+00
	512	16	5.4	383	27	AQ062349	CIT-HSP-2348I22.TF CIT 2.23e+00
c	513	16	5.4	384	17	AI040628	ox19h07.s1 Soares_feta 2.23e+00
c	514	16	5.4	384	17	AA878246	oi04e03.s1 NCI_CGAP_HN 2.23e+00
c	515	16	5.4	384	11	AA635423	nt68d08.s1 NCI_CGAP_Pr 2.23e+00
c	516	16	5.4	385	16	H17600	ym36g04.r1 Homo sapien 2.23e+00
	517	16	5.4	386	17	AI129929	qc45e03.x1 Soares_preg 2.23e+00
c	518	16	5.4	386	21	N92365	zb62a07.s1 Soares feta 2.23e+00
c	519	16	5.4	386	21	W37626	zc10a11.s1 Soares para 2.23e+00
c	520	16	5.4	387	12	AA765691	oa09c04.s1 NCI_CGAP_GC 2.23e+00
	521	16	5.4	388	16	H08533	yl89b11.r1 Homo sapien 2.23e+00
c	522	16	5.4	388	17	AI026002	ow03c05.s1 Soares_para 2.23e+00
	523	16	5.4	389	15	AA926979	om26b07.s1 Soares_NFL_ 2.23e+00
c	524	16	5.4	389	10	AA647037	ve45f05.r1 Soares mous 2.23e+00
	525	16	5.4	391	18	AI076559	os17e11.x1 Soares_feta 2.23e+00
c	526	16	5.4	391	27	AQ045077	RPC111-34N15.TK RPC111 2.23e+00
	527	16	5.4	392	22	AA023888	mh93h06.r1 Soares mous 2.23e+00
c	528	16	5.4	392	29	AQ150918	HS_3110_B1_E11_MR CIT 2.23e+00
c	529	16	5.4	392	9	AA551691	nf98f11.s1 NCI_CGAP_Co 2.23e+00
c	530	16	5.4	392	13	AA860641	aj82e06.s1 Soares para 2.23e+00
	531	16	5.4	393	28	AQ191867	HS_3233_B1_G05_MR CIT 2.23e+00
c	532	16	5.4	393	5	R50403	yj62d06.s1 Homo sapien 2.23e+00
c	533	16	5.4	393	14	AA959567	ub54g06.s1 Soares mous 2.23e+00
c	534	16	5.4	393	13	AA843278	aj16c03.s1 Soares para 2.23e+00
c	535	16	5.4	393	23	AA047874	zf50b06.s1 Soares reti 2.23e+00
c	536	16	5.4	393	18	AI168692	ox70d10.s1 Soares_NhHM 2.23e+00
c	537	16	5.4	393	15	AI048330	uh59b05.r1 Soares mous 2.23e+00
	538	16	5.4	394	11	AA708902	xf94d08.s1 Soares pine 2.23e+00
	539	16	5.4	395	14	AA951970	LD32893.3prime LD Dros 2.23e+00
c	540	16	5.4	395	29	AQ138447	HS_3075_B1_E08_MR CIT 2.23e+00
c	541	16	5.4	395	16	H19482	yn59a02.s1 Homo sapien 2.23e+00
c	542	16	5.4	395	16	H11176	ym09h04.r1 Homo sapien 2.23e+00
	543	16	5.4	396	16	R78451	yi82g10.s1 Homo sapien 2.23e+00
	544	16	5.4	396	14	AA978956	LD33172.5prime LD Dros 2.23e+00
c	545	16	5.4	396	14	C92065	Dictyostelium discoide 2.23e+00
c	546	16	5.4	396	17	AI016536	ot31h10.s1 Soares_test 2.23e+00
c	547	16	5.4	396	28	AQ094959	HS_3028_A2_D02_MF CIT 2.23e+00
	548	16	5.4	397	5	T98207	ye56h11.s1 Homo sapien 2.23e+00
	549	16	5.4	397	14	AA978608	LD32070.3prime LD Dros 2.23e+00
	550	16	5.4	397	17	AI051431	ox42h02.s1 Soares_tota 2.23e+00
c	551	16	5.4	397	27	B80935	CIT-HSP-2025L4.TF8 CIT 2.23e+00
	552	16	5.4	398	15	AI061765	LD34871.3prime LD Dros 2.23e+00

	553	16	5.4	398	6	AA387831	vc82h09.r1 Ko mouse em 2.23e+00
c	554	16	5.4	398	27	B94902	CIT-HSP-2170H5.TR CIT- 2.23e+00
c	555	16	5.4	398	28	AQ122895	HS_3094_B1_A02_MR CIT 2.23e+00
c	556	16	5.4	398	12	AA536282	LD16796.5prime LD Dros 2.23e+00
	557	16	5.4	399	12	T04497	544 Lambda-PRL2 Arabid 2.23e+00
	558	16	5.4	399	14	AA950895	LD31056.3prime LD Dros 2.23e+00
c	559	16	5.4	400	28	AQ216221	HS_3240_B2_H06_MR CIT 2.23e+00
c	560	16	5.4	401	16	H02867	yj43h01.s1 Homo sapien 2.23e+00
c	561	16	5.4	401	7	AA474713	vd56c05.r1 Knowles Sol 2.23e+00
c	562	16	5.4	401	19	N33043	yy05d06.s1 Homo sapien 2.23e+00
c	563	16	5.4	403	22	H96489	yt98f06.r1 Soares pine 2.23e+00
	564	16	5.4	404	16	H03302	yj47c10.r1 Homo sapien 2.23e+00
	565	16	5.4	404	18	AI154662	ud37c09.r1 Soares mous 2.23e+00
c	566	16	5.4	404	28	AQ212593	HS_3108_B2_E04_MR CIT 2.23e+00
c	567	16	5.4	405	19	H73455	yu48h09.s1 Homo sapien 2.23e+00
	568	16	5.4	406	11	AA076546	zm19e02.r1 Stratagene 2.23e+00
	569	16	5.4	406	28	AQ111367	CIT-HSP-2371M16.TR CIT 2.23e+00
	570	16	5.4	407	21	N94344	zb75f08.s1 Soares sene 2.23e+00
c	571	16	5.4	407	8	AA252201	zr64f12.s1 Soares NhHM 2.23e+00
	572	16	5.4	408	21	W29818	mc06h01.r1 Soares mous 2.23e+00
	573	16	5.4	408	9	AA553422	nk77b04.s1 NCI_CGAP_Sc 2.23e+00
	574	16	5.4	408	6	AA387824	vc82f09.r1 Ko mouse em 2.23e+00
c	575	16	5.4	408	14	AA938716	oc09h03.s1 NCI_CGAP_GC 2.23e+00
c	576	16	5.4	408	11	AA695695	GM04310.5prime GM Dros 2.23e+00
c	577	16	5.4	408	13	AA832180	oc91c11.s1 NCI_CGAP_GC 2.23e+00
c	578	16	5.4	408	14	AA864620	od45c08.s1 NCI_CGAP_GC 2.23e+00
	579	16	5.4	409	18	AI141236	qa45g01.s1 Soares_NhHM 2.23e+00
c	580	16	5.4	409	10	AA613041	nq40e09.s1 NCI_CGAP_Co 2.23e+00
	581	16	5.4	410	5	R20038	yg31h05.r1 Homo sapien 2.23e+00
	582	16	5.4	410	26	AQ021551	CIT-HSP-2301F3.TR CIT- 2.23e+00
c	583	16	5.4	410	19	N63096	yz32g04.s1 Homo sapien 2.23e+00
c	584	16	5.4	410	27	BB6153	RPCI11-22D12.TP RPCI11 2.23e+00
c	585	16	5.4	410	9	AA588341	nm93b12.s1 NCI_CGAP_Co 2.23e+00
	586	16	5.4	411	23	AA124373	mq76h11.r1 Stratagene 2.23e+00
	587	16	5.4	411	27	AQ043907	CIT-HSP-2318K11.TR CIT 2.23e+00
	588	16	5.4	412	20	W44023	mc71f09.r1 Soares mous 2.23e+00
c	589	16	5.4	412	22	N59135	yz62d10.s1 Soares mult 2.23e+00
c	590	16	5.4	412	16	R44862	yg25b10.s1 Homo sapien 2.23e+00
	591	16	5.4	413	11	AA076458	zm91h08.r1 Stratagene 2.23e+00
	592	16	5.4	413	21	AA033241	mi39c07.r1 Soares mous 2.23e+00
c	593	16	5.4	413	26	B17074	344E12.TPB CIT978SKA1 2.23e+00
	594	16	5.4	415	6	AA410439	zv12d04.s1 Soares NhHM 2.23e+00
	595	16	5.4	415	10	AA418347	zv96d12.r1 Soares NhHM 2.23e+00
c	596	16	5.4	415	28	AQ184010	HS_3213_A2_G12_MR CIT 2.23e+00
	597	16	5.4	416	28	AQ210488	HS_2202_B2_D12_MF CIT 2.23e+00
	598	16	5.4	416	6	AA405468	zw36c02.s1 Soares tota 2.23e+00
c	599	16	5.4	416	28	AQ200657	RPCI11-47H23.TJ RPCI11 2.23e+00
c	600	16	5.4	417	14	AA987529	os11e12.s1 NCI_CGAP_Lu 2.23e+00
	601	16	5.4	418	12	AA820320	LD23880.3prime LD Dros 2.23e+00
c	602	16	5.4	418	15	AA987318	or81h12.s1 NCI_CGAP_Lu 2.23e+00
c	603	16	5.4	418	21	W80789	zd83e04.s1 Soares feta 2.23e+00

c	604	16	5.4	420	26	B96038	F23D6TFB IGF Arabidops 2.23e+00
c	605	16	5.4	421	7	AA460673	zx64d06.s1 Soares tota 2.23e+00
c	606	16	5.4	421	6	AA433264	EST78 Drosophila melan 2.23e+00
c	607	16	5.4	422	22	AA088266	zl82e08.s1 Stratagene 2.23e+00
c	608	16	5.4	422	8	AA258495	zr59a09.r1 Soares NhHM 2.23e+00
	609	16	5.4	423	8	AA279679	zs92a11.r1 NCI_CGAP_GC 2.23e+00
c	610	16	5.4	423	17	AI086901	oz58g09.x1 Soares_sene 2.23e+00
c	611	16	5.4	423	28	AQ181108	HS_3218_A1_D05_MR CIT 2.23e+00
c	612	16	5.4	424	23	AA124085	mp72g08.r1 Soares 2NbM 2.23e+00
c	613	16	5.4	424	12	AA721096	nx90d07.s1 NCI_CGAP_GC 2.23e+00
	614	16	5.4	425	23	AA210129	mu42d06.r1 Soares 2NbM 2.23e+00
	615	16	5.4	425	29	AQ136138	HS_3060_B1_E07_MR CIT 2.23e+00
	616	16	5.4	426	19	N72856	za03e07.s1 Homo sapien 2.23e+00
	617	16	5.4	426	7	AA462741	vf91b04.r1 Soares mous 2.23e+00
	618	16	5.4	426	19	N50950	yy93g01.s1 Homo sapien 2.23e+00
c	619	16	5.4	426	18	AI079461	oz39f01.s1 Soares_NhHM 2.23e+00
c	620	16	5.4	426	13	AA653454	ag66c05.s1 Gessler Wil 2.23e+00
	621	16	5.4	427	20	AA005380	zh88g04.s1 Soares feta 2.23e+00
c	622	16	5.4	427	15	AA988592	oz79g05.s1 NCI_CGAP_Lu 2.23e+00
c	623	16	5.4	427	6	AA292378	zt51g11.s1 Soares ovar 2.23e+00
	624	16	5.4	428	16	R88997	yp97c12.s1 Homo sapien 2.23e+00
	625	16	5.4	428	17	AI106528	bb03a08_p1 ZF adult he 2.23e+00
c	626	16	5.4	428	10	AA419281	zv35f01.s1 Soares ovar 2.23e+00
c	627	16	5.4	428	6	AA406120	zv16c08.r1 Soares NhHM 2.23e+00
	628	16	5.4	429	6	AA132041	zl38b01.s1 Soares preg 2.23e+00
	629	16	5.4	429	10	AA595251	no33b11.s1 NCI_CGAP_Pr 2.23e+00
c	630	16	5.4	429	12	AA801961	GM03639.5prime GM Dros 2.23e+00
	631	16	5.4	430	29	AQ240443	CIT-HSP-2386N8.TR.1 CI 2.23e+00
	632	16	5.4	430	22	AA187933	zp75b05.s1 Stratagene 2.23e+00
c	633	16	5.4	430	9	AA554306	nl03b03.s1 NCI_CGAP_Co 2.23e+00
c	634	16	5.4	430	13	AA812239	nr82a05.s1 NCI_CGAP_Pr 2.23e+00
	635	16	5.4	431	15	AA961460	op44a06.s1 Soares_NFL_ 2.23e+00
	636	16	5.4	431	12	AA735230	LD01640.5prime LD Dros 2.23e+00
	637	16	5.4	431	12	AA735229	LD01640.3prime LD Dros 2.23e+00
	638	16	5.4	431	15	AI041882	ox57g08.x1 Soares_NhHM 2.23e+00
c	639	16	5.4	431	19	N62639	yz75h02.s1 Homo sapien 2.23e+00
c	640	16	5.4	431	14	AA916656	oh85e08.s1 NCI_CGAP_Co 2.23e+00
c	641	16	5.4	431	19	H03343	yw82d08.r1 Homo sapien 2.23e+00
	642	16	5.4	432	22	AA147635	zl52f01.s1 Soares preg 2.23e+00
	643	16	5.4	432	5	R54837	yj74b02.s1 Homo sapien 2.23e+00
c	644	16	5.4	432	16	H05006	yi74e09.s1 Homo sapien 2.23e+00
c	645	16	5.4	432	14	AA876428	oj24h02.s1 NCI_CGAP_Ki 2.23e+00
	646	16	5.4	433	12	AA816322	LD01195.5prime LD Dros 2.23e+00
	647	16	5.4	433	6	AA029108	zk09b06.r1 Soares preg 2.23e+00
	648	16	5.4	434	19	N66746	yz33h05.s1 Homo sapien 2.23e+00
c	649	16	5.4	434	10	AA619085	vc67e10.r1 Soares mous 2.23e+00
c	650	16	5.4	434	21	AA052830	mb32h09.r1 Soares mous 2.23e+00
c	651	16	5.4	434	7	AA471584	MBAFCX4B08T3 Brugia ma 2.23e+00
c	652	16	5.4	434	7	AA463464	zx98d07.s1 Soares NhHM 2.23e+00
c	653	16	5.4	434	29	AQ139228	HS_3090_A1_E11_MR CIT 2.23e+00
	654	16	5.4	435	21	AA053671	zl74g04.s1 Stratagene 2.23e+00

	655	16	5.4	435	26	B97555	T33P23TR TAMU Arabidop 2.23e+00
	656	16	5.4	435	29	AQ231012	HS_2037_B2_G01_MR CIT 2.23e+00
	657	16	5.4	435	15	AA909131	oml3b01.s1 Soares_NFL_ 2.23e+00
c	658	16	5.4	435	7	AF001163	Xenopus laevis clone n 2.23e+00
c	659	16	5.4	435	12	AA774956	as84b07.s1 Stratagene 2.23e+00
	660	16	5.4	436	28	AQ180166	HS_3200_B1_E01_MR CIT 2.23e+00
	661	16	5.4	436	12	AA732008	nw61d05.s1 NCI_CGAP_GC 2.23e+00
c	662	16	5.4	436	15	AA946787	oq50f07.s1 NCI_CGAP_Ki 2.23e+00
c	663	16	5.4	436	27	AQ009421	RPCI11-22D12.TPD RPCI1 2.23e+00
	664	16	5.4	437	15	AI033795	ox20b07.x1 Soares_feta 2.23e+00
c	665	16	5.4	437	13	AA829393	od06c07.s1 NCI_CGAP_GC 2.23e+00
c	666	16	5.4	437	10	AA421016	zu09c09.s1 Soares test 2.23e+00
	667	16	5.4	439	19	N40992	yy80e06.s1 Homo sapien 2.23e+00
	668	16	5.4	439	16	R72518	yj90f11.s1 Homo sapien 2.23e+00
	669	16	5.4	439	28	AQ196578	CIT-HSP-2380D10.TF CIT 2.23e+00
	670	16	5.4	441	14	AA941896	LD27228.5prime LD Dros 2.23e+00
c	671	16	5.4	441	12	AA735772	GM09828.5prime GM Dros 2.23e+00
c	672	16	5.4	441	16	H02770	yj36g10.s1 Homo sapien 2.23e+00
	673	16	5.4	443	28	AQ207663	HS_2249_A2_F11_MR CIT 2.23e+00
c	674	16	5.4	443	17	AI080734	ox54f01.x1 Soares_tota 2.23e+00
c	675	16	5.4	443	19	N70058	za70d05.s1 Homo sapien 2.23e+00
	676	16	5.4	444	13	AA806255	oc21g01.s1 NCI_CGAP_GC 2.23e+00
c	677	16	5.4	444	22	N63075	yz32d02.s1 Soares mult 2.23e+00
c	678	16	5.4	445	22	N76759	yv45b06.r1 Soares feta 2.23e+00
c	679	16	5.4	445	6	AA412026	zt68c09.s1 Soares test 2.23e+00
c	680	16	5.4	445	24	AA288121	vb12c10.r1 Soares mous 2.23e+00
c	681	16	5.4	445	18	AI073809	oy69d04.x1 NCI_CGAP_CL 2.23e+00
	682	16	5.4	446	8	AA468619	ne08b10.s1 NCI_CGAP_Co 2.23e+00
c	683	16	5.4	446	18	AI170988	EST216932 Normalized r 2.23e+00
c	684	16	5.4	447	16	H24327	ym51e03.s1 Homo sapien 2.23e+00
	685	16	5.4	448	17	AI076082	oy80d05.x1 NCI_CGAP_CL 2.23e+00
	686	16	5.4	448	18	AI155625	ue02e11.r1 Soares mous 2.23e+00
c	687	16	5.4	448	14	AA848019	od24h04.s1 NCI_CGAP_GC 2.23e+00
	688	16	5.4	449	19	N49538	yv23f12.r1 Homo sapien 2.23e+00
	689	16	5.4	449	18	AI174601	an47b09.s1 Gessler Wil 2.23e+00
	690	16	5.4	449	6	AA127393	zn81b04.s1 Stratagene 2.23e+00
c	691	16	5.4	449	6	AA040119	zk45b12.s1 Soares preg 2.23e+00
	692	16	5.4	451	13	AA872134	oi07c09.s1 NCI_CGAP_GC 2.23e+00
c	693	16	5.4	451	27	AQ080936	CIT-HSP-2367N4.TF CIT- 2.23e+00
	694	16	5.4	452	19	H93925	yv06f07.r1 Homo sapien 2.23e+00
c	695	16	5.4	453	26	AQ016187	CIT-HSP-2314O10.TF CIT 2.23e+00
c	696	16	5.4	454	18	AI185184	qe35f09.s1 Soares_feta 2.23e+00
	697	16	5.4	455	10	T21936	3944 Lambda-PRL2 Arabi 2.23e+00
	698	16	5.4	455	13	AA831739	oc93b05.s1 NCI_CGAP_GC 2.23e+00
c	699	16	5.4	455	13	AA825566	od55d05.s1 NCI_CGAP_GC 2.23e+00
c	700	16	5.4	455	13	AA814079	ob95b07.s1 NCI_CGAP_GC 2.23e+00
	701	16	5.4	456	15	AA989070	or88d05.s1 NCI_CGAP_Lu 2.23e+00
	702	16	5.4	456	13	AA806448	oc23c05.s1 NCI_CGAP_GC 2.23e+00
	703	16	5.4	456	12	AA779853	af45h04.s1 Soares tota 2.23e+00
	704	16	5.4	457	18	AI076079	oy80d01.x1 NCI_CGAP_CL 2.23e+00
c	705	16	5.4	457	6	AA436162	zv22c05.s1 Soares NhHM 2.23e+00

c	706	16	5.4	457	15	AI013016	EST207467 Normalized r 2.23e+00
c	707	16	5.4	457	12	AA802556	GM05818.5prime GM Dros 2.23e+00
	708	16	5.4	458	21	W46996	zc43g01.s1 Soares sene 2.23e+00
	709	16	5.4	458	19	N68991	yz79b11.s1 Homo sapien 2.23e+00
c	710	16	5.4	458	7	AA508939	MBAFCW1D09T3 Brugia ma 2.23e+00
c	711	16	5.4	459	27	AQ060802	CIT-HSP-2349J11.TF CIT 2.23e+00
	712	16	5.4	460	26	B53355	CIT-HSP-2010C8.TF CIT- 2.23e+00
	713	16	5.4	461	12	AA774669	ac13g01.s1 Stratagene 2.23e+00
c	714	16	5.4	461	22	AA053921	ze75c06.s1 Soares feta 2.23e+00
c	715	16	5.4	461	18	AI149215	qc81g02.x1 Soares_plac 2.23e+00
c	716	16	5.4	461	17	AI093466	qb08d10.x1 Soares_preg 2.23e+00
c	717	16	5.4	461	13	AA828282	ob60b03.s1 NCI_CGAP_GC 2.23e+00
	718	16	5.4	462	26	B41806	HS-1054-B1-A01-MF.abi 2.23e+00
c	719	16	5.4	462	28	AQ114350	CIT-HSP-2378D2.TR CIT- 2.23e+00
c	720	16	5.4	462	21	W42797	zc24a03.r1 Soares sene 2.23e+00
	721	16	5.4	464	17	AI086431	qf21g12.x1 NCI_CGAP_Br 2.23e+00
	722	16	5.4	464	23	AA231407	mw35b04.r1 Soares mous 2.23e+00
c	723	16	5.4	464	22	AA019480	ze57c06.r1 Soares reti 2.23e+00
c	724	16	5.4	464	21	W47031	zc43g01.r1 Soares sene 2.23e+00
c	725	16	5.4	464	11	AA698077	HL03739.5prime HL Dros 2.23e+00
c	726	16	5.4	464	27	AQ003956	CIT-HSP-2290I5.TR CIT- 2.23e+00
	727	16	5.4	465	17	AI077547	oz33e09.x1 Soares_tota 2.23e+00
c	728	16	5.4	465	19	N29323	yw84b12.s1 Homo sapien 2.23e+00
	729	16	5.4	467	18	AI147229	qb36e02.x1 Soares_preg 2.23e+00
	730	16	5.4	467	22	AA152407	zo07f02.s1 Stratagene 2.23e+00
c	731	16	5.4	467	16	R46368	yg50a08.s1 Homo sapien 2.23e+00
	732	16	5.4	470	29	AQ226305	HS_2010_A2_B04_T7 CIT 2.23e+00
	733	16	5.4	470	15	AA974339	oq10h09.s1 NCI_CGAP_GC 2.23e+00
c	734	16	5.4	470	16	H11977	ym15f12.s1 Homo sapien 2.23e+00
	735	16	5.4	472	19	N52761	yz18a05.s1 Homo sapien 2.23e+00
c	736	16	5.4	472	17	AI126081	qc54e02.x1 Soares_plac 2.23e+00
c	737	16	5.4	473	29	AQ227789	HS_2012_B2_H08_MR CIT 2.23e+00
c	738	16	5.4	473	7	AA500167	vg07a05.r1 Soares mous 2.23e+00
c	739	16	5.4	474	20	W51515	AS4SLC946T3 Brugia mal 2.23e+00
	740	16	5.4	475	16	R87815	yq33g04.r1 Homo sapien 2.23e+00
	741	16	5.4	475	14	AA955033	oq66h05.s1 NCI_CGAP_Ki 2.23e+00
	742	16	5.4	476	17	AI042050	ox60c08.x1 Soares_NhHM 2.23e+00
	743	16	5.4	476	8	AA413693	vc67a06.s1 Knowles Sol 2.23e+00
	744	16	5.4	476	19	N63039	yy70d11.s1 Homo sapien 2.23e+00
c	745	16	5.4	476	18	AI186348	qd30c04.x1 Soares_plac 2.23e+00
c	746	16	5.4	476	19	N62766	yz82e06.s1 Homo sapien 2.23e+00
c	747	16	5.4	476	19	N43443	AS4SLC542T3 Brugia mal 2.23e+00
c	748	16	5.4	476	17	AI031565	ow48d11.x1 Soares_para 2.23e+00
	749	16	5.4	477	29	AQ151698	HS_3095_A2_D07_T7 CIT 2.23e+00
c	750	16	5.4	477	16	RB7351	ym88b05.r1 Homo sapien 2.23e+00
c	751	16	5.4	477	5	R50813	yg62a08.s1 Homo sapien 2.23e+00
c	752	16	5.4	478	20	W51568	AS4SLC749T3 Brugia mal 2.23e+00
	753	16	5.4	479	10	AA636402	vq86b08.s1 Knowles Sol 2.23e+00
	754	16	5.4	479	18	AI160292	qb66b01.x1 Soares_feta 2.23e+00
c	755	16	5.4	479	28	AQ179012	HS_3071_B1_H04_MR CIT 2.23e+00
c	756	16	5.4	479	13	AA811209	oa76g09.s1 NCI_CGAP_GC 2.23e+00

c	757	16	5.4	479	13	AA846832	aj99h09.s1 Soares para	2.23e+00
c	758	16	5.4	480	22	W94745	ze13a11.s1 Soares feta	2.23e+00
	759	16	5.4	481	7	HUMHBC4571	similar to none, mRNA	2.23e+00
c	760	16	5.4	481	19	N30561	yw66b03.s1 Homo sapien	2.23e+00
	761	16	5.4	482	29	AQ140284	HS_3113_A2_G09_MR CIT	2.23e+00
	762	16	5.4	482	19	H92965	yv07a03.s1 Homo sapien	2.23e+00
c	763	16	5.4	482	19	N43412	AS4SLC534T3 Brugia mal	2.23e+00
c	764	16	5.4	483	10	AA581872	nn43f10.s1 NCI_CGAP_GC	2.23e+00
	765	16	5.4	484	18	AI094741	qa08h01.x1 NCI_CGAP_Br	2.23e+00
c	766	16	5.4	484	28	AQ215911	HS_3243_B2_E03_T7 CIT	2.23e+00
	767	16	5.4	485	15	AI010067	EST204518 Normalized r	2.23e+00
	768	16	5.4	486	11	T20979	2987 Lambda-PRL2 Arabi	2.23e+00
c	769	16	5.4	486	29	AQ253629	HS_3026_A1_G05_MF CIT	2.23e+00
	770	16	5.4	487	12	AA764630	vp10h12.r1 Soares mous	2.23e+00
c	771	16	5.4	487	11	AA713281	32841 Lambda-PRL2 Arab	2.23e+00
c	772	16	5.4	487	12	AA781886	ai66e08.s1 Soares test	2.23e+00
	773	16	5.4	488	14	AA981806	ua31h03.r1 Soares mous	2.23e+00
c	774	16	5.4	489	11	AA567504	HL01357.5prime HL Dros	2.23e+00
	775	16	5.4	490	12	AA539962	LD19272.5prime LD Dros	2.23e+00
	776	16	5.4	491	29	AQ233709	HS_2053_A2_C02_T7 CIT	2.23e+00
c	777	16	5.4	491	26	B54555	CIT-HSP-347C6.TVC CIT-	2.23e+00
c	778	16	5.4	492	27	B99842	CIT-HSP-2288K14.TR CIT	2.23e+00
c	779	16	5.4	492	13	AA830080	oc44b12.s1 NCI_CGAP_GC	2.23e+00
	780	16	5.4	493	15	AI025579	ov87d12.s1 Soares_test	2.23e+00
	781	16	5.4	493	26	AQ017527	CIT-HSP-2306H13.TR CIT	2.23e+00
c	782	16	5.4	494	18	AI102281	EST211570 Normalized r	2.23e+00
	783	16	5.4	495	22	AA034430	ze75c06.r1 Soares feta	2.23e+00
	784	16	5.4	495	12	AA739520	285 PtIPG2 Pinus taeda	2.23e+00
	785	16	5.4	495	7	AA443023	zx98d06.r1 Soares NhHM	2.23e+00
	786	16	5.4	496	17	AI039717	ox32h08.s1 Soares_tota	2.23e+00
c	787	16	5.4	496	29	AQ227534	HS_2019_B2_D10_MR CIT	2.23e+00
c	788	16	5.4	496	23	AA260299	va95d11.r1 Soares mous	2.23e+00
c	789	16	5.4	497	24	AA275022	vb05d04.r1 Soares mous	2.23e+00
	790	16	5.4	498	10	AA584424	nn81c10.s1 NCI_CGAP_Co	2.23e+00
	791	16	5.4	498	23	AA233088	zr68c02.r1 Soares NhHM	2.23e+00
c	792	16	5.4	498	22	N54820	yv31f05.s1 Soares feta	2.23e+00
c	793	16	5.4	500	8	AA549845	0918m3 gmbPfHB3.1, G.	2.23e+00
c	794	16	5.4	500	12	AA140871	CK01090.5prime CK Dros	2.23e+00
c	795	16	5.4	502	16	H06839	yl83g10.s1 Homo sapien	2.23e+00
c	796	16	5.4	503	15	AI036891	ub66h12.r1 Soares mous	2.23e+00
	797	16	5.4	505	5	T79591	yd71b08.s1 Homo sapien	2.23e+00
c	798	16	5.4	505	18	AI171539	EST217504 Normalized r	2.23e+00
	799	16	5.4	506	24	AA266656	mx90e03.r1 Soares mous	2.23e+00
	800	16	5.4	507	14	AA915364	vz29b10.r1 Soares 2NbM	2.23e+00
	801	16	5.4	507	13	AA643468	nr59e02.s1 NCI_CGAP_Ly	2.23e+00
c	802	16	5.4	507	28	AQ189020	HS_3183_B2_E10_T7 CIT	2.23e+00
c	803	16	5.4	507	28	AQ169564	HS_3176_B2_C04_T7 CIT	2.23e+00
	804	16	5.4	508	19	N36894	yy37h02.s1 Homo sapien	2.23e+00
	805	16	5.4	510	13	AA811613	ob74b06.s2 NCI_CGAP_GC	2.23e+00
	806	16	5.4	512	23	AA125566	mq77h11.r1 Stratagene	2.23e+00
c	807	16	5.4	512	8	AA482508	zv29h03.s1 Soares ovar	2.23e+00

	808	16	5.4	513	27	AQ057830	CIT-HSP-2339F12.TF CIT	2.23e+00
c	809	16	5.4	513	19	N23413	yx84c11.s1 Homo sapien	2.23e+00
c	810	16	5.4	513	6	AA151136	zl44f12.s1 Soares preg	2.23e+00
c	811	16	5.4	513	14	C92225	Dictyostelium discoide	2.23e+00
	812	16	5.4	514	18	AI180782	ub76c04.r1 Soares mous	2.23e+00
	813	16	5.4	515	28	AQ164409	HS_3001_B2_G02_MF CIT	2.23e+00
	814	16	5.4	515	18	AI156765	ue54c12.r1 Soares mous	2.23e+00
	815	16	5.4	515	28	AQ207176	HS_3239_B1_C07_T7 CIT	2.23e+00
c	816	16	5.4	516	26	AQ010938	F23D6TPC IGP Arabidops	2.23e+00
	817	16	5.4	517	14	AA915899	oh86b11.s1 NCI_CGAP_Co	2.23e+00
c	818	16	5.4	517	20	W02449	za47a06.r1 Soares feta	2.23e+00
	819	16	5.4	520	17	AI113892	GH10388.3prime GH Dros	2.23e+00
c	820	16	5.4	520	26	AQ021727	CIT-HSP-2310N19.TR CIT	2.23e+00
	821	16	5.4	521	29	AQ143212	HS_3067_A1_A11_MF CIT	2.23e+00
	822	16	5.4	521	28	AQ199899	RPCI11-59F4.TJ RPCI11	2.23e+00
c	823	16	5.4	521	12	AA819766	UI-R-A0-av-h-07-0-UI.s	2.23e+00
	824	16	5.4	522	12	AA816608	LD02664.3prime LD Dros	2.23e+00
	825	16	5.4	522	12	AA390297	LD03020.5prime LD Dros	2.23e+00
c	826	16	5.4	524	10	AA615239	vn99f03.r1 Stratagene	2.23e+00
c	827	16	5.4	526	13	AA854903	aj78f09.s1 Soares para	2.23e+00
	828	16	5.4	527	10	AA415730	vd25c10.s1 Knowles Sol	2.23e+00
	829	16	5.4	527	12	AA441643	LD16642.3prime LD Dros	2.23e+00
c	830	16	5.4	529	14	AA900730	UI-R-E0-di-a-08-0-UI.s	2.23e+00
	831	16	5.4	530	12	AA540934	LD20127.5prime LD Dros	2.23e+00
c	832	16	5.4	531	28	AQ124086	HS_3121_A1_G06_T7 CIT	2.23e+00
c	833	16	5.4	531	5	R41765	yg11g06.s1 Homo sapien	2.23e+00
c	834	16	5.4	532	27	FR0033471	Fugu rubripes GSS sequ	2.23e+00
c	835	16	5.4	532	26	AQ011334	F27B20TPC IGP Arabidop	2.23e+00
c	836	16	5.4	532	18	AI180191	EST223932 Normalized r	2.23e+00
	837	16	5.4	533	14	AA942516	LD26774.5prime LD Dros	2.23e+00
	838	16	5.4	534	14	AA950006	LD28609.3prime LD Dros	2.23e+00
	839	16	5.4	536	12	AA540703	LD20479.5prime LD Dros	2.23e+00
c	840	16	5.4	536	28	AQ182524	HS_3077_B1_C10_MR CIT	2.23e+00
c	841	16	5.4	536	18	AI144653	UI-R-BT0-pl-e-10-0-UI.	2.23e+00
	842	16	5.4	538	28	B58763	CIT-HSP-2013D19.TR CIT	2.23e+00
c	843	16	5.4	538	12	AA802380	GM04290.5prime GM Dros	2.23e+00
c	844	16	5.4	539	28	AQ213734	HS_3250_A1_H01_T7 CIT	2.23e+00
c	845	16	5.4	541	11	AA695592	GM03091.5prime GM Dros	2.23e+00
	846	16	5.4	542	18	AI056865	oz03e04.x1 Soares_feta	2.23e+00
	847	16	5.4	542	8	AA526900	ni09f10.s1 NCI_CGAP_Br	2.23e+00
	848	16	5.4	543	14	AA948815	LD27468.3prime LD Dros	2.23e+00
c	849	16	5.4	543	17	AU022825	Mus musculus unfertili	2.23e+00
c	850	16	5.4	544	28	AQ189530	HS_3203_B2_C10_T7 CIT	2.23e+00
c	851	16	5.4	545	12	AA779670	af42g04.s1 Soares tota	2.23e+00
c	852	16	5.4	545	15	AI062812	GH02155.5prime GH Dros	2.23e+00
	853	16	5.4	547	5	T70156	yc17g06.r1 Homo sapien	2.23e+00
	854	16	5.4	548	23	AA238724	my35a11.r1 Barstead mo	2.23e+00
c	855	16	5.4	548	7	AA447655	zw97h03.s1 Soares tota	2.23e+00
	856	16	5.4	551	17	AU019743	Mus musculus 8-cell em	2.23e+00
c	857	16	5.4	554	29	AQ153218	HS_2218_A1_D02_MF CIT	2.23e+00
c	858	16	5.4	555	27	B83845	RPCI11-20J6.TV RPCI11	2.23e+00

	859	16	5.4	557	17	AI107941	GH05841.5prime GH Dros	2.23e+00
	860	16	5.4	557	19	N73934	za74b03.s1 Homo sapien	2.23e+00
c	861	16	5.4	557	28	AQ215917	HS_3243_B2_D12_T7 CIT	2.23e+00
c	862	16	5.4	560	15	AI061825	GH02936.5prime GH Dros	2.23e+00
	863	16	5.4	561	11	H36258	14780 Lambda-PRL2 Arab	2.23e+00
c	864	16	5.4	561	17	AI108068	GH06793.5prime GH Dros	2.23e+00
	865	16	5.4	565	14	AA950509	LD30331.3prime LD Dros	2.23e+00
	866	16	5.4	565	12	AA390305	LD03003.3prime LD Dros	2.23e+00
	867	16	5.4	567	23	AA057432	zf57d12.s1 Soares reti	2.23e+00
	868	16	5.4	569	23	AA200480	mu36h07.r1 Soares 2NbM	2.23e+00
	869	16	5.4	570	17	AI107560	GH05304.3prime GH Dros	2.23e+00
	870	16	5.4	570	17	AI108161	GH06922.5prime GH Dros	2.23e+00
	871	16	5.4	570	17	AI108086	GH06822.5prime GH Dros	2.23e+00
	872	16	5.4	571	18	AI135587	GH13401.3prime GH Dros	2.23e+00
	873	16	5.4	572	27	AQ003448	RPCI11-28N6.TPB RPCI11	2.23e+00
	874	16	5.4	574	14	AA950988	LD31190.5prime LD Dros	2.23e+00
	875	16	5.4	574	15	C80879	Mus musculus 3.5-dpc b	2.23e+00
	876	16	5.4	574	17	AI109455	GH08663.3prime GH Dros	2.23e+00
c	877	16	5.4	574	26	B21842	T3E8-T7 TAMU Arabidops	2.23e+00
	878	16	5.4	575	14	AA933408	SWBmL3SBH06E10SK Brugi	2.23e+00
	879	16	5.4	575	14	AA951665	LD32316.5prime LD Dros	2.23e+00
	880	16	5.4	576	8	AA527856	nh85h01.s1 NCI_CGAP_Br	2.23e+00
	881	16	5.4	576	22	AA164257	zq46g05.s1 Stratagene	2.23e+00
	882	16	5.4	578	12	AA817225	LD22411.5prime LD Dros	2.23e+00
	883	16	5.4	578	12	AA518195	LD18853.3prime LD Dros	2.23e+00
	884	16	5.4	579	17	AI107859	GH05730.3prime GH Dros	2.23e+00
	885	16	5.4	579	17	AI113643	GH10040.3prime GH Dros	2.23e+00
	886	16	5.4	579	17	AI113634	GH10028.3prime GH Dros	2.23e+00
	887	16	5.4	579	17	AI107320	GH04122.3prime GH Dros	2.23e+00
	888	16	5.4	579	17	AI113734	GH10172.3prime GH Dros	2.23e+00
	889	16	5.4	579	17	AI107995	GH05916.3prime GH Dros	2.23e+00
	890	16	5.4	579	18	AI135359	GH13055.3prime GH Dros	2.23e+00
c	891	16	5.4	579	12	AA769084	oa74e01.s1 NCI_CGAP_GC	2.23e+00
	892	16	5.4	580	17	AI109477	GH08691.3prime GH Dros	2.23e+00
	893	16	5.4	580	17	AI109367	GH08530.3prime GH Dros	2.23e+00
	894	16	5.4	581	12	AA735210	LD01608.3prime LD Dros	2.23e+00
c	895	16	5.4	581	12	AA536381	LD16970.5prime LD Dros	2.23e+00
	896	16	5.4	583	17	AI109689	GH08942.3prime GH Dros	2.23e+00
	897	16	5.4	584	17	AI108748	GH07979.3prime GH Dros	2.23e+00
c	898	16	5.4	584	17	AU021616	Mus musculus 8-cell em	2.23e+00
c	899	16	5.4	586	13	AA206016	zq54d02.s1 Stratagene	2.23e+00
	900	16	5.4	587	13	C87571	Mus musculus fertilize	2.23e+00
	901	16	5.4	589	15	C78704	Mus musculus 3.5-dpc b	2.23e+00
	902	16	5.4	589	12	AA264460	LD08979.5prime LD Dros	2.23e+00
c	903	16	5.4	589	28	AQ159435	nbxb0014A01f CUGI Rice	2.23e+00
c	904	16	5.4	590	18	AI179947	EST223678 Normalized r	2.23e+00
c	905	16	5.4	592	27	B79759	CIT-HSP-2050C6.TR CIT-	2.23e+00
c	906	16	5.4	592	18	AI179139	EST222823 Normalized r	2.23e+00
c	907	16	5.4	592	12	AA539630	LD17902.5prime LD Dros	2.23e+00
c	908	16	5.4	593	19	N32579	yw86g06.s1 Homo sapien	2.23e+00
	909	16	5.4	596	17	AU024166	Mus musculus unfertili	2.23e+00

	910	16	5.4	597	20	W02282	za57g07.r1 Soares feta	2.23e+00
c	911	16	5.4	598	15	AI062741	GH02055.5prime GH Dros	2.23e+00
c	912	16	5.4	600	17	AU024513	Mus musculus unfertili	2.23e+00
c	913	16	5.4	601	12	AA440351	LD14927.5prime LD Dros	2.23e+00
	914	16	5.4	603	12	AA816271	LD01124.3prime LD Dros	2.23e+00
	915	16	5.4	603	14	AA990700	LD34530.5prime LD Dros	2.23e+00
c	916	16	5.4	604	12	AA802076	GM03818.5prime GM Dros	2.23e+00
c	917	16	5.4	604	18	AI169293	EST215128 Normalized r	2.23e+00
	918	16	5.4	605	12	AA141585	CK02035.3prime CK Dros	2.23e+00
	919	16	5.4	609	27	FR0031727	Fugu rubripes GSS sequ	2.23e+00
	920	16	5.4	609	28	AQ200546	RPCI11-63P11.TK RPCI11	2.23e+00
	921	16	5.4	611	17	AU024165	Mus musculus unfertili	2.23e+00
	922	16	5.4	612	27	B93170	CIT-HSP-2173L11.TF CIT	2.23e+00
	923	16	5.4	613	27	AQ053633	RPCI11-51M12.TK RPCI11	2.23e+00
	924	16	5.4	615	12	AA817204	LD22380.3prime LD Dros	2.23e+00
	925	16	5.4	616	14	AA979047	LD33295.5prime LD Dros	2.23e+00
	926	16	5.4	618	15	AI062215	GH01349.5prime GH Dros	2.23e+00
c	927	16	5.4	619	12	AA817098	LD22176.5prime LD Dros	2.23e+00
	928	16	5.4	622	14	AA800792	EST190289 Normalized r	2.23e+00
	929	16	5.4	625	15	AI063655	GH03591.3prime GH Dros	2.23e+00
	930	16	5.4	628	28	AQ194816	RPCI11-65P7.TK RPCI11	2.23e+00
	931	16	5.4	630	14	AA979647	LD34350.5prime LD Dros	2.23e+00
	932	16	5.4	632	26	AQ003708	CpG0346B CpIOWAgDNA1 C	2.23e+00
c	933	16	5.4	632	26	AQ003708	CpG0346B CpIOWAgDNA1 C	2.23e+00
c	934	16	5.4	633	18	AI135581	GH13390.5prime GH Dros	2.23e+00
c	935	16	5.4	633	12	AA792252	ve40d11.r1 Soares mous	2.23e+00
c	936	16	5.4	634	15	AI064272	GH04636.5prime GH Dros	2.23e+00
	937	16	5.4	635	15	AI063641	GH03574.3prime GH Dros	2.23e+00
c	938	16	5.4	635	27	AQ002540	CIT-HSP-2283I18.TF CIT	2.23e+00
c	939	16	5.4	635	18	AI135402	GH13125.5prime GH Dros	2.23e+00
c	940	16	5.4	640	9	AA524254	ng34b04.s1 NCI_CGAP_Co	2.23e+00
	941	16	5.4	642	14	AA979037	LD33283.5prime LD Dros	2.23e+00
c	942	16	5.4	643	27	B70078	CIT-HSP-2063G9.TF CIT-	2.23e+00
	943	16	5.4	646	15	AI061965	LD35073.5prime LD Dros	2.23e+00
c	944	16	5.4	646	12	AA440400	LD14987.5prime LD Dros	2.23e+00
c	945	16	5.4	646	26	AQ003639	CpG0296A CpIOWAgDNA1 C	2.23e+00
c	946	16	5.4	646	17	AI110099	GH09561.5prime GH Dros	2.23e+00
	947	16	5.4	647	29	AQ249014	F26K9-Sp6.1 IGF Arabid	2.23e+00
	948	16	5.4	648	15	AI064562	GH05022.3prime GH Dros	2.23e+00
	949	16	5.4	649	12	AA539015	LD18713.5prime LD Dros	2.23e+00
	950	16	5.4	650	12	AA735044	ET2387 Trypanosoma bru	2.23e+00
c	951	16	5.4	652	14	C91169	Dictyostelium discoide	2.23e+00
c	952	16	5.4	656	15	AI001224	MEST5-A5.TW1412.Seq IS	2.23e+00
c	953	16	5.4	657	28	AQ203032	RPCI11-44J20.TJ RPCI11	2.23e+00
	954	16	5.4	659	28	AQ156116	nbxb0002D02f CUGI Rice	2.23e+00
	955	16	5.4	660	26	C11907	Ciona intestinalis gen	2.23e+00
	956	16	5.4	663	26	AQ012872	CIT-HSP-2300O5.TF CIT-	2.23e+00
	957	16	5.4	665	28	AQ157002	nbxb0008F04f CUGI Rice	2.23e+00
	958	16	5.4	667	26	B77182	T31I2TR TAMU Arabidops	2.23e+00
c	959	16	5.4	667	26	B75299	RPCI11-15D7.TP RPCI11	2.23e+00
	960	16	5.4	668	12	AA820547	LD24338.3prime LD Dros	2.23e+00

	961	16	5.4	673	27	AQ051587	nbxb0002cB01f CUGI Ric	2.23e+00
c	962	16	5.4	673	28	AQ194996	RPCI11-43L10.TJ RPCI11	2.23e+00
	963	16	5.4	674	26	B54914	CIT-HSP-384022.TR CIT-	2.23e+00
c	964	16	5.4	674	15	AI062012	GH01049.5prime GH Dros	2.23e+00
	965	16	5.4	675	27	AQ077206	CIT-HSP-2359P3.TR CIT-	2.23e+00
	966	16	5.4	676	15	AI062709	GH02007.5prime GH Dros	2.23e+00
	967	16	5.4	678	22	AA129719	zl16d12.s1 Soares preg	2.23e+00
c	968	16	5.4	680	12	AA816607	LD02659.5prime LD Dros	2.23e+00
c	969	16	5.4	681	8	AA550176	1291m3 gmbPfHB3.1, G.	2.23e+00
c	970	16	5.4	682	27	B65981	CIT-HSP-2023K16.TR CIT	2.23e+00
	971	16	5.4	692	12	AA440502	LD13759.3prime LD Dros	2.23e+00
c	972	16	5.4	696	23	AA221075	mv75c04.r1 Soares mous	2.23e+00
	973	16	5.4	704	12	AA776878	ac40a04.s1 Stratagene	2.23e+00
	974	16	5.4	708	13	AA872732	ob11d07.s1 NCI_CGAP_Ki	2.23e+00
c	975	16	5.4	721	29	AQ254586	CpG0711B CpIOWAgDNA1 C	2.23e+00
c	976	16	5.4	730	26	B00767	cSRL-119h2-u cSRL flow	2.23e+00
c	977	16	5.4	734	14	C90870	Dictyostelium discoide	2.23e+00
c	978	16	5.4	769	26	AG006958	Homo sapiens genomic D	2.23e+00
c	979	16	5.4	771	27	AQ053901	CIT-HSP-2345N21.TR CIT	2.23e+00
c	980	16	5.4	775	26	AG006960	Homo sapiens genomic D	2.23e+00
c	981	16	5.4	779	26	AG006959	Homo sapiens genomic D	2.23e+00
c	982	16	5.4	790	26	B21471	F10A20-Sp6 IGF Arabido	2.23e+00
	983	16	5.4	796	28	AQ156060	nbxb0002M13f CUGI Rice	2.23e+00
	984	16	5.4	799	26	AG006950	Homo sapiens genomic D	2.23e+00
	985	16	5.4	824	26	B08384	F1A17-T7.1 IGF Arabido	2.23e+00
	986	16	5.4	829	26	B21337	T29N16-T7 TAMU Arabido	2.23e+00
c	987	16	5.4	838	22	AA203150	zx53g12.r1 Soares feta	2.23e+00
c	988	16	5.4	855	11	AA697762	HL03118.5prime HL Dros	2.23e+00
	989	16	5.4	856	28	AF011197	Homo sapiens chromosom	2.23e+00
	990	16	5.4	867	26	B03061	cSRL-167C11-u cSRL flo	2.23e+00
	991	16	5.4	881	27	AQ050662	nbxb0004bH11r CUGI Ric	2.23e+00
	992	16	5.4	887	26	B10424	F1A17-T7.2 IGF Arabido	2.23e+00
	993	16	5.4	888	12	AA781144	aj24a09.s1 Soares test	2.23e+00
c	994	16	5.4	911	28	AG014122	Homo sapiens genomic D	2.23e+00
c	995	16	5.4	911	26	AG006133	Homo sapiens genomic D	2.23e+00
	996	16	5.4	923	21	W34252	mb01h02.r1 Soares mous	2.23e+00
	997	16	5.4	931	26	B12525	F1A17-T7 IGF Arabidops	2.23e+00
c	998	16	5.4	1120	26	B11580	F21D10-T7 IGF Arabidop	2.23e+00
	999	16	5.4	1145	26	B08791	F1N21-T7 IGF Arabidops	2.23e+00
	1000	16	5.4	1201	26	B10260	F12K15-Sp6 IGF Arabido	2.23e+00

ALIGNMENTS

```
RESULT    1
LOCUS       HUMGS00628     296 bp    mRNA            EST       17-JUN-1996
DEFINITION  Human HL60 3'directed MboI cDNA, HUMGS00628, clone mp0960, mRNA
            sequence.
ACCESSION   D19673
NID         g499989
KEYWORDS    EST; EST(expressed sequence tag); HL60; gene signature(GS);
```



```
            granulocyte; macrophage; monocyte; promyelocyte.
SOURCE      Homo sapiens adult female promyelocyte cell_line:HL60 cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (sites)
  AUTHORS   Okubo,K., Itoh,K., Fukushima,A., Yoshii,J. and Matsubara,K.
  TITLE     Monitoring cell physiology by expression profiles and discovering
            cell type-specific genes by compiled expression profiles
  JOURNAL   Genomics 30 (2), 178-186 (1995)
  MEDLINE   96163870
REFERENCE   2  (bases 1 to 296)
  AUTHORS   Okubo,K., Fukushima,A., Yoshii,J., Niiyama,T., Kojima,Y.,
            Yoshinari,H., Arimoto,J. and Matsubara,K.
  TITLE     Gene expression of human promyelocytic cell line HL60before and
            after induction of differentiation. A new application of
            3'directedcDNA sequencing
  JOURNAL   Unpublished (1993)
COMMENT     Submitted (10-Sep-1993) to DDBJ by:
            Kousaku Okubo
            Institute for Molecular and Cellular Biology
            Osaka University
            3-1, Yamadaoka
            Suita, Osaka, 565
            Japan
            Phone: 06-877-5111
            Fax  : 06-875-1922.
FEATURES              Location/Qualifiers
     source           1. .296
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /cell_line="HL60"
                      /cell_type="promyelocyte"
                      /dev_stage="adult"
                      /sex="female"
BASE COUNT      102 a     42 c     38 g    114 t
ORIGIN

  Query Match          100.0%;  Score 296;  DB 7;  Length 296;
  Best Local Similarity 100.0%;  Pred. No. 0.00e+00;
  Matches   296;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db       1 GATCTGCCATAAGAAAATCTAGTTCAACTCTAATTTTATGTAGTAAATAAATTGGCAGGT 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy       1 GATCTGCCATAAGAAAATCTAGTTCAACTCTAATTTTATGTAGTAAATAAATTGGCAGGT 60

Db      61 AATTGTTTTTACAAAGAATCCACCTGACTTCCCCTAATGCATTAAAAATATTTTTATTTA 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      61 AATTGTTTTTACAAAGAATCCACCTGACTTCCCCTAATGCATTAAAAATATTTTTATTTA 120
```



```
Db     121 AATAACTTTATTTATAACTTTTAGAAACATGTAGTATTGTTTAAACATCATTTGTTCTTC 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     121 AATAACTTTATTTATAACTTTTAGAAACATGTAGTATTGTTTAAACATCATTTGTTCTTC 180

Db     181 AGTATTTTTCATTTGGAAGTCCAATAGGGCAAATTGAATGAAGTATTATTATCTGTCTCT 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     181 AGTATTTTTCATTTGGAAGTCCAATAGGGCAAATTGAATGAAGTATTATTATCTGTCTCT 240

Db     241 TGTAGTACAATGTATCCAACAGACACTCAATAAACTTTTTGGTTGTTAAACTGAAA 296
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     241 TGTAGTACAATGTATCCAACAGACACTCAATAAACTTTTTGGTTGTTAAACTGAAA 296


RESULT    2
LOCUS       AA906845       413 bp    mRNA            EST       10-JUN-1998
DEFINITION  ol20a09.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
            IMAGE:1523992 3', mRNA sequence.
ACCESSION   AA906845
NID         g3042089
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 413)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1095   Std Error: 0.00
            Seq primer: -40ml3 fwd. ET from Amersham
            High quality sequence stop: 351.
FEATURES              Location/Qualifiers
     source           1. .413
                      /organism="Homo sapiens"
                      /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                      a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                      Equal amounts of plasmid DNA from three normalized
                      libraries (fetal lung NbHL19W, testis NHT, and B-cell
                      NCI_CGAP_GCB1) were mixed, and ss circles were made in
                      vitro. Following HAP purification, this DNA was used as
                      tracer in a subtractive hybridization reaction. The driver
                      was PCR-amplified cDNAs from pools of 5,000 clones made
                      from the same 3 libraries. The pools consisted of
                      I.M.A.G.E. clones 297480-302087, 682632-687239,
```



```
                      726408-728711, and 729096-731399. Subtraction by Bento
                      Soares and M. Fatima Bonaldo."
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1523992"
                      /clone_lib="Soares_NFL_T_GBC_S1"
                      /lab_host="DH10B"
BASE COUNT      155 a     53 c     54 g    151 t
ORIGIN

  Query Match          100.0%;  Score 296;  DB 15;  Length 413;
  Best Local Similarity 100.0%;  Pred. No. 0.00e+00;
  Matches   296;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db       7 TTTCAGTTTAACAACCAAAAAGTTTATTGAGTGTCTGTTGGATACATTGTACTACAAGAG 66
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     296 TTTCAGTTTAACAACCAAAAAGTTTATTGAGTGTCTGTTGGATACATTGTACTACAAGAG 237

Db      67 ACAGATAATAATACTTCATTCAATTTGCCCTATTGGACTTCCAAATGAAAAATACTGAAG 126
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     236 ACAGATAATAATACTTCATTCAATTTGCCCTATTGGACTTCCAAATGAAAAATACTGAAG 177

Db     127 AACAAATGATGTTTAAACAATACTACATGTTTCTAAAAGTTATAAATAAAGTTATTTAAA 186
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     176 AACAAATGATGTTTAAACAATACTACATGTTTCTAAAAGTTATAAATAAAGTTATTTAAA 117

Db     187 TAAAAATATTTTTAATGCATTAGGGGAAGTCAGGTGGATTCTTTGTAAAAACAATTACCT 246
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     116 TAAAAATATTTTTAATGCATTAGGGGAAGTCAGGTGGATTCTTTGTAAAAACAATTACCT 57

Db     247 GCCAATTTATTTACTACATAAAATTAGAGTTGAACTAGATTTTCTTATGGCAGATC 302
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp      56 GCCAATTTATTTACTACATAAAATTAGAGTTGAACTAGATTTTCTTATGGCAGATC 1


RESULT    3
LOCUS       AA836401       447 bp    mRNA            EST       31-MAR-1998
DEFINITION  od37d12.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1370135,
            mRNA sequence.
ACCESSION   AA836401
NID         g2910720
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 447)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
```

```
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
            Ph.D., Gerald Marti, M.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 960   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 418.
FEATURES            Location/Qualifiers
     source         1..447
                    /organism="Homo sapiens"
                    /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                    polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                    was prepared from human tonsillar cells enriched for
                    germinal center B cells by flow sorting (CD20+, IgD-),
                    provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                    (NCI) and Dr. Gerald Marti (CBER).  cDNA synthesis was
                    primed with a Not I - oligo(dT) primer
                    [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTT-
                    3']. Double-stranded cDNA was ligated to Eco RI adaptors
                    (Pharmacia), digested with Not I and cloned into the Not I
                    and Eco RI sites of the modified pT7T3 vector. Library
                    went through one round of normalization, and was
                    constructed by Bento Soares and M. Fatima Bonaldo."
                    /db_xref="taxon:9606"
                    /clone="IMAGE:1370135"
                    /clone_lib="NCI_CGAP_GCB1"
                    /tissue_type="germinal center B cell"
                    /lab_host="DH10B"
BASE COUNT      164 a     55 c     56 g    172 t
ORIGIN

  Query Match          100.0%;  Score 296;  DB 14;  Length 447;
  Best Local Similarity 100.0%;  Pred. No. 0.00e+00;
  Matches   296;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      23 TTTCAGTTTAACAACCAAAAAGTTTATTGAGTGTCTGTTGGATACATTGTACTACAAGAG 82
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     296 TTTCAGTTTAACAACCAAAAAGTTTATTGAGTGTCTGTTGGATACATTGTACTACAAGAG 237

Db      83 ACAGATAATAATACTTCATTCAATTTGCCCTATTGGACTTCCAAATGAAAAATACTGAAG 142
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     236 ACAGATAATAATACTTCATTCAATTTGCCCTATTGGACTTCCAAATGAAAAATACTGAAG 177
```



```
Db     143 AACAAATGATGTTTAAACAATACTACATGTTTCTAAAAGTTATAAATAAAGTTATTTAAA 202
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     176 AACAAATGATGTTTAAACAATACTACATGTTTCTAAAAGTTATAAATAAAGTTATTTAAA 117

Db     203 TAAAAATATTTTTAATGCATTAGGGGAAGTCAGGTGGATTCTTTGTAAAAACAATTACCT 262
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     116 TAAAAATATTTTTAATGCATTAGGGGAAGTCAGGTGGATTCTTTGTAAAAACAATTACCT 57

Db     263 GCCAATTTATTTACTACATAAAATTAGAGTTGAACTAGATTTTCTTATGGCAGATC 318
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp      56 GCCAATTTATTTACTACATAAAATTAGAGTTGAACTAGATTTTCTTATGGCAGATC 1


RESULT    4
LOCUS       AA748329      342 bp    mRNA            EST       23-JAN-1998
DEFINITION  oa54e10.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1308810,
            mRNA sequence.
ACCESSION   AA748329
NID         g2788287
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 342)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
            Ph.D., Gerald Marti, M.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 1525   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 321.
FEATURES            Location/Qualifiers
     source         1..342
                    /organism="Homo sapiens"
                    /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
```



```
                    polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                    was prepared from human tonsillar cells enriched for
                    germinal center B cells by flow sorting (CD20+, IgD-),
                    provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                    (NCI) and Dr. Gerald Marti (CBER).  cDNA synthesis was
                    primed with a Not I - oligo(dT) primer
                    [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTT-
                    3']. Double-stranded cDNA was ligated to Eco RI adaptors
                    (Pharmacia), digested with Not I and cloned into the Not I
                    and Eco RI sites of the modified pT7T3 vector. Library
                    went through one round of normalization, and was
                    constructed by Bento Soares and M. Fatima Bonaldo."
                    /db_xref="taxon:9606"
                    /clone="IMAGE:1308810"
                    /clone_lib="NCI_CGAP_GCB1"
                    /tissue_type="germinal center B cell"
                    /lab_host="DH10B"
BASE COUNT      128 a     42 c     46 g    126 t
ORIGIN

  Query Match          98.0%;  Score 290;  DB 12;  Length 342;
  Best Local Similarity 100.0%;  Pred. No. 0.00e+00;
  Matches   290;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db       8 TTTAACAACCAAAAAGTTTATTGAGTGTCTGTTGGATACATTGTACTACAAGAGACAGAT 67
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     290 TTTAACAACCAAAAAGTTTATTGAGTGTCTGTTGGATACATTGTACTACAAGAGACAGAT 231

Db      68 AATAATACTTCATTCAATTTGCCCTATTGGACTTCCAAATGAAAAATACTGAAGAACAAA 127
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     230 AATAATACTTCATTCAATTTGCCCTATTGGACTTCCAAATGAAAAATACTGAAGAACAAA 171

Db     128 TGATGTTTAAACAATACTACATGTTTCTAAAAGTTATAAATAAAGTTATTTAAATAAAAA 187
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     170 TGATGTTTAAACAATACTACATGTTTCTAAAAGTTATAAATAAAGTTATTTAAATAAAAA 111

Db     188 TATTTTTAATGCATTAGGGGAAGTCAGGTGGATTCTTTGTAAAAACAATTACCTGCCAAT 247
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     110 TATTTTTAATGCATTAGGGGAAGTCAGGTGGATTCTTTGTAAAAACAATTACCTGCCAAT 51

Db     248 TTATTTACTACATAAAATTAGAGTTGAACTAGATTTTCTTATGGCAGATC 297
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Cp      50 TTATTTACTACATAAAATTAGAGTTGAACTAGATTTTCTTATGGCAGATC 1


RESULT    5
LOCUS       AA808708      403 bp    mRNA            EST       12-FEB-1998
DEFINITION  oa47d05.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1308105,
            mRNA sequence.
ACCESSION   AA808708
NID         g2878114
```



```
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 403)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
            Ph.D., Gerald Marti, M.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 394.
FEATURES            Location/Qualifiers
     source         1..403
                    /organism="Homo sapiens"
                    /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                    polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                    was prepared from human tonsillar cells enriched for
                    germinal center B cells by flow sorting (CD20+, IgD-),
                    provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                    (NCI) and Dr. Gerald Marti (CBER).  cDNA synthesis was
                    primed with a Not I - oligo(dT) primer
                    [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTT-
                    3']. Double-stranded cDNA was ligated to Eco RI adaptors
                    (Pharmacia), digested with Not I and cloned into the Not I
                    and Eco RI sites of the modified pT7T3 vector. Library
                    went through one round of normalization, and was
                    constructed by Bento Soares and M. Fatima Bonaldo."
                    /db_xref="taxon:9606"
                    /clone="IMAGE:1308105"
                    /clone_lib="NCI_CGAP_GCB1"
                    /tissue_type="germinal center B cell"
                    /lab_host="DH10B"
BASE COUNT      145 a     58 c     53 g    147 t
ORIGIN

  Query Match          73.0%;  Score 216;  DB 12;  Length 403;
```

```
  Best Local Similarity  99.3%;  Pred. No. 0.00e+00;
  Matches   276;  Conservative    0;  Mismatches   2;  Indels   0;  Gaps   0;

Db      36 AAAGTTTATTGAGTGTCTGTTGGATACATTGTACTACAAGAGACAGATAATAATACTTCA 95
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     278 AAAGTTTATTGAGTGTCTGTTGGATACATTGTACTACAAGAGACAGATAATAATACTTCA 219

Db      96 TTCAATTTGCCCTATTGGACTTCCAAATGAAAAATACTGAAGAACAAATGATGTTTAAAC 155
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp     218 TTCAATTTGCCCTATTGGACTTCCAAATGAAAAATACTGAAGAACAAATGATGTTTAAAC 159

Db     156 AATACTACATGTTTCTAAAAGTTATACATAAAGTTATTTAAATACAAATATTTTTAATGC 215
           ||||||||||||||||||||||||| |||||||||||||||| ||||||||||||||||
Cp     158 AATACTACATGTTTCTAAAAGTTATAAATAAAGTTATTTAAATAAAAATATTTTTAATGC 99

Db     216 ATTAGGGGAAGTCAGGTGGATTCTTTGTAAAAACAATTACCTGCCAATTTATTTACTACA 275
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp      98 ATTAGGGGAAGTCAGGTGGATTCTTTGTAAAAACAATTACCTGCCAATTTATTTACTACA 39

Db     276 TAAAATTAGAGTTGAACTAGATTTTCTTATGGCAGATC 313
           |||||||||||||||||||||||||||||||||||||
Cp      38 TAAAATTAGAGTTGAACTAGATTTTCTTATGGCAGATC 1
```

```
RESULT    6
LOCUS       T81784         116 bp    mRNA            EST       15-MAR-1995
DEFINITION  yd30d06.s1 Homo sapiens cDNA clone 109739 3' similar to gb:S52450
            NONSPECIFIC LIPID-TRANSFER PROTEIN PRECURSOR (HUMAN);.
ACCESSION   T81784
NID         g704791
KEYWORDS    EST.
SOURCE      human clone=109739 library=Soares fetal liver spleen 1NFLS
            vector=pT7T3D (Pharmacia) with a modified polylinker host=DH10B
            (ampicillin resistant) primer=T3 Rsite1=Pac I Rsite2=Eco RI Liver
            and spleen from a 20 week-post conception male fetus. 1st strand
            cDNA was primed with a Pac I - oligo(dT) primer [5'
            AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTT 3'], double-stranded
            cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac
            I and cloned into the Pac I and Eco RI sites of the modified pT7T3
            vector.  Library went through one round of normalization. Library
            constructed by Bento Soares and M.Fatima Bonaldo.
  ORGANISM  Homo sapiens
            Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 116)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
```



```
  JOURNAL   Unpublished (1995)
COMMENT     Other_ESTs: yd30d06.r1
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence starts: 1
            High quality sequence stops: 1
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Trace considered overall poor quality.
FEATURES              Location/Qualifiers
     source           1. .116
                      /organism="Homo sapiens"
                      /clone="109739"
BASE COUNT       41 a      13 c      16 g      41 t       5 others
ORIGIN

  Query Match           6.8%;  Score 20;  DB 5;  Length 116;
  Best Local Similarity  95.2%;  Pred. No. 1.36e-05;
  Matches    20;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db      16 TTAAAAATATTTTTATTTNAA 36
           |||||||||||||||||| ||
Qy     102 TTAAAAATATTTTTATTTAAA 122


RESULT    7
LOCUS       AA547753       163 bp    mRNA            EST       06-AUG-1997
DEFINITION  EST188662 B. glabrata adult whole body Biomphalaria glabrata cDNA
            5' similar to ribosomal protein L13, mRNA sequence.
ACCESSION   AA547753
NID         g2314897
KEYWORDS    EST.
SOURCE      bloodfluke planorb.
  ORGANISM  Biomphalaria glabrata
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Mollusca;
            Gastropoda; Pulmonata; Basommatophora; Planorbidae; Biomphalaria.
REFERENCE   1  (bases 1 to 163)
  AUTHORS   Knight,M., Miller,A.N., Geoghagen,N.S.M., Lewis,F.A. and
            Kerlavag,A.R.
  TITLE     Expressed Sequence Tags (ESTs) of Biomphalaria glabrata, an
            intermediate snail host of Schistosoma mansoni: Use in the
            identification of RFLP markers
  JOURNAL   Malacologia (1997) In press
COMMENT
            Contact: Kerlavage, AR
```



```
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlave@tigr.org
            Seq primer: M13 Reverse.
FEATURES              Location/Qualifiers
     source           1. .163
                      /organism="Biomphalaria glabrata"
                      /strain="BS-90"
                      /note="Organ: whole body; Vector: pBluescript SK-; Site_1:
                      EcoRI; Site_2: XhoI; poly(A)+ RNA was purified from adult
                      whole snails.  cDNA was constructed using an oligo-dT
                      primer and directionally cloned using the Lambda ZAP II
                      Vector Kit by Stratagene."
                      /db_xref="taxon:6526"
                      /clone_lib="B. glabrata adult whole body"
                      /tissue_type="whole body"
     mRNA             <1. .>163
BASE COUNT       39 a      40 c      23 g      59 t       2 others
ORIGIN

  Query Match           6.8%;  Score 20;  DB 8;  Length 163;
  Best Local Similarity 100.0%;  Pred. No. 1.36e-05;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;.

Db      10 AAAAATATTTTTATTTAAAT 29
           ||||||||||||||||||||
Qy     104 AAAAATATTTTTATTTAAAT 123


RESULT    8
LOCUS       AA186021       589 bp    mRNA            EST       17-FEB-1997
DEFINITION  mt35d04.r1 Soares mouse 3NbMS Mus musculus cDNA clone 623047 5'
            similar to SW:RU2B_HUMAN P08579 U2 SMALL NUCLEAR RIBONUCLEOPROTEIN
            B ;, mRNA sequence.
ACCESSION   AA186021
NID         g1772162
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Mus.
REFERENCE   1  (bases 1 to 589)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
```



```
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:383871
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 494.
FEATURES              Location/Qualifiers
     source           1. .589
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /note="Vector: pT7T3D-Pac.(Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                      was primed with a Not I - oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCGCTGTTTTTTTTTTTTTTTTTTTT
                      3']; double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Not I and cloned into the Not I
                      and Eco RI sites of the modified pT7T3 vector. RNA
                      provided by Dr. Bertrand Jordan. Library went through
                      three rounds of normalization, and was constructed by
                      Bento Soares and M.Fatima Bonaldo."
                      /db_xref="taxon:10090"
                      /clone="623047"
                      /clone_lib="Soares mouse 3NbMS"
                      /sex="male"
                      /tissue_type="Spleen"
                      /dev_stage="4 weeks"
                      /lab_host="DH10B"
     mRNA             <1. .>589
BASE COUNT      200 a      87 c     116 g     186 t
ORIGIN

  Query Match           6.8%;  Score 20;  DB 23;  Length 589;
  Best Local Similarity 100.0%;  Pred. No. 1.36e-05;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     202 TTGGACTTCCAAATGAAAAA 221
           ||||||||||||||||||||
Cp     204 TTGGACTTCCAAATGAAAAA 185


RESULT    9
LOCUS       B53879         625 bp    DNA             GSS       19-JUN-1998
```

```
DEFINITION  CIT-HSP-2013H3.TF CIT-HSP Homo sapiens genomic clone 2013H3,
            genomic survey sequence.
ACCESSION   B53879
NID         g2608213
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 625)
  AUTHORS   Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K.,
            Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
            Simon,M. and Venter,J.C.
  TITLE     Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: CIT-HSP-2013H3.TR
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: M13-21
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .625
                     /organism="Homo sapiens"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII"
                     /db_xref="taxon:9606"
                     /clone="2013H3"
                     /clone_lib="CIT-HSP"
                     /sex="Male"
                     /cell_type="Sperm"
BASE COUNT      240 a      71 c      63 g     251 t
ORIGIN

  Query Match             6.8%;  Score 20;  DB 26;  Length 625;
  Best Local Similarity 100.0%;  Pred. No. 1.36e-05;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      79 AAAATATTTTTATTTAAATA 98
           ||||||||||||||||||||
Qy     105 AAAATATTTTTATTTAAATA 124
```



```
RESULT   10
LOCUS       W80705         181 bp    mRNA            EST       07-MAY-1997
DEFINITION  zh51a06.s1 Soares fetal liver spleen 1NFLS S1 Homo sapiens cDNA
            clone 415570 3', mRNA sequence.
ACCESSION   W80705
NID         g1391723
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 181)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
            Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 314   Std Error: 0.00
            Seq primer: mob.REGA+ET.
FEATURES             Location/Qualifiers
     source          1. .181
                     /organism="Homo sapiens"
                     /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                     with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                     This is a subtracted version of the original Soares fetal
                     liver spleen 1NFLS library.   1st strand cDNA was primed
                     with a Pac I - oligo(dT) primer [5'
                     AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Pac I and cloned into the Pac I
                     and Eco RI sites of the modified pT7T3 vector.  Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
                     /db_xref="taxon:9606"
                     /clone="415570"
                     /clone_lib="Soares fetal liver spleen 1NFLS S1"
                     /sex="male"
                     /dev_stage="20 week-post conception fetus"
                     /lab_host="DH10B (ampicillin resistant)"
```



```
     mRNA            complement(<1. .>181)
                     /db_xref="GDB:1324040"
BASE COUNT       41 a      29 c      51 g      60 t
ORIGIN

  Query Match             6.4%;  Score 19;  DB 6;  Length 181;
  Best Local Similarity 100.0%;  Pred. No. 3.32e-04;
  Matches    19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      75 TTATCTGTCTCTTGTAGTA 93
           |||||||||||||||||||
Qy     229 TTATCTGTCTCTTGTAGTA 247


RESULT   11
LOCUS       AA678036       254 bp    mRNA            EST       19-DEC-1997
DEFINITION  zi25d03.s1 Soares fetal liver spleen 1NFLS S1 Homo sapiens cDNA
            clone 431813 3', mRNA sequence.
ACCESSION   AA678036
NID         g2658558
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 254)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
            Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
            Theising,B., White,T., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 218.
FEATURES             Location/Qualifiers
     source          1. .254
                     /organism="Homo sapiens"
                     /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                     with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                     This is a subtracted version of the original Soares fetal
                     liver spleen 1NFLS library.   1st strand cDNA was primed
                     with a Pac I - oligo(dT) primer [5'
```



```
                     AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Pac I and cloned into the Pac I
                     and Eco RI sites of the modified pT7T3 vector.  Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
                     /db_xref="taxon:9606"
                     /clone="431813"
                     /clone_lib="Soares fetal liver spleen 1NFLS S1"
                     /sex="male"
                     /dev_stage="20 week-post conception fetus"
                     /lab_host="DH10B (ampicillin resistant)"
BASE COUNT       57 a      40 c      69 g      88 t
ORIGIN

  Query Match             6.4%;  Score 19;  DB 11;  Length 254;
  Best Local Similarity 100.0%;  Pred. No. 3.32e-04;
  Matches    19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      74 TTATCTGTCTCTTGTAGTA 92
           |||||||||||||||||||
Qy     229 TTATCTGTCTCTTGTAGTA 247


RESULT   12
LOCUS       R81084         273 bp    mRNA            EST       12-JUN-1996
DEFINITION  EST22b023 WATM1 Homo sapiens cDNA clone 22b023, mRNA sequence.
ACCESSION   R81084
NID         g857687
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 273)
  AUTHORS   Bouillaud,F.
  TITLE     Study of expressed sequences tags in adipose tissue 1995
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Frederic Bouillaud
            Centre de Recherche sur l'Endocrinologie moleculaire et le
            Developpement
            CNRS
            9, Rue Jules Hetzel, Meudon Bellevue, 92190 France
            Tel: 33 1 45 07 52 87
            Fax: 33 1 45 07 58 90
            Email: bouillau@infobiogen.fr
            Southern : single band human DNA EcoRI 5.4 Kb.
FEATURES             Location/Qualifiers
     source          1. .273
                     /organism="Homo sapiens"
```

```
                     /strain="caucasian"
                     /note="Vector: lambda gt11; Site_1: EcoRI; White adipose
                     tissue, sub cutaneous, adult, female. Purification of
                     polyA mRNA, first strand priming with random
                     oligonucleotides. Ligation of adaptors EcoRI/NotI, cloning
                     in EcoRI site of lambda gt11."
                     /db_xref="taxon:9606"
                     /clone="22b023"
                     /clone_lib="WATM1"
     mRNA            <1. .>273
BASE COUNT       83 a     56 c     45 g     89 t
ORIGIN

  Query Match           6.4%;  Score 19;  DB 20;  Length 273;
  Best Local Similarity 100.0%;  Pred. No. 3.32e-04;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    212 TTAAAAATATTTTTATTTA 230
          ||||||||||||||||||||
Qy    102 TTAAAAATATTTTTATTTA 120


RESULT   13
LOCUS       AA064804       289 bp    mRNA            EST       23-DEC-1997
DEFINITION  sm50a08.s1 Stratagene fibroblast (#937212) Homo sapiens cDNA clone
            529046 3', mRNA sequence.
ACCESSION   AA064804
NID         g1558952
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 289)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
            Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            WARNING: There is evidence that suggests that the 384-well parent
            plate of this clone contains both human and mouse derived clones.
            Thus, the origin of this clone is uncertain. This caution should be
            kept in mind should you use this clone.
```



```
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40M13 fwd. from Amersham
            High quality sequence stop: 164.
FEATURES             Location/Qualifiers
     source          1. .289
                     /organism="Homo sapiens"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; Cloned unidirectionally.  Primer: Oligo dT. W138
                     cell line. Average insert size: 0.8 kb; Uni-ZAP XR Vector;
                     -5' adaptor sequence: 5' GAATTCGGCACGAG 3' -3' adaptor
                     sequence: 5' CTCGAGTTTTTTTTTTTTTTTTT 3'"
                     /db_xref="GDB:3918557"
                     /db_xref="taxon:9606"
                     /clone="529046"
                     /clone_lib="Stratagene fibroblast (#937212)"
                     /lab_host="SOLR cells (kanamycin resistant)"
BASE COUNT       62 a     45 c     78 g    101 t      3 others
ORIGIN

  Query Match           6.4%;  Score 19;  DB 11;  Length 289;
  Best Local Similarity 100.0%;  Pred. No. 3.32e-04;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     84 TTATCTGTCTCTTGTAGTA 102
          ||||||||||||||||||||
Qy    229 TTATCTGTCTCTTGTAGTA 247


RESULT   14
LOCUS       R69529         293 bp    mRNA            EST       01-JUN-1995
DEFINITION  yj82c11.s1 Homo sapiens cDNA clone 155252 3'.
ACCESSION   R69529
NID         g843046
KEYWORDS    EST.
SOURCE      human clone=155252 library=Soares breast 2NbHBst vector=pT7T3D
            (Pharmacia) with a modified polylinker host=DH10B (ampicillin
            resistant) primer=Promega -21m13 Rsite1=Not I Rsite2=Eco RI Adult
            female. 1st strand cDNA was primed with a Not I - oligo(dT) primer
            [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTT 3'),
            double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia),
            digested with Not I and cloned into the Not I and Eco RI sites of a
            modified pT7T3 vector (Pharmacia). Library went through one round
            of normalization to a Cot = 230. Library constructed by Bento
            Soares and M.Fatima Bonaldo.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
```



```
REFERENCE   1  (bases 1 to 293)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 185
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES             Location/Qualifiers
     source          1. .293
                     /organism="Homo sapiens"
                     /clone="155252"
BASE COUNT       63 a     46 c     76 g    105 t      3 others
ORIGIN

  Query Match           6.4%;  Score 19;  DB 16;  Length 293;
  Best Local Similarity 100.0%;  Pred. No. 3.32e-04;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     41 TTATCTGTCTCTTGTAGTA 59
          ||||||||||||||||||||
Qy    229 TTATCTGTCTCTTGTAGTA 247


RESULT   15
LOCUS       AA262768       294 bp    mRNA            EST       13-AUG-1997
DEFINITION  zs24d12.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:686135 3',
            mRNA sequence.
ACCESSION   AA262768
NID         g1898323
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 294)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
```



```
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -41m13 fwd. ET from Amersham
            High quality sequence stop: 222.
FEATURES             Location/Qualifiers
     source          1. .294
                     /organism="Homo sapiens"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from human tonsillar cells enriched for
                     germinal center B cells by flow sorting (CD20+, IgD-),
                     provided by Dr. Louis M. Staudt (NCI). Dr. David Allman
                     (NCI) and Dr. Gerald Marti (CBER).  cDNA synthesis was
                     primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTT-
                     3']. Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo."
                     /db_xref="taxon:9606"
                     /clone="IMAGE:686135"
                     /clone_lib="NCI_CGAP_GCB1"
                     /tissue_type="germinal center B cell"
                     /lab_host="DH10B"
     mRNA            complement(<1. .>294)
BASE COUNT       64 a     45 c     80 g    105 t
ORIGIN

  Query Match           6.4%;  Score 19;  DB 8;  Length 294;
  Best Local Similarity 100.0%;  Pred. No. 3.32e-04;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     82 TTATCTGTCTCTTGTAGTA 100
          ||||||||||||||||||||
Qy    229 TTATCTGTCTCTTGTAGTA 247


RESULT   16
LOCUS       AA845909       307 bp    mRNA            EST       04-MAR-1998
DEFINITION  ak86b11.s1 Barstead spleen HPLRB2 Homo sapiens cDNA clone
            IMAGE:1414749 3', mRNA sequence.
ACCESSION   AA845909
NID         g2932049
KEYWORDS    EST.
```

```
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 307)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
            Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
            Theising,B., White,T., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40m13 fwd. ET from Amersham.
FEATURES              Location/Qualifiers
     source           1. .307
                      /organism="Homo sapiens"
                      /note="Organ: spleen; Vector: pT7T3D-Pac (Pharmacia) with
                      a modified polylinker; Site_1: EcoRI; Site_2: NotI; 1st
                      strand cDNA was primed with a Not I - oligo(dT) primer [5'
                      TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTTTTTT
                      3']; double-stranded cDNA was ligated to Eco RI adaptors
                      [AATTCGGATCCTTG], digested with Not I and cloned into the
                      Not I and Eco RI sites of the modified pT7T3 vector.
                      Library constructed by Bob Barstead."
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1414749"
                      /clone_lib="Barstead spleen HPLRB2"
                      /sex="male"
                      /dev_stage="adult, 17 years"
                      /lab_host="DH10B"
BASE COUNT       66 a     46 c     84 g    111 t
ORIGIN

  Query Match             6.4%;  Score 19;  DB 13;  Length 307;
  Best Local Similarity 100.0%;  Pred. No. 3.32e-04;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     86 TTATCTGTCTCTTGTAGTA 104
          ||||||||||||||||||
Qy    229 TTATCTGTCTCTTGTAGTA 247


RESULT   17
```



```
LOCUS       AA887742      318 bp    mRNA            EST       07-APR-1998
DEFINITION  nq97f10.s1 NCI_CGAP_Co10 Homo sapiens cDNA clone IMAGE:1160299 3',
            mRNA sequence.
ACCESSION   AA887742
NID         g1003417
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 318)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck,
            M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 289   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 274.
FEATURES              Location/Qualifiers
     source           1. .318
                      /organism="Homo sapiens"
                      /note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a
                      modified polylinker; 1st strand cDNA was prepared from
                      RER+ colon tumor, and was then primed with a Not I -
                      oligo(dT) primer. Double-stranded cDNA was ligated to Eco
                      RI adaptors (Pharmacia), digested with Not I and cloned
                      into the Not I and Eco RI sites of the modified pT7T3
                      vector.  Library is normalized. Library was constructed by
                      Bento Soares and M. Fatima Bonaldo (N-Soares4)."
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1160299"
                      /clone_lib="NCI_CGAP_Co10"
                      /tissue_type="colon tumor RER+"
                      /lab_host="DH10B"
BASE COUNT       68 a     50 c     84 g    116 t
ORIGIN

  Query Match             6.4%;  Score 19;  DB 14;  Length 318;
```



```
  Best Local Similarity 100.0%;  Pred. No. 3.32e-04;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     78 TTATCTGTCTCTTGTAGTA 96
          ||||||||||||||||||
Qy    229 TTATCTGTCTCTTGTAGTA 247


RESULT   18
LOCUS       AQ042161      321 bp    DNA             GSS       13-JUL-1998
DEFINITION  CIT-HSP-2330A19.TR CIT-HSP Homo sapiens genomic clone 2330A19,
            genomic survey sequence.
ACCESSION   AQ042161
NID         g3311355
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 321)
  AUTHORS   Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K.,
            Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
            Simon,M. and Venter,J.C.
  TITLE     Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: CIT-HSP-2330A19.TF
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13 Reverse
            Class: BAC ends.
FEATURES              Location/Qualifiers
     source           1. .321
                      /organism="Homo sapiens"
                      /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                      HindIII"
                      /db_xref="taxon:9606"
                      /clone="2330A19"
                      /clone_lib="CIT-HSP"
                      /sex="Male"
                      /cell_type="Sperm"
BASE COUNT       90 a     70 c     46 g    115 t
ORIGIN
```



```
  Query Match             6.4%;  Score 19;  DB 27;  Length 321;
  Best Local Similarity 100.0%;  Pred. No. 3.32e-04;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    191 AAATATTTTTATTTAAATA 209
          ||||||||||||||||||
Qy    106 AAATATTTTTATTTAAATA 124


RESULT   19
LOCUS       AA634259      334 bp    mRNA            EST       06-MAR-1998
DEFINITION  ac75g01.s1 Stratagene lung (#937210) Homo sapiens cDNA clone 868464
            3', mRNA sequence.
ACCESSION   AA634259
NID         g2557473
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 334)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
            Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
            Theising,B., White,T., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 274   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham.
FEATURES              Location/Qualifiers
     source           1. .334
                      /organism="Homo sapiens"
                      /note="Organ: lung; Vector: pBluescript SK-; Site_1:
                      EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                      Oligo dT  normal lung.  Average insert size: 1.0 Kb;
                      Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCACGAG
                      3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTTT 3'"
                      /db_xref="taxon:9606"
                      /clone="868464"
                      /clone_lib="Stratagene lung (#937210)"
                      /sex="male"
```

```
                    /dev_stage="72 years"
                    /lab_host="SOLR cells (kanamycin resistant)"
  BASE COUNT       71 a      54 c      86 g     123 t
  ORIGIN

    Query Match           6.4%;  Score 19;  DB 13;  Length 334;
    Best Local Similarity 100.0%;  Pred. No. 3.32e-04;
    Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

  Db      82 TTATCTGTCTCTTGTAGTA 100
             |||||||||||||||||||
  Qy     229 TTATCTGTCTCTTGTAGTA 247


  RESULT   20
  LOCUS       T09855        334 bp    DNA              EST       29-NOV-1993
  DEFINITION  0428m7 Plasmodium falciparum genomic clone 0428m.
  ACCESSION   T09855
  NID         g319687
  KEYWORDS    EST.
  SOURCE      Malaria, falciparum clone=0428m library=gmbPfHB3.1, G. Roman Reddy
              vector=pBlueScript SK(+) host=E. coli XL1-Blue primer=T7 Genomic
              DNA, from asynchronous blood stage parasites of the cloned Honduran
              HB3 isolate cultured in vitro, was digested with mung bean nuclease
              in the presence of 30% formamide at 50oC (Vernick, K.D., Imberski,
              R.B., and McCutchan, T.F. 1988. Nucleic Acids Research
              16:6883-6896).  The ends of the fragments were polished using T4
              DNA polymerase, and the fragments were ligated to EcoR V-cleaved
              and dephosphorylated pBlueScript SK(+).  Recombinant plasmids
              transformed E. coli XL1-Blue.
    ORGANISM  Plasmodium falciparum
              Eucaryotae; Protozoa; Apicomplexa; Sporozoa; Coccidia;
              Eucoccidiida; Haemosporina; Plasmodium.
  REFERENCE   1  (bases 1 to 334)
    AUTHORS   Reddy,G.R., Chakrabarti,D., Schuster,S.M., Perl,R.J., Almira,E.C.
              and Dame,J.B.
    TITLE     Gene sequence tags from Plasmodium falciparum genomic DNA fragments
              prepared by the genease activity of mung bean nuclease
    JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 90, 9867-9871 (1993)
  COMMENT     *Other_ESTs: 0428m3
              Contact: Dame JB
              Department of Infectious Diseases
              University of Florida
              Bldg. 471 Mowry Rd., Gainesville, FL  32611-0880
              Tel: 9043924700
              Fax: 9043929704
              Email: dame@icbr.ifas.ufl.edu.
  FEATURES              Location/Qualifiers
       source           1. .334
                        /organism="Plasmodium falciparum"
                        /clone="0428m"
```



```
  BASE COUNT      122 a     26 c     33 g    146 t      7 others
  ORIGIN

    Query Match           6.4%;  Score 19;  DB 5;  Length 334;
    Best Local Similarity 100.0%;  Pred. No. 3.32e-04;
    Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

  Db     246 ATTAAAAATATTTTTATTT 264
             |||||||||||||||||||
  Qy     101 ATTAAAAATATTTTTATTT 119


  RESULT   21
  LOCUS       AA083747      339 bp    mRNA             EST       23-DEC-1997
  DEFINITION  zm62b12.r1 Stratagene fibroblast (#937212) Homo sapiens cDNA clone
              530207 5', mRNA sequence.
  ACCESSION   AA083747
  NID         g1625945
  KEYWORDS    EST.
  SOURCE      human.
    ORGANISM  Homo sapiens
              Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
              Primates; Catarrhini; Hominidae; Homo.
  REFERENCE   1  (bases 1 to 339)
    AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
              Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
              Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
              Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
    TITLE     WashU-Merck EST Project
    JOURNAL   Unpublished (1995)
  COMMENT     Contact: Wilson RK
              Washington University School of Medicine
              4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
              Tel: 314 286 1800
              Fax: 314 286 1810
              Email: est@watson.wustl.edu
              WARNING: There is evidence that suggests that the 384-well parent
              plate of this clone contains both human and mouse derived clones.
              Thus, the origin of this clone is uncertain. This caution should be
              kept in mind should you use this clone.

              This clone is available royalty-free through LLNL ; contact the
              IMAGE Consortium (info@image.llnl.gov) for further information.
              Seq primer: -28M13 rev2 from Amersham
              High quality sequence stop: 311.
  FEATURES              Location/Qualifiers
       source           1. .339
                        /organism="Homo sapiens"
                        /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                        XhoI; Cloned unidirectionally.  Primer: Oligo dT. W138
```



```
                    cell line. Average insert size: 0.8 kb; Uni-ZAP XR Vector;
                    -5' adaptor sequence: 5' GAATTCGGCACGAG 3' -3' adaptor
                    sequence: 5' CTCGAGTTTTTTTTTTTTTTTTT 3'"
                    /db_xref="GDB:391971l8"
                    /db_xref="taxon:9606"
                    /clone="530207"
                    /clone_lib="Stratagene fibroblast (#937212)"
                    /lab_host="SOLR cells (kanamycin resistant)"
  BASE COUNT      123 a     88 c     56 g     72 t
  ORIGIN

    Query Match           6.4%;  Score 19;  DB 11;  Length 339;
    Best Local Similarity 100.0%;  Pred. No. 3.32e-04;
    Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

  Db     244 TACTACAAGAGACAGATAA 262
             |||||||||||||||||||
  Cp     247 TACTACAAGAGACAGATAA 229


  RESULT   22
  LOCUS       AA960086      344 bp    mRNA             EST       08-MAY-1998
  DEFINITION  vw59f12.s1 Soares mouse mammary gland NMLMG Mus musculus cDNA clone
              1248143 3', mRNA sequence.
  ACCESSION   AA960086
  NID         g3125986
  KEYWORDS    EST.
  SOURCE      house mouse.
    ORGANISM  Mus musculus
              Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
              Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  REFERENCE   1  (bases 1 to 344)
    AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
              Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
              Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
              Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
              Waterston,R.
    TITLE     The WashU-HHMI Mouse EST Project
    JOURNAL   Unpublished (1996)
  COMMENT     Contact: Marra M/Mouse EST Project
              WashU-HHMI Mouse EST Project
              Washington University School of MedicineP
              4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
              Tel: 314 286 1800
              Fax: 314 286 1810
              Email: mouseest@watson.wustl.edu
              This clone is available royalty-free through LLNL ; contact the
              IMAGE Consortium (info@image.llnl.gov) for further information.
              MGI:661831
              Seq primer: -28m13 rev2 ET from Amersham
```



```
              High quality sequence stop: 330.
  FEATURES              Location/Qualifiers
       source           1. .344
                        /organism="Mus musculus"
                        /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                        polylinker; 1st strand cDNA was prepared from mammary
                        gland tissue from a lactating female, and was then primed
                        with a Not I - oligo(dT) primer. Double-stranded  cDNA was
                        ligated to Eco RI adaptors (Pharmacia), digested with Not
                        I  and cloned into the Not I and Eco RI sites of the
                        modified pT7T3 vector.  Library is normalized.  Library
                        was constructed by Bento Soares and M. Fatima Bonaldo."
                        /db_xref="taxon:10090"
                        /clone="1248143"
                        /clone_lib="Soares mouse mammary gland NMLMG"
                        /sex="female (lactating)"
                        /tissue_type="mammary gland"
                        /lab_host="DH10B"
  BASE COUNT      147 a     63 c     50 g     84 t
  ORIGIN

    Query Match           6.4%;  Score 19;  DB 14;  Length 344;
    Best Local Similarity 100.0%;  Pred. No. 3.32e-04;
    Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

  Db     230 TAAATAAAAATATTTTTAA 248
             |||||||||||||||||||
  Cp     120 TAAATAAAAATATTTTTAA 102


  RESULT   23
  LOCUS       AA112884      350 bp    mRNA             EST       23-DEC-1997
  DEFINITION  zm62b12.s1 Stratagene fibroblast (#937212) Homo sapiens cDNA clone
              530207 3', mRNA sequence.
  ACCESSION   AA112884
  NID         g1664235
  KEYWORDS    EST.
  SOURCE      human.
    ORGANISM  Homo sapiens
              Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
              Primates; Catarrhini; Hominidae; Homo.
  REFERENCE   1  (bases 1 to 350)
    AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
              Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
              Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
              Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
    TITLE     WashU-Merck EST Project
    JOURNAL   Unpublished (1995)
  COMMENT     Contact: Wilson RK
              Washington University School of Medicine
```

```
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            WARNING: There is evidence that suggests that the 384-well parent
            plate of this clone contains both human and mouse derived clones.
            Thus, the origin of this clone is uncertain. This caution should be
            kept in mind should you use this clone.

            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40M13 fwd. from Amersham
            High quality sequence stop: 287.
FEATURES            Location/Qualifiers
     source         1. .350
                    /organism="Homo sapiens"
                    /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                    XhoI; Cloned unidirectionally.  Primer: Oligo dT. WI38
                    cell line. Average insert size: 0.8 kb; Uni-ZAP XR Vector;
                    -5' adaptor sequence: 5' GAATTCGGCACGAG 3' -3' adaptor
                    sequence: 5' CTCGAGTTTTTTTTTTTTTTTTT 3'"
                    /db_xref="GDB:3919718"
                    /db_xref="taxon:9606"
                    /clone="530207"
                    /clone_lib="Stratagene fibroblast (#937212)"
                    /lab_host="SOLR cells (kanamycin resistant)"
BASE COUNT       72 a     58 c     91 g    129 t
ORIGIN

  Query Match           6.4%;  Score 19;  DB 11;  Length 350;
  Best Local Similarity 100.0%;  Pred. No. 3.32e-04;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      82 TTATCTGTCTCTTGTAGTA 100
           ||||||||||||||||||||
Qy     229 TTATCTGTCTCTTGTAGTA 247


RESULT   24
LOCUS       AA991247      353 bp    mRNA            EST       03-JUN-1998
DEFINITION  os40h07.s1 NCI_CGAP_Br2 Homo sapiens cDNA clone IMAGE:1607869 3',
            mRNA sequence.
ACCESSION   AA991247
NID         g1177736
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 353)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
```



```
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 342.
FEATURES            Location/Qualifiers
     source         1. .353
                    /organism="Homo sapiens"
                    /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                    polylinker; 1st strand cDNA was prepared from pooled bulk
                    breast tumor tissue, and was then primed with a Not I -
                    oligo(dT) primer. Double-stranded cDNA was ligated to Eco
                    RI adaptors (Pharmacia), digested with Not I and cloned
                    into the Not I and Eco RI sites of the modified pT7T3
                    vector.  This library is the normalized version of
                    NCI_CGAP_Br1.1. Library was constructed by Bento Soares
                    and M. Fatima Bonaldo."
                    /db_xref="taxon:9606"
                    /clone="IMAGE:1607869"
                    /clone_lib="NCI_CGAP_Br2"
                    /sex="female, pooled"
                    /tissue_type="breast"
                    /lab_host="DH10B"
BASE COUNT      158 a     47 c     56 g     92 t
ORIGIN

  Query Match           6.4%;  Score 19;  DB 14;  Length 353;
  Best Local Similarity 100.0%;  Pred. No. 3.32e-04;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db       7 AAATAACTTTATTTATAAC 25
           ||||||||||||||||||||
Qy     120 AAATAACTTTATTTATAAC 138


RESULT   25
LOCUS       H67871        359 bp    mRNA            EST       11-DEC-1995
DEFINITION  yu67g08.s1 Homo sapiens cDNA clone 238910 3'.
```



```
ACCESSION   H67871
NID         g1026611
KEYWORDS    EST.
SOURCE      human clone=238910 primer=Promega -21m13 library=Weizmann Olfactory
            Epithelium vector=pBluescript SK- host=SOLR cells (kanamycin
            resistant) Reite1=EcoRI Reite2=XhoI From 35 year old female. The
            cDNA was oligo (dT) primed with an XhoI restriction  enzyme
            recognition site and an 18 base poly dT sequence.   For the 5' end,
            the synthesized cDNA termini were treated with T4 DNA polymerase
            and EcoRI adaptors were ligated to the blunt ends.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 359)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
            Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 273
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES            Location/Qualifiers
     source         1. .359
                    /organism="Homo sapiens"
                    /clone="238910"
     mRNA           <1. .>359
BASE COUNT       71 a     61 c     89 g    134 t      4 others
ORIGIN

  Query Match           6.4%;  Score 19;  DB 19;  Length 359;
  Best Local Similarity 100.0%;  Pred. No. 3.32e-04;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      62 TTATCTGTCTCTTGTAGTA 80
           ||||||||||||||||||||
Qy     229 TTATCTGTCTCTTGTAGTA 247
```



```
RESULT   26
LOCUS       AQ121777      365 bp    DNA             GSS       21-SEP-1998
DEFINITION  HS_3079_B2_C06_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3079 Col=12 Row=F, genomic survey
            sequence.
ACCESSION   AQ121777
NID         g3498943
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 365)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Furlong,J.,
            Shaker,R., Schmidt,S., Traicoff,R. and Hood,L.E.
  TITLE     Construction of a Characterized Clone Resource for Genomic
            Sequencing
  JOURNAL   Unpublished (1998)
COMMENT
            Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3079  row: F  column: 12
            Class: BAC ends
            High quality sequence stop: 365.
FEATURES            Location/Qualifiers
     source         1. .365
                    /organism="Homo sapiens"
                    /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                    E-Coli DH10B"
                    /db_xref="taxon:9606"
                    /clone="Plate=3079 Col=12 Row=F"
                    /clone_lib="CIT Approved Human Genomic Sperm Library D"
                    /sex="male"
BASE COUNT      133 a     48 c     49 g    130 t      5 others
ORIGIN

  Query Match           6.4%;  Score 19;  DB 28;  Length 365;
  Best Local Similarity 100.0%;  Pred. No. 3.32e-04;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     241 AATGCATTAAAAATATTTT 259
           ||||||||||||||||||||
Qy      96 AATGCATTAAAAATATTTT 114
```

```
RESULT   27
LOCUS       AA399557      377 bp    mRNA            EST       16-MAY-1997
DEFINITION  zt88d05.s1 Soares testis NHT Homo sapiens cDNA clone 729417 3',
            mRNA sequence.
ACCESSION   AA399557
NID         g2052561
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 377)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J.,
            Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-Merck EST Project 1997
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -41ml3 fwd. ET from Amersham
            High quality sequence stop: 346.
FEATURES                Location/Qualifiers
     source             1. .377
                        /organism="Homo sapiens"
                        /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                        polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                        was prepared from mRNA obtained from Clontech
                        Laboratories, Inc., and primed with  a Not I - oligo(dT)
                        primer [5'
                        TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3']. 
                        Double-stranded cDNA was ligated to Eco RI adaptors
                        (Pharmacia), digested with Not I and cloned into the Not I
                        and Eco RI sites of the modified pT7T3 vector. Library
                        went through one round of normalization to Cot5, and was
                        constructed  by Bento Soares and M. Fatima Bonaldo."
                        /db_xref="taxon:9606"
                        /clone="729417"
                        /clone_lib="Soares testis NHT"
                        /sex="male"
                        /lab_host="DH10B"
     mRNA               complement(<1. .>377)
```



```
                        /db_xref="GDB:5925946"
BASE COUNT       74 a      64 c      92 g     147 t
ORIGIN

  Query Match           6.4%;  Score 19;  DB 6;  Length 377;
  Best Local Similarity 100.0%;  Pred. No. 3.32e-04;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      79 TTATCTGTCTCTTGTAGTA 97
           ||||||||||||||||||||
Qy     229 TTATCTGTCTCTTGTAGTA 247


RESULT   28
LOCUS       AI150715      388 bp    mRNA            EST       30-SEP-1998
DEFINITION  qb77d10.x1 Soares_fetal_heart_NbHH19W Homo sapiens cDNA clone
            IMAGE:1706131 3', mRNA sequence.
ACCESSION   AI150715
NID         g3679184
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 388)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40ml3 fwd. ET from Amersham.
FEATURES                Location/Qualifiers
     source             1. .388
                        /organism="Homo sapiens"
                        /note="Organ: heart; Vector: pT7T3D (Pharmacia) with a
                        modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                        strand cDNA was primed with a Not I - oligo(dT) primer [5'
                        TGTTACCAATCTGAAGTGGGAGCGGCCGCATCTTTTTTTTTTTTTTTT 3'],
                        double-stranded cDNA was size selected, ligated to Eco RI
                        adapters (Pharmacia), digested with Not I and cloned into
                        the Not I and Eco RI sites of a modified pT7T3 vector
                        (Pharmacia). Library went through one round of
                        normalization to a Cot = 5. Library constructed by
                        M.Fatima Bonaldo. This library was constructed from the
                        same fetus as the fetal lung library, Soares fetal lung
                        NbHL19W."
```



```
                        /db_xref="taxon:9606"
                        /clone="IMAGE:1706131"
                        /clone_lib="Soares_fetal_heart_NbHH19W"
                        /sex="unknown"
                        /dev_stage="19 weeks"
                        /lab_host="DH10B (ampicillin resistant)"
BASE COUNT       75 a      65 c      92 g     156 t
ORIGIN

  Query Match           6.4%;  Score 19;  DB 18;  Length 388;
  Best Local Similarity 100.0%;  Pred. No. 3.32e-04;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      82 TTATCTGTCTCTTGTAGTA 100
           ||||||||||||||||||||
Qy     229 TTATCTGTCTCTTGTAGTA 247


RESULT   29
LOCUS       AI070901      390 bp    mRNA            EST       05-AUG-1998
DEFINITION  UI-R-C2-mt-b-06-0-UI.s1 UI-R-C2 Rattus norvegicus cDNA clone
            UI-R-C2-mt-b-06-0-UI 3', mRNA sequence.
ACCESSION   AI070901
NID         g3397152
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 390)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT
            Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            cDNA Library Preparation: M. Fatima Bonaldo, Ph.D. Clone
            distribution: clones will be available through Research Genetics
            Seq primer: M13 Forward.
FEATURES                Location/Qualifiers
     source             1. .390
                        /organism="Rattus norvegicus"
                        /strain="Sprague-Dawley"
                        /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
```



```
                        polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-C2
                        library is a subtracted library derived from the UI-R-C1
                        library, which is a subtracted library derived from the
                        UI-R-C0 library.  The UI-R-C0 library consisted of a
                        mixture of individually tagged normalized  libraries
                        constructed from rat placenta, adult lung, brain, liver,
                        kidney,  heart, spleen, ovary, muscle, 8, 12 and 18-day
                        embryo. The tag is a string of  3-5 nucleotides present
                        between the Not I site and the oligo-dT track which allows
                        identification of the library of origin of a clone within
                        the mixture. The subtracted library (UI-R-C2) was
                        constructed as follows: PCR amplified cDNA inserts from
                        UI-R-C1 clones from which 3' ESTs had been derived was
                        used as a driver in a hybridization with the UI-R-C1
                        library in the form of single-stranded circles. The
                        remaining single-stranded circles (subtracted library) was
                        purified by hydroxyapatite column chromatography,
                        converted to double-stranded circles and electroporated
                        into DH10B bacteria (Life Technologies) to generate the
                        UI-R-C2 library. This procedure has been previously
                        described (Bonaldo, Lennon and Soares, Genome Research 6:
                        791-806, 1996)"
                        /db_xref="taxon:10116"
                        /clone="UI-R-C2-mt-b-06-0-UI"
                        /clone_lib="UI-R-C2"
                        /dev_stage="adult"
                        /lab_host="DH10B (Life Technologies)"
BASE COUNT      124 a      81 c      75 g     110 t
ORIGIN

  Query Match           6.4%;  Score 19;  DB 17;  Length 390;
  Best Local Similarity 100.0%;  Pred. No. 3.32e-04;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      21 TATTTAAATAACTTTATTT 39
           ||||||||||||||||||||
Qy     115 TATTTAAATAACTTTATTT 133


RESULT   30
LOCUS       C25635        395 bp    mRNA            EST       24-JUL-1997
DEFINITION  Dictyostelium discoideum slug cDNA, clone SSA826, mRNA sequence.
ACCESSION   C25635
NID         g2281851
KEYWORDS    EST; EST(expressed sequence tag).
SOURCE      Dictyostelium discoideum (strain:AX4) Slug cDNA to mRNA,
            clone_lib:SS.
  ORGANISM  Dictyostelium discoideum
            Eukaryotae; mitochondrial eukaryotes; Dictyosteliida;
            Dictyostelium.
REFERENCE   1  (bases 1 to 395)
```

```
  AUTHORS   Urushihara,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1997) to the DDBJ/EMBL/GenBank databases. Hideko
            Urushihara, University of Tsukuba, Institute of Biological
            Sciences; Ten-noudai 1-1-1, Tsukuba-shi 305, Japan
            (E-mail:qxurushi@bank.dna.affrc.go.jp, Tel:0298-53-4664,
            Fax:0298-53-6614)
REFERENCE   2  (sites)
  AUTHORS   Saitoh,T., Morio,T., Tanaka,Y. and Ochiai,H.
  TITLE     Developmental cDNA in Dictyostelium discoideum(970723)
  JOURNAL   Unpublished (1997)
COMMENT     PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES             Location/Qualifiers
     source          1. .395
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone_lib="SS"
                     /dev_stage="Slug"
BASE COUNT      157 a     71 c     37 g    130 t
ORIGIN

  Query Match           6.4%;  Score 19;  DB 7;  Length 395;
  Best Local Similarity 100.0%;  Pred. No. 3.32e-04;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    266 ATTAAAAATATTTTTATTT 284
          |||||||||||||||||||
Qy    101 ATTAAAAATATTTTTATTT 119


RESULT   31
LOCUS       AA171915      415 bp    mRNA            EST       09-MAR-1998
DEFINITION  zo95d07.r1 Stratagene ovarian cancer (#937219) Homo sapiens cDNA
            clone 594637 5', mRNA sequence.
ACCESSION   AA171915
NID         g1751036
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 415)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
            Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
            Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Wilson RK
```



```
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1364   Std Error: 0.00
            Seq primer: -28M13 rev2 from Amersham.
FEATURES             Location/Qualifiers
     source          1. .415
                     /organism="Homo sapiens"
                     /note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; Cloned unidirectionally.  Primer: Oligo dT.
                     Papillary serous carcinoma, isolated from ascites, 64 year
                     old caucasian. Average insert size: 0.8 kb; Uni-ZAP XR
                     Vector; -5' adaptor sequence: 5' GAATTCGGCACGAG 3' -3'
                     adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTTT 3'"
                     /db_xref="taxon:9606"
                     /clone="594637"
                     /clone_lib="Stratagene ovarian cancer (#937219)"
                     /sex="female"
                     /dev_stage="adult, 64 years"
                     /lab_host="SOLR (kanamycin resistant)"
BASE COUNT      149 a     98 c     78 g     90 t
ORIGIN

  Query Match           6.4%;  Score 19;  DB 13;  Length 415;
  Best Local Similarity 100.0%;  Pred. No. 3.32e-04;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    250 TACTACAAGAGACAGATAA 268
          |||||||||||||||||||
Cp    247 TACTACAAGAGACAGATAA 229


RESULT   32
LOCUS       W42584        417 bp    mRNA            EST       21-MAY-1996
DEFINITION  zc23a05.s1 Soares senescent fibroblasts NbHSF Homo sapiens cDNA
            clone 323120 3', mRNA sequence.
ACCESSION   W42584
NID         g1327014
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 417)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
```



```
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: mob.REGA+ET
            High quality sequence stop: 299.
FEATURES             Location/Qualifiers
     source          1. .417
                     /organism="Homo sapiens"
                     /note="Vector: pT7T3D (Pharmacia) with a modified
                     polylinker V_TYPE: phagemid; Site_1: Not I; Site_2: Eco
                     RI; TGTTACCAATCTGAAGTGGGAGCGGCCGCATTTTTTTTTTTTTTTTTTT
                     3'), double-stranded cDNA was size selected, ligated to
                     Eco RI adapters (Pharmacia), digested with Not I and
                     cloned into the Not I and Eco RI sites of a modified pT7T3
                     vector (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by Bento
                     Soares and M.Fatima Bonaldo."
                     /db_xref="taxon:9606"
                     /clone="323120"
                     /clone_lib="Soares senescent fibroblasts NbHSF"
                     /tissue_type="senescent fibroblast"
                     /lab_host="DH10B (ampicillin resistant)"
     mRNA            complement(<1. .>417)
BASE COUNT       81 a     73 c    100 g    160 t      3 others
ORIGIN

  Query Match           6.4%;  Score 19;  DB 20;  Length 417;
  Best Local Similarity 100.0%;  Pred. No. 3.32e-04;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     82 TTATCTGTCTCTTGTAGTA 100
          |||||||||||||||||||
Qy    229 TTATCTGTCTCTTGTAGTA 247


RESULT   33
LOCUS       AA649293      418 bp    mRNA            EST       13-NOV-1997
DEFINITION  ns31g01.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1185264,
            mRNA sequence.
ACCESSION   AA649293
```



```
NID         g2575722
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 418)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            unknown library type
            Insert Length: 604   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 412.
FEATURES             Location/Qualifiers
     source          1. .418
                     /organism="Homo sapiens"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from human tonsillar cells enriched for
                     germinal center B cells by flow sorting (CD20+, IgD-),
                     provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                     (NCI) and Dr. Gerald Marti (CBER).  cDNA synthesis was
                     primed with a Not I - oligo(dT) primer
                     (5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCTCATTTTTTTTTTTTTTTTT-
                     3'). Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo."
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1185264"
                     /clone_lib="NCI_CGAP_GCB1"
                     /tissue_type="germinal center B cell"
                     /lab_host="DH10B"
BASE COUNT       80 a     68 c     99 g    171 t
ORIGIN

  Query Match           6.4%;  Score 19;  DB 10;  Length 418;
  Best Local Similarity 100.0%;  Pred. No. 3.32e-04;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     91 TTATCTGTCTCTTGTAGTA 109
          |||||||||||||||||||
Qy    229 TTATCTGTCTCTTGTAGTA 247
```

```
RESULT   34
LOCUS       AI058966      421 bp    mRNA            EST       20-JUL-1998
DEFINITION  UI-R-C1-lr-h-08-0-UI.s1 UI-R-C1 Rattus norvegicus cDNA clone
            UI-R-C1-lr-h-08-0-UI 3', mRNA sequence.
ACCESSION   AI058966
NID         g3332743
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 421)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT
            Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence tag present in the cDNA between the NotI site and the
            oligo-dT track served to identify it as a clone from the normalized
            adult Spleen library. cDNA Library Preparation: M. Fatima Bonaldo,
            Ph.D. Clone distribution: clones will be available through Research
            Genetics
            Seq primer: M13 Forward.
FEATURES             Location/Qualifiers
     source          1. .421
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-C1
                     library is a subtracted library derived from the UI-R-C0
                     library, which is a subtracted library derived from the
                     UI-R-A1 and UI-R-E1 libraries. The UI-R-A1 library
                     consisted of a mixture of  individually tagged normalized
                     libraries constructed from rat placenta,  adult lung,
                     brain, liver, kidney, heart, spleen, ovary, and muscle.
                     The UI-R-E1 library  consisted of a mixture of
                     individually tagged normalized libraries constructed from
                     8, 12 and 18-day embryo. The tag is a string of 3-5
                     nucleotides present between the Not I site and the
                     oligo-dT track which allows identification of the library
                     of origin of a clone within the mixture. The subtracted
                     library (UI-R-C1) was constructed as follows: PCR
```



```
                     amplified cDNA inserts from  UI-R-C0 clones from which 3'
                     ESTs had been derived was used as a driver in a
                     hybridization with the UI-R-C0 library in the form of
                     single-stranded circles. The remaining single-stranded
                     circles (subtracted library) was purified by
                     hydroxyapatite column chromatography, converted to
                     double-stranded circles and electroporated into DH10B
                     bacteria (Life Technologies) to generate the UI-R-C1
                     library. This procedure has been previously described
                     (Bonaldo, Lennon and Soares, Genome Research 6: 791-806,
                     1996)"
                     /db_xref="taxon:10116"
                     /clone="UI-R-C1-lr-h-08-0-UI"
                     /clone_lib="UI-R-C1"
                     /dev_stage="Adult"
                     /lab_host="DH10B (Life Technologies)"
BASE COUNT      133 a     90 c     81 g    117 t
ORIGIN

  Query Match           6.4%;  Score 19;  DB 15;  Length 421;
  Best Local Similarity 100.0%;  Pred. No. 3.32e-04;
  Matches    19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     21 TATTTAAATAACTTTATTT 39
          ||||||||||||||||||||
Qy    115 TATTTAAATAACTTTATTT 133


RESULT   35
LOCUS       R10366        445 bp    mRNA            EST       06-APR-1995
DEFINITION  yf37a04.s1 Homo sapiens cDNA clone 129006 3'.
ACCESSION   R10366
NID         g762322
KEYWORDS    EST.
SOURCE      human clone=129006 library=Soares fetal liver spleen 1NFLS
            vector=pT7T3D (Pharmacia) with a modified polylinker host=DH10B
            (ampicillin resistant) primer=-21m13 Rsitel=Pac I Rsite2=Eco RI
            Liver and spleen from a 20 week-post conception male fetus. 1st
            strand cDNA was primed with a Pac I - oligo(dT) primer [5'
            AACTGGAAGAATTAATAAAGATCTTTTTTTTTTTTTTTT 3'], double-stranded
            cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac
            I and cloned into the Pac I and Eco RI sites of the modified pT7T3
            vector.  Library went through one round of normalization. Library
            constructed by Bento Soares and M.Fatima Bonaldo.
  ORGANISM  Homo sapiens
            Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 445)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
```



```
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 306
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES             Location/Qualifiers
     source          1. .445
                     /organism="Homo sapiens"
                     /clone="129006"
BASE COUNT      122 a     93 c     86 g    129 t     15 others
ORIGIN

  Query Match           6.4%;  Score 19;  DB 5;  Length 445;
  Best Local Similarity 86.4%;  Pred. No. 3.32e-04;
  Matches    19;  Conservative    0;  Mismatches   3;  Indels   0;  Gaps   0;

Db     22 ANTTTTANTTAAATAANTTTAT 43
          | ||||| |||||||| |||||
Qy    110 ATTTTTATTTAAATAACTTTAT 131


RESULT   36
LOCUS       AI150662      449 bp    mRNA            EST       30-SEP-1998
DEFINITION  qc06c11.x1 Soares_fetal_heart_NbHH19W Homo sapiens cDNA clone
            IMAGE:1708820 3', mRNA sequence.
ACCESSION   AI150662
NID         g1679131
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 449)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
```



```
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40m13 fwd. ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .449
                     /organism="Homo sapiens"
                     /note="Organ: heart; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGGCATCTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by
                     M.Fatima Bonaldo. This library was constructed from the
                     same fetus as the fetal lung library, Soares fetal lung
                     NbHL19W."
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1708820"
                     /clone_lib="Soares_fetal_heart_NbHH19W"
                     /sex="unknown"
                     /dev_stage="19 weeks"
                     /lab_host="DH10B (ampicillin resistant)"
BASE COUNT       86 a     83 c    101 g    178 t       1 others
ORIGIN

  Query Match           6.4%;  Score 19;  DB 18;  Length 449;
  Best Local Similarity 100.0%;  Pred. No. 3.32e-04;
  Matches    19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     82 TTATCTGTCTCTTGTAGTA 100
          ||||||||||||||||||||
Qy    229 TTATCTGTCTCTTGTAGTA 247


RESULT   37
LOCUS       H22455        450 bp    mRNA            EST       06-JUL-1995
DEFINITION  yn69e02.s1 Homo sapiens cDNA clone 173690 3'.
ACCESSION   H22455
NID         g891150
KEYWORDS    EST.
SOURCE      Homo sapiens clone=173690 library=Soares adult brain N2b5HB55Y
            vector=pT7T3D (Pharmacia) with a modified polylinker host=DH10B
            (ampicillin resistant) primer=Promega -21m13 Rsitel=Not I
            Rsite2=Eco RI 55-year old male. 1st strand cDNA was primed with a
            Not I - oligo(dT) primer [5'
            TGTTACCAATCTGAAGTGGGAGCGGCCGCGCTTTTTTTTTTTTTTTTT 3'],
            double-stranded cDNA was size selected, ligated to Eco RI adapters
            (Pharmacia), digested with Not I and cloned into the Not I and Eco
```

```
            RI sites of a modified pT7T3 vector (Pharmacia). Library went
            through one round of normalization to a Cot = 53. Library
            constructed by Bento Soares and M.Fatima Bonaldo. The adult brain
            RNA was provided by Dr. Donald H. Gilden. Tissue was acquired 17-18
            hours after death which occurred in consequence of a ruptured
            aortic aneurysm. RNA was prepared from a pool of tissues
            representing the following areas of the brain: frontal, parietal,
            temporal and occipital cortex from the left and right hemispheres,
            subcortical white matter, basal ganglia, thalamus, cerebellum,
            midbrain, pons and medulla.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 450)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 142
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES             Location/Qualifiers
     source          1. .450
                     /organism="Homo sapiens"
                     /clone="173690"
BASE COUNT      108 a     95 c    140 g     98 t      9 others
ORIGIN

  Query Match           6.4%;  Score 19;  DB 16;  Length 450;
  Best Local Similarity  95.0%;  Pred. No. 3.32e-04;
  Matches    19;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db      6 TTAAAAATNTTTTTATTTAA 25
          |||||||| ||||||||||||
Qy    102 TTAAAAATATTTTTATTTAA 121
```



```
RESULT   38
LOCUS       AA154911      452 bp    mRNA            EST       11-FEB-1997
DEFINITION  ms02d03.r1 Stratagene mouse embryonic carcinoma (#937317) Mus
            musculus cDNA clone 605765 5', mRNA sequence.
ACCESSION   AA154911
NID         g1726531
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Mus.
REFERENCE   1  (bases 1 to 452)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:371197
            Putative full length read
            vector to vector length is 455
            Seq primer: -28m13 revl ET from Amersham
            High quality sequence stop: 416.
FEATURES             Location/Qualifiers
     source          1. .452
                     /organism="Mus musculus"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; Cloned unidirectionally.  Primer: Oligo dT. P19 cell
                     line. Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'
                     adaptor sequence: 5' GAATTCGGCACGAG 3' -3' adaptor
                     sequence: 5' CTCGAGTTTTTTTTTTTTTTTTT 3'"
                     /db_xref="taxon:10090"
                     /clone="605765"
                     /clone_lib="Stratagene mouse embryonic carcinoma
                     (#937317)"
                     /dev_stage="P19 cell line (ATCC CRL1825)"
                     /lab_host="SOLR (kanamycin resistant)"
     mRNA            <1. .>452
BASE COUNT      115 a     97 c    102 g    138 t
```



```
ORIGIN

  Query Match           6.4%;  Score 19;  DB 23;  Length 452;
  Best Local Similarity 100.0%;  Pred. No. 3.32e-04;
  Matches    19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    431 AAATAAAGTTATTTAAATA 449
          |||||||||||||||||||
Cp    133 AAATAAAGTTATTTAAATA 115


RESULT   39
LOCUS       R59260        500 bp    mRNA            EST       24-MAY-1995
DEFINITION  yg97e02.r1 Homo sapiens cDNA clone 41480 5'.
ACCESSION   R59260
NID         g829955
KEYWORDS    EST.
SOURCE      human clone=41480 library=Soares infant brain 1NIB vector=Lafmid BA
            host=DH10B (ampicillin resistant) primer=M13RP1 RsiteI=Not I
            Rsite2=Hind III Whole brain from a 73 days post natal female. 1st
            strand cDNA was primed with a Not I - oligo(dT) primer [5'
            AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTTT 3']; double-stranded
            cDNA was ligated to Hind III adaptors (Pharmacia), digested with
            Not I and directionally cloned into the Not I and Hind III sites of
            the Lafmid BA vector. Library went through one round of
            normalization. Library constructed by Bento Soares and M.Fatima
            Bonaldo.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 500)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            GDB: G00-414-021
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 347
            Source: IMAGE Consortium, LLNL
```



```
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES             Location/Qualifiers
     source          1. .500
                     /organism="Homo sapiens"
                     /clone="41480"
BASE COUNT      164 a     79 c    101 g    150 t      6 others
ORIGIN

  Query Match           6.4%;  Score 19;  DB 16;  Length 500;
  Best Local Similarity 100.0%;  Pred. No. 3.32e-04;
  Matches    19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    361 TTATTTAAATAAAAATATT 379
          |||||||||||||||||||
Cp    125 TTATTTAAATAAAAATATT 107


RESULT   40
LOCUS       AQ253742      504 bp    DNA             GSS       08-OCT-1998
DEFINITION  HS_3059_B2_A04_MF CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3059 Col=8 Row=B, genomic survey
            sequence.
ACCESSION   AQ253742
NID         g3725096
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 504)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Furlong,J.,
            Shaker,R., Schmidt,S., Traicoff,R. and Hood,L.E.
  TITLE     Construction of a Characterized Clone Resource for Genomic
            Sequencing
  JOURNAL   Unpublished (1998)
COMMENT
            Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3059  row: B  column: 8
            Class: BAC ends
            High quality sequence stop: 504.
FEATURES             Location/Qualifiers
     source          1. .504
                     /organism="Homo sapiens"
```

```
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
                     /db_xref="taxon:9606"
                     /clone="Plate=3059 Col=8 Row=B"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
BASE COUNT       141 a      97 c      98 g     168 t
ORIGIN

  Query Match           6.4%;  Score 19;  DB 29;  Length 504;
  Best Local Similarity 100.0%;  Pred. No. 3.32e-04;
  Matches    19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    197 AAAAATATTTTTAATGCAT 215
          |||||||||||||||||||
Cp    115 AAAAATATTTTTAATGCAT 97


RESULT   41
LOCUS       AI144725      509 bp    mRNA            EST       28-SEP-1998
DEFINITION  UI-R-BT0-po-c-01-0-UI.s1 UI-R-BT0 Rattus norvegicus cDNA clone
            UI-R-BT0-po-c-01-0-UI 3', mRNA sequence.
ACCESSION   AI144725
NID         g3666524
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 509)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT
            Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoaresblue.weeg.uiowa.edu
            The sequence tag present in the cDNA between the NotI site and the
            oligo-dT track served to identify it as a clone from the normalised
            adult NRBcorpus-striatum library. cDNA Library Preparation: M.
            Fatima Bonaldo, Ph.D. Clone distribution: clones will be available
            through Research Genetics The following repetitive elements were
            found in this cDNA sequence: 1-39, >AT_rich#Low_complexity
            Seq primer: M13 Forward.
FEATURES              Location/Qualifiers
```



```
     source           1. .509
                      /organism="Rattus norvegicus"
                      /strain="Sprague-Dawley"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; This library
                      (UI-R-BT0) consists of a mixture of individually tagged
                      normalized libraries constructed from rat hippocampus,
                      thalamus, mid-brain, medulla, corpus striatum, cerebral
                      cortex and testis. The tag used to  identify the source
                      tissue is a string of 3-6 nucleotides present  between the
                      Not I site and the oligo-dT track which allows
                      identification  of the library of origin of a clone within
                      the mixture. This library was then subtracted using a
                      driver consisting of a mixture of all clones from UI-R-A0,
                      UI-R-A1, UI-R-E0, UI-R-E1, UI-R-C0, UI-R-C1, UI-R-C2 and
                      UI-R-C2p."
                      /db_xref="taxon:10116"
                      /clone="UI-R-BT0-po-c-01-0-UI"
                      /clone_lib="UI-R-BT0"
                      /dev_stage="adult"
                      /lab_host="DH10B (Life Technologies)"
BASE COUNT       149 a     115 c      98 g     146 t       1 others
ORIGIN

  Query Match           6.4%;  Score 19;  DB 18;  Length 509;
  Best Local Similarity 100.0%;  Pred. No. 3.32e-04;
  Matches    19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     21 TATTTAAATAACTTTATTT 39
          |||||||||||||||||||
Qy    115 TATTTAAATAACTTTATTT 133


RESULT   42
LOCUS       AA876015      520 bp    mRNA            EST       31-MAR-1998
DEFINITION  ob93d12.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1338935 3',
            mRNA sequence.
ACCESSION   AA876015
NID         g2984856
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 520)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
```



```
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
            Ph.D., Gerald Marti, M.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 832  Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 483.
FEATURES              Location/Qualifiers
     source           1. .520
                      /organism="Homo sapiens"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                      was prepared from human tonsillar cells enriched for
                      germinal center B cells by flow sorting (CD20+, IgD-),
                      provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                      (NCI) and Dr. Gerald Marti (CBER).  cDNA synthesis was
                      primed with a Not I - oligo(dT) primer
                      [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTT-
                      3']. Double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Not I and cloned into the Not I
                      and Eco RI sites of the modified pT7T3 vector. Library
                      went through one round of normalization, and was
                      constructed by Bento Soares and M. Fatima Bonaldo."
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1338935"
                      /clone_lib="NCI_CGAP_GCB1"
                      /tissue_type="germinal center B cell"
                      /lab_host="DH10B"
BASE COUNT       99 a      96 c     115 g     210 t
ORIGIN

  Query Match           6.4%;  Score 19;  DB 14;  Length 520;
  Best Local Similarity 100.0%;  Pred. No. 3.32e-04;
  Matches    19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     91 TTATCTGTCTCTTGTAGTA 109
          |||||||||||||||||||
Qy    229 TTATCTGTCTCTTGTAGTA 247


RESULT   43
LOCUS       AA398463      625 bp    mRNA            EST       16-MAY-1997
DEFINITION  zt62e11.s1 Soares testis NHT Homo sapiens cDNA clone 726956 3',
```



```
            mRNA sequence.
ACCESSION   AA398463
NID         g2051573
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 625)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J.,
            Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-Merck EST Project 1997
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -41m13 fwd. ET from Amersham
            High quality sequence stop: 410.
FEATURES              Location/Qualifiers
     source           1. .625
                      /organism="Homo sapiens"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                      was prepared from mRNA obtained from Clontech
                      Laboratories, Inc., and primed with  a Not I - oligo(dT)
                      primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3']."
                      Double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Not I and cloned into the Not I
                      and Eco RI sites of the modified pT7T3 vector. Library
                      went through one round of normalization to Cot5, and was
                      constructed  by Bento Soares and M. Fatima Bonaldo."
                      /db_xref="taxon:9606"
                      /clone="726956"
                      /clone_lib="Soares testis NHT"
                      /sex="male"
                      /lab_host="DH10B"
     mRNA             complement(<1. .>625)
                      /db_xref="GDB:5923867"
BASE COUNT       128 a     129 c     145 g     223 t
ORIGIN
```

```
  Query Match           6.4%;  Score 19;  DB 6;  Length 625;
  Best Local Similarity 100.0%;  Pred. No. 3.32e-04;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      80 TTATCTGTCTCTTGTAGTA 98
           ||||||||||||||||||||
Qy     229 TTATCTGTCTCTTGTAGTA 247


RESULT   44
LOCUS       AG012998      687 bp    DNA             GSS       25-AUG-1998
DEFINITION  Homo sapiens genomic DNA, 21q region, clone: 72018Bm54, genomic
            survey sequence.
ACCESSION   AG012998
NID         g3452613
KEYWORDS    GSS.
SOURCE      Homo sapiens DNA, clone:72018Bm54.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 687)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Shiba,T. and Sakaki,Y.
  TITLE     Homo sapiens genomic DNA, chromosome 21q
  JOURNAL   Published Only in DataBase (1998) In press
REFERENCE   2  (bases 1 to 687)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Shiba,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-AUG-1998) to the DDBJ/EMBL/GenBank databases.
            Masahira Hattori, Kitasato University, Department of Science, JST
            Sequencing Laboratory; Kitasato 1-15-1, Sagamihara 228, Japan
            (E-mail:hattori@hgc.ims.u-tokyo.ac.jp, Tel:0427-78-9732,
            Fax:0427-78-9561)
FEATURES              Location/Qualifiers
     source           1. .687
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="21"
                      /clone="72018Bm54"
                      /map="21q"
BASE COUNT      238 a    127 c    117 g    205 t
ORIGIN

  Query Match           6.4%;  Score 19;  DB 27;  Length 687;
  Best Local Similarity 100.0%;  Pred. No. 3.32e-04;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     415 AATGAAAAATACTGAAGAA 433
           ||||||||||||||||||||
Cp     193 AATGAAAAATACTGAAGAA 175
```



```
RESULT   45
LOCUS       AG005353      687 bp    DNA             GSS       26-FEB-1998
DEFINITION  Homo sapiens genomic DNA, 21q region, clone: 72018Bm54, genomic
            survey sequence.
ACCESSION   AG005353
NID         g2913929
KEYWORDS    GSS.
SOURCE      Homo sapiens DNA, clone:72018Bm54.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 687)
  AUTHORS   Hattori,M., Ishii,K., Shiba,T. and Sakaki,Y.
  TITLE     Homo sapiens genomic DNA, chromosome 21q
  JOURNAL   Published Only in DataBase (1998) In press
REFERENCE   2  (bases 1 to 687)
  AUTHORS   Hattori,M., Ishii,K., Shiba,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-FEB-1998) to the DDBJ/EMBL/GenBank databases.
            Masahira Hattori, Kitasato University, Department of Science, JST
            Sequencing Laboratory; Kitasato 1-15-1, Sagamihara 228, Japan
            (E-mail:hattori@hgc.ims.u-tokyo.ac.jp, Tel:0427-78-9732,
            Fax:0427-78-9561)
FEATURES              Location/Qualifiers
     source           1. .687
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="21"
                      /clone="72018Bm54"
                      /map="21q"
BASE COUNT      238 a    127 c    117 g    205 t
ORIGIN

  Query Match           6.4%;  Score 19;  DB 26;  Length 687;
  Best Local Similarity 100.0%;  Pred. No. 3.32e-04;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     415 AATGAAAAATACTGAAGAA 433
           ||||||||||||||||||||
Cp     193 AATGAAAAATACTGAAGAA 175


RESULT   46
LOCUS       AG005363      706 bp    DNA             GSS       26-FEB-1998
DEFINITION  Homo sapiens genomic DNA, 21q region, clone: 72018Bm54, genomic
            survey sequence.
ACCESSION   AG005363
NID         g2913939
KEYWORDS    GSS.
SOURCE      Homo sapiens DNA, clone:72018Bm54.
```



```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 706)
  AUTHORS   Hattori,M., Ishii,K., Shiba,T. and Sakaki,Y.
  TITLE     Homo sapiens genomic DNA, chromosome 21q
  JOURNAL   Published Only in DataBase (1998) In press
REFERENCE   2  (bases 1 to 706)
  AUTHORS   Hattori,M., Ishii,K., Shiba,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-FEB-1998) to the DDBJ/EMBL/GenBank databases.
            Masahira Hattori, Kitasato University, Department of Science, JST
            Sequencing Laboratory; Kitasato 1-15-1, Sagamihara 228, Japan
            (E-mail:hattori@hgc.ims.u-tokyo.ac.jp, Tel:0427-78-9732,
            Fax:0427-78-9561)
FEATURES              Location/Qualifiers
     source           1. .706
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="21"
                      /clone="72018Bm54"
                      /map="21q"
BASE COUNT      213 a    116 c    137 g    239 t      1 others
ORIGIN

  Query Match           6.4%;  Score 19;  DB 26;  Length 706;
  Best Local Similarity 100.0%;  Pred. No. 3.32e-04;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      53 TTCTTCAGTATTTTCATT 71
           ||||||||||||||||||||
Qy     175 TTCTTCAGTATTTTCATT 193


RESULT   47
LOCUS       AG013008      706 bp    DNA             GSS       25-AUG-1998
DEFINITION  Homo sapiens genomic DNA, 21q region, clone: 72018Bm54, genomic
            survey sequence.
ACCESSION   AG013008
NID         g3452623
KEYWORDS    GSS.
SOURCE      Homo sapiens DNA, clone:72018Bm54.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 706)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Shiba,T. and Sakaki,Y.
  TITLE     Homo sapiens genomic DNA, chromosome 21q
  JOURNAL   Published Only in DataBase (1998) In press
REFERENCE   2  (bases 1 to 706)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Shiba,T. and Sakaki,Y.
```



```
  TITLE     Direct Submission
  JOURNAL   Submitted (25-AUG-1998) to the DDBJ/EMBL/GenBank databases.
            Masahira Hattori, Kitasato University, Department of Science, JST
            Sequencing Laboratory; Kitasato 1-15-1, Sagamihara 228, Japan
            (E-mail:hattori@hgc.ims.u-tokyo.ac.jp, Tel:0427-78-9732,
            Fax:0427-78-9561)
FEATURES              Location/Qualifiers
     source           1. .706
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="21"
                      /clone="72018Bm54"
                      /map="21q"
BASE COUNT      213 a    116 c    137 g    239 t      1 others
ORIGIN

  Query Match           6.4%;  Score 19;  DB 27;  Length 706;
  Best Local Similarity 100.0%;  Pred. No. 3.32e-04;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      53 TTCTTCAGTATTTTCATT 71
           ||||||||||||||||||||
Qy     175 TTCTTCAGTATTTTCATT 193


RESULT   48
LOCUS       B08181        931 bp    DNA             GSS       13-MAY-1997
DEFINITION  F15B10-Sp6 IGF Arabidopsis thaliana genomic clone F15B10, genomic
            survey sequence.
ACCESSION   B08181
NID         g2089303
KEYWORDS    GSS.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryotae; mitochondrial eukaryotes; Viridiplantae;
            Charophyta/Embryophyta group; Embryophyta; vascular plants; seed
            plants; Magnoliophyta; Magnoliopsida; Capparales; Brassicaceae;
            Arabidopsis.
REFERENCE   1  (bases 1 to 931)
  AUTHORS   Feng,J., Dewar,K., Buehler,E., Kim,C., Li,Y., Shinn,P., Sun,H. and
            Ecker,J.
  TITLE     BAC End Sequences at ATGC
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: F15B10-Sp6.1, F15B10-T7
            Contact: Ecker J.
            Arabidopsis Thaliana Genome Center
            University of Pennsylvania
            Dept. of Biology, University of Pennsylvania, Philadelphia, PA
            19104
            Tel: 215-898-9384
            Fax: 215-898-8780
```

```
            Email: jecker@atgenome.bio.upenn.edu
            Seq primer: Sp6
            Class: BAC ends
            High quality sequence start: 435
            High quality sequence stop: 512.
FEATURES              Location/Qualifiers
     source           1. .931
                      /organism="Arabidopsis thaliana"
                      /strain="Columbia"
                      /note="Vector: BeloBACII; Site_1: EcoRI; Site_2: EcoRI;
                      Produced by Thomas Altmann"
                      /db_xref="taxon:3702"
                      /clone="F15B10"
                      /clone_lib="IGF"
                      /sex="hermaphrodite"
BASE COUNT       331 a    172 c    118 g    291 t     19 others
ORIGIN

  Query Match           6.4%;  Score 19;  DB 26;  Length 931;
  Best Local Similarity 100.0%;  Pred. No. 3.32e-04;
  Matches    19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    258 AATGAAAAATACTGAAGAA 276
          ||||||||||||||||||
Cp    193 AATGAAAAATACTGAAGAA 175


RESULT   49
LOCUS       AA102128      152 bp    mRNA            EST       28-OCT-1996
DEFINITION  zl80a03.s1 Stratagene colon (#937204) Homo sapiens cDNA clone
            510892 3' similar to gb:S52450 NONSPECIFIC LIPID-TRANSFER PROTEIN
            PRECURSOR (HUMAN);, mRNA sequence.
ACCESSION   AA102128
NID         g1646264
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 152)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
            Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
```



```
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Trace considered overall poor quality
            Seq primer: -40M13 fwd. from Amersham
            High quality sequence stop: 1.
FEATURES              Location/Qualifiers
     source           1. .152
                      /organism="Homo sapiens"
                      /note="Organ: colon; Vector: pBluescript SK-; Site_1:
                      EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                      Oligo dT. T-84 colonic epithelial cell line.  Average
                      insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor
                      sequence: 5' GAATTCGGCACGAG 3' -3' adaptor sequence: 5'
                      CTCGAGTTTTTTTTTTTTTTTT 3'"
                      /db_xref="taxon:9606"
                      /clone="510892"
                      /clone_lib="Stratagene colon (#937204)"
                      /lab_host="SOLR cells (kanamycin resistant)"
     mRNA             complement(<1. .>152)
BASE COUNT       49 a     20 c     21 g     56 t      6 others
ORIGIN

  Query Match           6.1%;  Score 18;  DB 22;  Length 152;
  Best Local Similarity 100.0%;  Pred. No. 7.21e-03;
  Matches    18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     99 TTAAAAATATTTTTATTT 116
          ||||||||||||||||||
Qy    102 TTAAAAATATTTTTATTT 119


RESULT   50
LOCUS       AA283357      173 bp    mRNA            EST       10-MAR-1998
DEFINITION  TH222 HTCDL1 Homo sapiens cDNA 5'/3' similar to lipid-transfer
            protein precursor, mRNA sequence.
ACCESSION   AA283357
NID         g2947662
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 173)
  AUTHORS   Sohn,U., Park,D.S., Lee,C.M., Cho,W.K., Ahn,H.J., Lee,M.Y.,
            Hwang,M.Y. and Jin,S.W.
  TITLE     Human HTCDL1 library cDNAs
  JOURNAL   Unpublished (1994)
COMMENT
```



```
            Contact: Uik
            Sohn,D.S.Park,C.M.Lee,W.K.Cho,H.J.Ahn,M.Y.Lee,M.Y.Hwang,S.W.Jin
            Laboratory of Molecular Biology
            Kyungpook National University
            Dept. of Genetic Eng.,Kyungpook National Univ., Taegu 702-701,
            Korea
            Tel: 82-053-950-5382
            Fax: 82-053-955-5327
            Email: usohn@bh.kyungpook.ac.kr
            Seq primer: M13 Reverse/SK primer.
FEATURES              Location/Qualifiers
     source           1. .173
                      /organism="Homo sapiens"
                      /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
                      EcoRI; Poly(A)-mRNA  from the  2-year  old male  fetal
                      thymus, oligo(dT)  priming,  EcoRI cloning in the vector
                      pBluescipt (Stratagene)."
                      /db_xref="taxon:9606"
                      /clone_lib="HTCDL1"
                      /lab_host="XL1-Blue"
BASE COUNT       63 a     25 c     19 g     66 t
ORIGIN

  Query Match           6.1%;  Score 18;  DB 13;  Length 173;
  Best Local Similarity 100.0%;  Pred. No. 7.21e-03;
  Matches    18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     22 AAATAAAAATATTTTAA 39
          |||||||||||||||||
Cp    119 AAATAAAAATATTTTAA 102


Search completed: Tue Dec 22 19:45:03 1998
Job time : 529 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Tue Dec 22 19:45:23 1998; MasPar time 56.69 Seconds
710.426 Million cell updates/sec

Tabular output not generated.

Title: >US-08-530-112A-563
Description: (1-296) from US08530112A.seq
Perfect Score: 296
N.A. Sequence: 1 GATCTGCCATAAGAAAATCT.....TTTTGGTTGTTAAACTGAAA 296
Comp: CTAGACGGTATTCTTTTAGA.....AAAACCAACAATTTGACTTT

Scoring table: TABLE jmetric
Gap 60

Kmatch STD : Dbase 0; Query 0

Searched: 188442 seqs, 68026449 bases x 2

Post-processing: Minimum Match 0%
Listing first 1000 summaries

Database: n-geneseq32
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39 40:part40

Statistics: Mean 6.658; Variance 3.138; scale 2.122

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	Query Match	Length	DB	ID	Description	Pred. No.
1		296	100.0	296	18	T19563	Human gene signature	4.08e-241
2		17	5.7	399	13	Q78184	Maize associated regi	3.99e+00
3	c	17	5.7	870	17	Q98473	Carboxy terminus of M	3.99e+00
4		17	5.7	3441	34	T87922	Rat cerebellum derive	3.99e+00
5	c	17	5.7	3441	34	T87922	Rat cerebellum derive	3.99e+00
6	c	17	5.7	580073	27	T58840	Mycoplasma genitalium	3.99e+00
7		16	5.4	265	23	T26241	Human gene signature	1.55e+01
8		16	5.4	296	40	V04278	Secreted protein C195	1.55e+01
9		16	5.4	364	19	T20533	Human gene signature	1.55e+01
10	c	16	5.4	399	13	Q78184	Maize associated regi	1.55e+01
11	c	16	5.4	1218	10	Q45345	Rat 20 alpha-hydroxys	1.55e+01
12		16	5.4	1393	2	N60225	Sequence of autonomou	1.55e+01
13		16	5.4	1694	23	T15232	Tumor necrosis factor	1.55e+01
14		16	5.4	1694	36	T94635	TNF-R1-DD ligand prot	1.55e+01
15	c	16	5.4	3543	3	Q14669	Dipteran active toxin	1.55e+01
16	c	16	5.4	3543	13	Q81178	B.t. toxin PS71M3 gen	1.55e+01
17	c	16	5.4	3940	1	N93054	Delta-endotoxin cryst	1.55e+01
18	c	16	5.4	4571	1	N93059	Delta-endotoxin cryst	1.55e+01
19		16	5.4	4704	1	Q05264	Sequence encoding bea	1.55e+01
20	c	16	5.4	4934	1	N81490	Insecticidal (Diptera	1.55e+01
21	c	16	5.4	4954	14	Q55753	Escherichia coli geno	1.55e+01
22	c	16	5.4	37895	24	T34137	Gene for biosynthetic	1.55e+01
23	c	16	5.4	37913	10	Q51128	Eicosapentaenoic acid	1.55e+01
24	c	16	5.4	133894	17	T13635	AcNPV genomic DNA clo	1.55e+01
25		16	5.4	134525	1	Q04525	Total base sequence o	1.55e+01
26		15	5.1	66	22	T25856	Human gene signature	5.76e+01
27	c	15	5.1	146	18	T19710	Human gene signature	5.76e+01
28	c	15	5.1	163	21	T24769	Human gene signature	5.76e+01
29	c	15	5.1	205	1	N90214	Malaria-specific DNA	5.76e+01
30		15	5.1	246	21	T25769	Human gene signature	5.76e+01
31		15	5.1	271	7	Q50336	Serglycin - 3' end.	5.76e+01
32	c	15	5.1	304	21	T26536	Human gene signature	5.76e+01
33	c	15	5.1	351	19	T19975	Human gene signature	5.76e+01
34		15	5.1	390	19	T22717	Human gene signature	5.76e+01
35		15	5.1	391	19	T20524	Human gene signature	5.76e+01
36		15	5.1	593	14	Q85373	Growth factor-activat	5.76e+01
37		15	5.1	609	29	T48043	Yeast chromosome IX-d	5.76e+01
38		15	5.1	709	1	Q03374	Partial sequence of g	5.76e+01
39	c	15	5.1	731	5	Q31693	RsaI restriction frag	5.76e+01
40	c	15	5.1	749	2	N81961	Sequence of rat reg c	5.76e+01
41	c	15	5.1	919	38	V06125	Viral infection gene	5.76e+01
42		15	5.1	975	40	V24837	H. pylori ORF hp3p108	5.76e+01

43		15	5.1	1116	35	T75679	Borrelia burgdorferi	5.76e+01
44		15	5.1	1183	11	Q63613	T-cell transmembrane	5.76e+01
45		15	5.1	1183	3	Q14534	Lov gene (cDNA 19.5).	5.76e+01
46	c	15	5.1	1198	35	T79694	BRCA2 cancer suscepti	5.76e+01
47		15	5.1	1301	33	T61407	MACH isoform alpha3 c	5.76e+01
48		15	5.1	1361	14	Q83838	B. burgdorferi 81 gen	5.76e+01
49		15	5.1	1443	33	T61411	MACH isoform beta5 co	5.76e+01
50	c	15	5.1	1626	35	T88774	Human receptor protei	5.76e+01
51		15	5.1	1864	2	N71405	Sequence of ANS-1 whi	5.76e+01
52		15	5.1	1948	2	Q12693	Toxin A gene fragment	5.76e+01
53	c	15	5.1	1969	1	N93087	Human basic fibroblas	5.76e+01
54		15	5.1	1977	2	Q11986	Transcription activat	5.76e+01
55	c	15	5.1	2015	38	V06386	B. thuringiensis 14 a	5.76e+01
56		15	5.1	2026	40	V09901	cDNA encoding a mouse	5.76e+01
57		15	5.1	2113	18	T00521	H6 promoted canine di	5.76e+01
58		15	5.1	2198	39	V05124	Fusion protein gene o	5.76e+01
59	c	15	5.1	2267	36	T94700	Murine Pisp12 cDNA.	5.76e+01
60		15	5.1	2322	37	T88849	Cell death-associated	5.76e+01
61	c	15	5.1	2330	4	Q26422	Gene for beta-IG-M2.	5.76e+01
62		15	5.1	2412	29	T60072	63B toxin coding sequ	5.76e+01
63		15	5.1	2412	5	Q30940	Encodes BT toxin 63B.	5.76e+01
64		15	5.1	2412	6	Q32164	BT toxin 63B	5.76e+01
65		15	5.1	2412	5	Q31413	Bacillus thuringiensi	5.76e+01
66		15	5.1	2619	33	T61406	MACH isoform alpha2 c	5.76e+01
67	c	15	5.1	2776	3	Q22780	Codes for beta-subuni	5.76e+01
68	c	15	5.1	2776	1	Q06050	Sequence encoding rec	5.76e+01
69	c	15	5.1	2776	1	N80863	Sequence of cDNA clon	5.76e+01
70		15	5.1	2887	33	T61405	MACH isoform alpha1 c	5.76e+01
71		15	5.1	2920	36	T96727	Human adducin-like 70	5.76e+01
72	c	15	5.1	2973	9	Q53501	Sequence of bean mali	5.76e+01
73		15	5.1	3031	1	N81541	Sequence of pGP5-5.	5.76e+01
74		15	5.1	3031	5	Q31624	pGP5-5 containing ge	5.76e+01
75		15	5.1	3081	1	Q05876	pGP5-5 plasmid sequen	5.76e+01
76		15	5.1	3467	19	T12324	DNA encoding sulfolob	5.76e+01
77		15	5.1	3591	10	Q55942	Human leukaemia inhib	5.76e+01
78		15	5.1	3591	9	Q58427	Human LIF-R.	5.76e+01
79		15	5.1	3591	16	Q92272	Human leukaemia inhib	5.76e+01
80		15	5.1	3661	20	T05902	Human IL-2 gene (p2IL	5.76e+01
81	c	15	5.1	3877	37	V02209	cDNA of leaderless pr	5.76e+01
82	c	15	5.1	3878	1	N81502	DNA encoding human ba	5.76e+01
83	c	15	5.1	3878	2	N70494	Sequence encoding hum	5.76e+01
84	c	15	5.1	4343	18	T00522	pMM126 coding sequenc	5.76e+01
85		15	5.1	4586	38	T96838	Intron 8 of human PRb	5.76e+01
86	c	15	5.1	4604	18	T00523	pC5LCDVHAF1 coding se	5.76e+01
87	c	15	5.1	4998	1	N81114	Non-A, non-B hepatiti	5.76e+01
88		15	5.1	5212	2	N80317	Transcription control	5.76e+01
89	c	15	5.1	5212	2	N80317	Transcription control	5.76e+01
90		15	5.1	5737	10	Q58034	Interleukin-2 gene.	5.76e+01
91		15	5.1	5737	16	Q74053	Human interleukin-2.	5.76e+01
92		15	5.1	6313	5	Q29965	Sugar beet chitinase	5.76e+01
93		15	5.1	6452	22	T30193	Peripheral nervous sy	5.76e+01

94	c	15	5.1	6767	1	Q05883	Plasmid clone pUC-SK1	5.76e+01
95		15	5.1	10670	38	T99449	Staphylococcus carnos	5.76e+01
96	c	15	5.1	11283	35	T84841	Human breast and ovar	5.76e+01
97	c	15	5.1	11385	36	T69707	Human breast cancer s	5.76e+01
98		15	5.1	11785	12	Q74694	Genomic RNA sequence	5.76e+01
99	c	15	5.1	12132	7	Q43847	Plasmid pAH807.	5.76e+01
100	c	15	5.1	12284	20	T29591	Hog cholera virus.	5.76e+01
101	c	15	5.1	12284	1	Q06001	Sequence encoding pro	5.76e+01
102	c	15	5.1	12284	12	Q71308	Sequence encoding 55	5.76e+01
103		15	5.1	17327	7	Q44278	Serglycin - proteogly	5.76e+01
104	c	15	5.1	19124	32	T72882	Plasmodium var-7 gene	5.76e+01
105		15	5.1	31122	7	Q40706	Bacillus subtilis srf	5.76e+01
106	c	15	5.1	35100	40	V20441	Human c-fms oncogene.	5.76e+01
107		15	5.1	53526	40	T94101	Human PKD1 gene.	5.76e+01
108		15	5.1	53577	28	T18551	Human polycystic kidn	5.76e+01
109		15	5.1	53577	40	T94108	Human PKD1 locus betw	5.76e+01
110		14	4.7	21	35	T92587	BRCA2 cancer suscepti	2.03e+02
111		14	4.7	24	2	Q12437	Primer HPU2_1 for PCR	2.03e+02
112	c	14	4.7	27	4	Q24751	V-gamma 1.4 sequence,	2.03e+02
113		14	4.7	78	19	T21245	Human gene signature	2.03e+02
114		14	4.7	80	19	T23866	Human gene signature	2.03e+02
115	c	14	4.7	84	19	T22087	Human gene signature	2.03e+02
116		14	4.7	102	21	T26575	Human gene signature	2.03e+02
117	c	14	4.7	102	22	T25782	Human gene signature	2.03e+02
118		14	4.7	143	22	T26784	Human gene signature	2.03e+02
119	c	14	4.7	144	5	Q33570	Sequence upstream of	2.03e+02
120	c	14	4.7	173	18	T19297	Human gene signature	2.03e+02
121		14	4.7	185	21	T23363	Human gene signature	2.03e+02
122		14	4.7	187	19	T23793	Human gene signature	2.03e+02
123	c	14	4.7	187	20	T23649	Human gene signature	2.03e+02
124		14	4.7	191	20	T23025	Human gene signature	2.03e+02
125		14	4.7	213	20	T22504	Human gene signature	2.03e+02
126		14	4.7	223	3	Q14661	Part of melon cellula	2.03e+02
127		14	4.7	240	11	Q77251	Human genome fragment	2.03e+02
128	c	14	4.7	246	21	T25769	Human gene signature	2.03e+02
129		14	4.7	270	20	T22686	Human gene signature	2.03e+02
130		14	4.7	279	30	T68015	H. pylori cell envelo	2.03e+02
131		14	4.7	296	20	T21886	Human gene signature	2.03e+02
132	c	14	4.7	296	18	T19563	Human gene signature	2.03e+02
133	c	14	4.7	297	19	T20029	Human gene signature	2.03e+02
134	c	14	4.7	300	18	T19166	Human gene signature	2.03e+02
135		14	4.7	304	21	T26536	Human gene signature	2.03e+02
136		14	4.7	326	8	Q59521	Human brain Expressed	2.03e+02
137		14	4.7	345	20	T24045	Human gene signature	2.03e+02
138		14	4.7	359	20	T22888	Human gene signature	2.03e+02
139	c	14	4.7	361	19	T20648	Human gene signature	2.03e+02
140	c	14	4.7	371	19	T19955	Human gene signature	2.03e+02
141	c	14	4.7	375	8	Q60949	Human brain Expressed	2.03e+02
142		14	4.7	385	7	Q42778	Ligand-induced gene,	2.03e+02
143		14	4.7	411	30	T67470	H. pylori cell envelo	2.03e+02
144		14	4.7	452	5	Q28760	Sequence of clone 19	2.03e+02

	145	14	4.7	452	5	Q28007	Sequence encoding clo 2.03e+02
	146	14	4.7	453	2	Q11804	Promoter activated by 2.03e+02
	147	14	4.7	458	17	T10808	Human papilloma virus 2.03e+02
	148	14	4.7	458	1	Q03906	L1 amplified region; 2.03e+02
	149	14	4.7	458	38	T91259	Human papilloma virus 2.03e+02
	150	14	4.7	458	10	Q56377	HPV isolate 238A L1 a 2.03e+02
	151	14	4.7	458	32	T77882	DNA sequence of L1 am 2.03e+02
	152	14	4.7	458	24	T44697	Human papillomavirus 2.03e+02
c	153	14	4.7	504	29	T67378	H. pylori surface mem 2.03e+02
c	154	14	4.7	522	12	Q73898	Borrelia VSDA antigen 2.03e+02
c	155	14	4.7	525	12	Q73900	Borrelia 20047 antige 2.03e+02
c	156	14	4.7	531	12	Q73891	Borrelia ACA1 antigen 2.03e+02
c	157	14	4.7	534	12	Q73892	Borrelia H9 antigen v 2.03e+02
c	158	14	4.7	549	35	T72981	DNA encoding sugar be 2.03e+02
c	159	14	4.7	550	15	Q87995	Sugar beet anti-micro 2.03e+02
c	160	14	4.7	573	12	Q73873	Borrelia VSDA antigen 2.03e+02
c	161	14	4.7	576	12	Q73875	Borrelia 20047 antige 2.03e+02
c	162	14	4.7	582	12	Q73866	Borrelia ACA1 antigen 2.03e+02
c	163	14	4.7	585	12	Q73867	Borrelia H9 antigen v 2.03e+02
c	164	14	4.7	607	28	T36876	Piscicolin 126 gene. 2.03e+02
c	165	14	4.7	608	23	T37055	Drosophila EcoRI geno 2.03e+02
	166	14	4.7	693	36	T96479	MSRV-1 gag gene seque 2.03e+02
	167	14	4.7	693	29	T67402	H. pylori inner membr 2.03e+02
c	168	14	4.7	717	1	Q04328	Probe for sequence en 2.03e+02
	169	14	4.7	749	1	Q04330	Probe for sequence en 2.03e+02
c	170	14	4.7	749	1	Q04311	Probe for sequence en 2.03e+02
	171	14	4.7	773	36	T91163	Strawberry fruit sucr 2.03e+02
c	172	14	4.7	779	35	T79332	DNA encoding whale ma 2.03e+02
c	173	14	4.7	792	7	Q42596	Sequence of variant o 2.03e+02
c	174	14	4.7	811	37	T80820	Staphylococcus aureus 2.03e+02
c	175	14	4.7	909	29	T65493	Wild type uricase cod 2.03e+02
c	176	14	4.7	909	9	Q50608	Sequence encoding uri 2.03e+02
c	177	14	4.7	959	4	Q24874	Alpha-ald gene 2. 2.03e+02
	178	14	4.7	1036	35	T74048	Human GRP17 cDNA clon 2.03e+02
	179	14	4.7	1065	27	T43381	Human cytokine respon 2.03e+02
c	180	14	4.7	1106	12	Q72602	Plant insecticidal pa 2.03e+02
	181	14	4.7	1116	9	Q52273	Soybean seed storage 2.03e+02
c	182	14	4.7	1146	17	T08172	Full length meadowfoa 2.03e+02
c	183	14	4.7	1172	12	Q72601	Plant insecticidal pa 2.03e+02
	184	14	4.7	1215	31	T43468	ATM gene exons 30-31. 2.03e+02
	185	14	4.7	1236	38	V06459	Esex sequence used fo 2.03e+02
c	186	14	4.7	1254	31	T43482	ATM gene exons 48-49. 2.03e+02
c	187	14	4.7	1278	30	T67906	H. pylori surface or 2.03e+02
c	188	14	4.7	1301	33	T61407	MACH isoform alpha3 c 2.03e+02
	189	14	4.7	1302	24	T43045	Helicobacter pylori p 2.03e+02
	190	14	4.7	1314	30	T68169	H. pylori inner membr 2.03e+02
c	191	14	4.7	1316	30	T63621	Interleukin-12 p35 su 2.03e+02
c	192	14	4.7	1364	40	V15228	Heterodimeric human i 2.03e+02
c	193	14	4.7	1364	26	T48099	Human interleukin-12 2.03e+02
c	194	14	4.7	1396	4	Q24371	Natural killer cell s 2.03e+02
	195	14	4.7	1406	29	T41619	Tomato endo-xylogluca 2.03e+02

	196	14	4.7	1426	17	T00937	Almond N-glycosidase 2.03e+02
c	197	14	4.7	1430	29	T60142	Arabidopsis terminal 2.03e+02
c	198	14	4.7	1431	2	N70264	Sequence encoding alp 2.03e+02
	199	14	4.7	1435	17	T00935	Almond N-glycosidase 2.03e+02
c	200	14	4.7	1443	33	T61411	MACH isoform beta5 co 2.03e+02
c	201	14	4.7	1515	39	T94216	Vibrio harveyi endogl 2.03e+02
c	202	14	4.7	1569	26	T28573	Bacterial antibiotic 2.03e+02
	203	14	4.7	1577	36	T96477	MSRV-1 clone G+E+A. 2.03e+02
c	204	14	4.7	1598	27	T28520	H. influenzae detecti 2.03e+02
c	205	14	4.7	1622	24	T34592	NTII-11 nerve protein 2.03e+02
c	206	14	4.7	1734	19	Q97800	Clone pCBL80 coding f 2.03e+02
c	207	14	4.7	1740	5	Q30046	Encodes Porcine Parvo 2.03e+02
c	208	14	4.7	1820	7	Q42760	Swine parvovirus B ge 2.03e+02
c	209	14	4.7	1831	2	N70038	Fragment of pLIP1 ins 2.03e+02
	210	14	4.7	1839	23	T35301	Bluegill sunfish sacc 2.03e+02
	211	14	4.7	1849	25	T44514	Arabidopsis S-adenosy 2.03e+02
c	212	14	4.7	1924	1	N90684	Gene encoding the 43 2.03e+02
c	213	14	4.7	1945	8	Q47816	Sequence encoding 43k 2.03e+02
	214	14	4.7	1961	36	T95211	Human disintegrin met 2.03e+02
	215	14	4.7	2004	39	V16167	DNA encoding an 80 kD 2.03e+02
	216	14	4.7	2004	38	T73995	B. cereus 80 kD VIP1A 2.03e+02
	217	14	4.7	2004	13	Q74680	Bacillus cereus 80 kD 2.03e+02
	218	14	4.7	2004	20	T13941	B. cereus VIP1 protei 2.03e+02
c	219	14	4.7	2044	6	Q37838	Sequence of the contr 2.03e+02
c	220	14	4.7	2044	18	T13167	Rat interleukin-1 typ 2.03e+02
c	221	14	4.7	2053	22	T37083	AR791 fragment encodi 2.03e+02
c	222	14	4.7	2060	11	Q64002	2 kilobase genomic se 2.03e+02
c	223	14	4.7	2060	3	N40068	Sequence of the codin 2.03e+02
c	224	14	4.7	2060	3	Q22615	Bce4 gene. 2.03e+02
c	225	14	4.7	2060	2	Q13861	B.campestris seed-spe 2.03e+02
c	226	14	4.7	2062	1	N70818	Pig parvo virus B cap 2.03e+02
c	227	14	4.7	2062	7	Q42754	pSY875 swine parvovir 2.03e+02
c	228	14	4.7	2062	1	N91150	Swine parvovirus B Ge 2.03e+02
c	229	14	4.7	2062	22	T31894	Swine parvovirus caps 2.03e+02
c	230	14	4.7	2081	2	Q13862	Bce4 genomic clone mu 2.03e+02
c	231	14	4.7	2120	1	N91155	Swine parvovirus B ge 2.03e+02
c	232	14	4.7	2131	3	Q20684	pY53.3 (NCIMB 40308). 2.03e+02
c	233	14	4.7	2182	38	T99933	Racillus subtilis rod 2.03e+02
c	234	14	4.7	2249	3	Q15169	Bacillus thuringiensi 2.03e+02
c	235	14	4.7	2281	38	T99936	Racillus subtilis rod 2.03e+02
	236	14	4.7	2299	10	Q58976	B.thuringiensis serov 2.03e+02
c	237	14	4.7	2322	37	T88849	Cell death-associated 2.03e+02
c	238	14	4.7	2335	1	N81781	Sequence of the circu 2.03e+02
c	239	14	4.7	2337	2	N60362	Sequence encoding CS 2.03e+02
c	240	14	4.7	2340	1	N81108	Sequence of the circu 2.03e+02
c	241	14	4.7	2360	4	Q24438	EGF/HB-EHM. 2.03e+02
	242	14	4.7	2400	7	Q42496	Full-length human pse 2.03e+02
	243	14	4.7	2445	7	Q48393	Full-length foetal hu 2.03e+02
	244	14	4.7	2445	2	N60111	Sequence encoding a p 2.03e+02
c	245	14	4.7	2538	4	Q27509	Encodes alpha-acetola 2.03e+02
c	246	14	4.7	2574	36	T73125	Mouse Tec tyrosine ki 2.03e+02

c	247	14	4.7	2619	33	T61406	MACH isoform alpha2 c 2.03e+02
	248	14	4.7	2629	4	Q24523	Vasoactive intestinal 2.03e+02
	249	14	4.7	2629	1	N81338	Protein A gene with n 2.03e+02
	250	14	4.7	2655	38	T73994	B. cereus VIP1A(a) co 2.03e+02
	251	14	4.7	2655	13	Q74679	Bacillus cereus 100 k 2.03e+02
	252	14	4.7	2655	39	V16166	DNA encoding a 100 kD 2.03e+02
	253	14	4.7	2655	20	T13940	B. cereus VIP1A(a) ge 2.03e+02
c	254	14	4.7	2696	2	N60909	Plasmid pAU157 insert 2.03e+02
c	255	14	4.7	2696	2	N60889	Plasmid pAU157 sequen 2.03e+02
c	256	14	4.7	2697	2	N60864	Sequence of plasmid p 2.03e+02
c	257	14	4.7	2735	25	T44351	H. pylori ureA + ureB 2.03e+02
	258	14	4.7	2750	6	Q38808	G6PD coding sequence. 2.03e+02
	259	14	4.7	2763	36	T95212	Human disintegrin met 2.03e+02
c	260	14	4.7	2767	2	Q12485	DNA encoding A and B 2.03e+02
c	261	14	4.7	2812	16	T02327	Human fibroblast acti 2.03e+02
c	262	14	4.7	2815	34	T89286	Human fibroblast acti 2.03e+02
c	263	14	4.7	2815	34	T90148	Human fibroblast acti 2.03e+02
c	264	14	4.7	2887	33	T61405	MACH isoform alpha1 c 2.03e+02
	265	14	4.7	2916	18	T15957	Cycloisomaltooligosac 2.03e+02
	266	14	4.7	2974	13	Q79630	Partial FRAXE region 2.03e+02
c	267	14	4.7	3024	6	Q39260	ICAM-1 cDNA. 2.03e+02
c	268	14	4.7	3024	17	T11672	Intracellular adhesio 2.03e+02
c	269	14	4.7	3024	1	Q04049	Sequence of clone HL- 2.03e+02
c	270	14	4.7	3024	9	Q57397	Human ICAM-1. 2.03e+02
c	271	14	4.7	3024	28	T47957	Human ICAM-1 (HL-60 c 2.03e+02
c	272	14	4.7	3024	29	T62835	Human ICAM-1 cDNA. 2.03e+02
	273	14	4.7	3045	34	T73866	Cotton fibre promoter 2.03e+02
c	274	14	4.7	3113	1	N80995	HL-cDNA clone encodin 2.03e+02
	275	14	4.7	3281	8	Q47743	Truncated IBV spike p 2.03e+02
c	276	14	4.7	3440	40	V15268	Brassica campestris B 2.03e+02
c	277	14	4.7	3440	3	Q13967	Bce-4 gene with regul 2.03e+02
c	278	14	4.7	3467	19	T12324	DNA encoding Sulfolob 2.03e+02
	279	14	4.7	3489	25	T18211	Major neutralising an 2.03e+02
	280	14	4.7	3504	1	Q05304	Clone Imd2 encoding i 2.03e+02
c	281	14	4.7	3528	21	T31824	Tomato pZ130 genomic 2.03e+02
c	282	14	4.7	3528	26	T48816	Calgene Lambda 140 ge 2.03e+02
	283	14	4.7	3545	2	N60680	Sequence encoding spi 2.03e+02
	284	14	4.7	3600	2	N70196	Sequence encoding a r 2.03e+02
	285	14	4.7	3645	2	N60679	Sequence encoding S1 2.03e+02
	286	14	4.7	3710	2	Q12164	Sequence encoding Hum 2.03e+02
	287	14	4.7	3797	6	Q36866	Coleoptera toxin from 2.03e+02
	288	14	4.7	3797	10	Q58975	B.thuringiensis serov 2.03e+02
	289	14	4.7	3848	36	T86133	Maize protoporphyrino 2.03e+02
	290	14	4.7	3848	38	V04311	Maize protox-1 promot 2.03e+02
	291	14	4.7	4032	4	Q28506	Tomato HCD promoter (2.03e+02
	292	14	4.7	4032	6	Q40416	DNA clone of lambdaUC 2.03e+02
	293	14	4.7	4041	20	T13944	VIP2A(a) and VIP1A(a) 2.03e+02
	294	14	4.7	4041	39	V16172	DNA sequence encoding 2.03e+02
	295	14	4.7	4041	38	T73999	B. cereus VIP1A(a)/VI 2.03e+02
c	296	14	4.7	4054	1	Q06869	Sequence encodes Plat 2.03e+02
c	297	14	4.7	4060	15	Q90768	BAV3 genome (left 11% 2.03e+02

c	298	14	4.7	4098	24	T43682	Medium chain-specific 2.03e+02
	299	14	4.7	4140	1	N92735	Sheep PrP gene for sc 2.03e+02
c	300	14	4.7	4383	2	Q10263	pZ130 contg. Calgene 2.03e+02
c	301	14	4.7	4383	6	Q34941	Calgene Lambda 140 ge 2.03e+02
c	302	14	4.7	4383	2	Q10319	Calgene lambda 140 ge 2.03e+02
c	303	14	4.7	4383	6	Q35143	Calgene lambda 140/pZ 2.03e+02
c	304	14	4.7	4487	2	N60943	Sequence encoding glu 2.03e+02
c	305	14	4.7	4526	8	Q49780	Prosystemin. 2.03e+02
	306	14	4.7	4656	1	N80747	Sequence of 2A11 geno 2.03e+02
c	307	14	4.7	4677	24	T49535	Human parvovirus geno 2.03e+02
c	308	14	4.7	4704	1	Q05264	Sequence encoding bea 2.03e+02
c	309	14	4.7	5361	32	T78868	P. falciparum liver s 2.03e+02
	310	14	4.7	5502	24	T30092	Neuronal apoptosis in 2.03e+02
	311	14	4.7	5502	32	T71263	Neuronal apoptosis in 2.03e+02
	312	14	4.7	5778	3	N40033	Sequence of portion o 2.03e+02
	313	14	4.7	6049	39	V16165	DNA encoding vegetati 2.03e+02
	314	14	4.7	6049	38	T73993	B. cereus VIP2A(a) co 2.03e+02
	315	14	4.7	6049	20	T13939	B. cereus VIP2A(a) ge 2.03e+02
c	316	14	4.7	6075	4	Q27451	Type A human platelet 2.03e+02
	317	14	4.7	6106	13	Q74678	Bacillus cereus vegat 2.03e+02
	318	14	4.7	6124	32	T71265	Neuronal apoptosis in 2.03e+02
	319	14	4.7	6133	32	T71264	Neuronal apoptosis in 2.03e+02
c	320	14	4.7	6152	32	T78867	P. falciparum liver s 2.03e+02
	321	14	4.7	6228	32	T71266	Neuronal apoptosis in 2.03e+02
c	322	14	4.7	6311	5	Q29965	Sugar beet chitinase 2.03e+02
	323	14	4.7	6371	22	T30194	Peripheral nervous sy 2.03e+02
	324	14	4.7	6404	22	T30195	Peripheral nervous sy 2.03e+02
c	325	14	4.7	6412	1	Q05989	TR4 cDNA clone relate 2.03e+02
	326	14	4.7	6471	18	T06306	Tomato pathogen resis 2.03e+02
	327	14	4.7	6480	5	Q29715	Partial sequence of h 2.03e+02
c	328	14	4.7	6763	34	T79728	Beta-lactoglobulin cl 2.03e+02
	329	14	4.7	6785	3	Q20065	Plasmid pAZ112. 2.03e+02
	330	14	4.7	7475	3	Q15312	Plasmid pHPB12 encodi 2.03e+02
	331	14	4.7	7753	4	Q26727	Encodes exons XV to X 2.03e+02
	332	14	4.7	8355	24	T35524	Human interferon-beta 2.03e+02
	333	14	4.7	8738	32	T72327	Lactobacillus bacteri 2.03e+02
	334	14	4.7	8841	14	Q89299	Tryptophan-2,3-dioxyg 2.03e+02
	335	14	4.7	8959	14	Q55133	Staphylococcus aureus 2.03e+02
	336	14	4.7	9295	29	T62125	Arabidopsis thaliana 2.03e+02
	337	14	4.7	9739	27	T58549	Human Immunodeficienc 2.03e+02
	338	14	4.7	9739	3	Q14751	HIV-1(MN) env protein 2.03e+02
	339	14	4.7	9746	3	Q14752	HIV-1(MN-ST1) env pro 2.03e+02
	340	14	4.7	9746	27	T58550	Human Immunodeficienc 2.03e+02
	341	14	4.7	10614	15	Q89555	Hamster cholesterol 7 2.03e+02
c	342	14	4.7	10807	34	T79725	Ovine beta-lactoglobu 2.03e+02
c	343	14	4.7	10807	22	T03855	Sheep beta-lactoglobu 2.03e+02
	344	14	4.7	11225	2	N91695	Sequence of fowlpox v 2.03e+02
	345	14	4.7	12001	13	Q76213	HSV L/ST region. 2.03e+02
c	346	14	4.7	12720	10	Q64211	L.lactis branched ami 2.03e+02
c	347	14	4.7	14507	32	T73568	Expression augmenting 2.03e+02
c	348	14	4.7	14602	31	T73236	Streptococcus thermop 2.03e+02

c	349	14	4.7	18177	1	N90490	DNA of human retinobl	2.03e+02
	350	14	4.7	22481	24	T11658	PEDF full length sequ	2.03e+02
c	351	14	4.7	53526	40	T94101	Human PKD1 gene.	2.03e+02
c	352	14	4.7	53577	40	T94108	Human PKD1 locus betw	2.03e+02
c	353	14	4.7	53577	28	T18551	Human polycystic kidn	2.03e+02
	354	14	4.7	133894	17	T13635	AcNPV genomic DNA clo	2.03e+02
	355	14	4.7	580073	27	T58840	Mycoplasma genitalium	2.03e+02
	356	13	4.4	18	2	Q12436	Primer HPU2 for PCR a	6.75e+02
	357	13	4.4	21	24	T30213	Antisense oligonucleo	6.75e+02
	358	13	4.4	21	15	T01741	Peptide Nucleic acid	6.75e+02
	359	13	4.4	21	3	Q22631	Antisense oligonucleo	6.75e+02
	360	13	4.4	25	38	V05637	Primer for test of hu	6.75e+02
c	361	13	4.4	28	11	Q66059	Japanese cedar pollen	6.75e+02
c	362	13	4.4	30	21	T32201	Yeast cAMP-dependent	6.75e+02
	363	13	4.4	40	6	Q38045	Oligonucleotide gel2,	6.75e+02
c	364	13	4.4	40	6	Q38080	Oligonucleotide gel37	6.75e+02
c	365	13	4.4	50	22	T25074	Human gene signature	6.75e+02
c	366	13	4.4	64	19	T22103	Human gene signature	6.75e+02
	367	13	4.4	66	2	N60756	Conserved sequence co	6.75e+02
c	368	13	4.4	68	19	T21591	Human gene signature	6.75e+02
c	369	13	4.4	69	10	Q44415	Special AT-rich (ATC)	6.75e+02
	370	13	4.4	73	6	Q35435	PCR primer C6B18G.	6.75e+02
	371	13	4.4	73	40	V17710	Avipox virus ALVAC DN	6.75e+02
c	372	13	4.4	84	19	T20757	Human gene signature	6.75e+02
c	373	13	4.4	86	25	T14776	C. coli G1-G3 region	6.75e+02
c	374	13	4.4	86	25	T14781	C. coli G1-G3 region	6.75e+02
c	375	13	4.4	97	29	T33723	Cellubrevin-1 partial	6.75e+02
c	376	13	4.4	102	21	T22566	Human gene signature	6.75e+02
c	377	13	4.4	109	22	T24833	Human gene signature	6.75e+02
c	378	13	4.4	111	19	T19769	Human gene signature	6.75e+02
c	379	13	4.4	116	20	T22044	Human gene signature	6.75e+02
c	380	13	4.4	117	25	T14777	C. coli G1-G3 region	6.75e+02
c	381	13	4.4	117	25	T14778	C. coli G1-G3 region	6.75e+02
c	382	13	4.4	117	25	T14779	C. coli G1-G3 region	6.75e+02
c	383	13	4.4	117	25	T14780	C. coli G1-G3 region	6.75e+02
c	384	13	4.4	122	21	T25427	Human gene signature	6.75e+02
	385	13	4.4	128	20	T22326	Human gene signature	6.75e+02
	386	13	4.4	129	21	T26553	Human gene signature	6.75e+02
	387	13	4.4	129	20	T22325	Human gene signature	6.75e+02
c	388	13	4.4	129	12	Q71384	DNA probe (AUTO 2) co	6.75e+02
	389	13	4.4	130	22	T25033	Human gene signature	6.75e+02
	390	13	4.4	131	22	T24215	Human gene signature	6.75e+02
c	391	13	4.4	134	23	T26803	Human gene signature	6.75e+02
c	392	13	4.4	139	19	T23860	Human gene signature	6.75e+02
c	393	13	4.4	143	20	T21482	Human gene signature	6.75e+02
c	394	13	4.4	145	19	T20287	Human gene signature	6.75e+02
	395	13	4.4	159	13	Q81767	Japanese Black Pine c	6.75e+02
	396	13	4.4	159	26	T40825	Serine protease nfSP7	6.75e+02
c	397	13	4.4	161	8	Q59750	Human brain Expressed	6.75e+02
c	398	13	4.4	170	20	T23494	Human gene signature	6.75e+02
c	399	13	4.4	171	19	T20237	Human gene signature	6.75e+02

	400	13	4.4	172	7	Q41241	3'-5' micF RNA strand	6.75e+02
	401	13	4.4	174	3	Q20669	micF RNA.	6.75e+02
	402	13	4.4	174	6	Q37973	micF-RNA.	6.75e+02
	403	13	4.4	174	9	Q53309	MicF-RNA complementar	6.75e+02
c	404	13	4.4	176	19	T20216	Human gene signature	6.75e+02
	405	13	4.4	180	20	T23036	Human gene signature	6.75e+02
	406	13	4.4	180	38	T91286	Lactococcal promoter	6.75e+02
c	407	13	4.4	186	34	T67149	P gene encoding sakac	6.75e+02
	408	13	4.4	191	20	T21962	Human gene signature	6.75e+02
	409	13	4.4	191	19	T21422	Human gene signature	6.75e+02
c	410	13	4.4	194	1	N90211	Malaria-specific DNA	6.75e+02
c	411	13	4.4	197	40	V20650	Sweet potato latent v	6.75e+02
	412	13	4.4	205	4	Q31930	Human scu-PA54 gene f	6.75e+02
c	413	13	4.4	211	10	Q76850	Human genome fragment	6.75e+02
	414	13	4.4	221	8	Q59563	Human brain Expressed	6.75e+02
c	415	13	4.4	226	40	V09976	Partial cDNA Incyte c	6.75e+02
	416	13	4.4	236	40	V09969	Partial cDNA Incyte c	6.75e+02
c	417	13	4.4	238	22	T22970	Human gene signature	6.75e+02
	418	13	4.4	242	13	Q80696	AcMNPV 6.9K gene 3' U	6.75e+02
	419	13	4.4	249	10	Q76653	Human genome fragment	6.75e+02
	420	13	4.4	251	40	V09952	Partial cDNA Incyte c	6.75e+02
	421	13	4.4	255	37	V02834	Human RHAMM genomic D	6.75e+02
	422	13	4.4	256	40	V09960	Partial cDNA Incyte c	6.75e+02
c	423	13	4.4	258	29	T67368	H. pylori secreted or	6.75e+02
	424	13	4.4	259	40	V09975	Partial cDNA Incyte c	6.75e+02
c	425	13	4.4	259	19	T20082	Human gene signature	6.75e+02
	426	13	4.4	263	40	V09974	Partial cDNA Incyte c	6.75e+02
	427	13	4.4	264	30	T67535	H. pylori flagella-as	6.75e+02
c	428	13	4.4	265	40	V09981	Partial cDNA Incyte c	6.75e+02
c	429	13	4.4	266	19	T20590	Human gene signature	6.75e+02
	430	13	4.4	268	10	Q76789	Human genome fragment	6.75e+02
c	431	13	4.4	271	19	T21518	Human gene signature	6.75e+02
c	432	13	4.4	275	23	T26212	Human gene signature	6.75e+02
c	433	13	4.4	276	22	T35828	M6A1-2.9 partial sequ	6.75e+02
c	434	13	4.4	276	7	Q48729	M. penetrans clone M6	6.75e+02
	435	13	4.4	279	40	V09962	Partial cDNA Incyte c	6.75e+02
	436	13	4.4	279	30	T68201	H. pylori flagella as	6.75e+02
c	437	13	4.4	279	30	T68015	H. pylori cell envelo	6.75e+02
c	438	13	4.4	280	40	V09979	Partial cDNA Incyte c	6.75e+02
	439	13	4.4	281	40	V09963	Partial cDNA Incyte c	6.75e+02
c	440	13	4.4	289	8	Q61330	Human brain Expressed	6.75e+02
c	441	13	4.4	289	21	T24606	Human gene signature	6.75e+02
c	442	13	4.4	292	1	Q03819	Marek's disease virus	6.75e+02
c	443	13	4.4	293	20	T22671	Human gene signature	6.75e+02
c	444	13	4.4	296	40	V04278	Secreted protein C195	6.75e+02
c	445	13	4.4	296	20	T21886	Human gene signature	6.75e+02
c	446	13	4.4	297	18	T20994	Human gene signature	6.75e+02
c	447	13	4.4	305	22	T18872	Genotyping marker C16	6.75e+02
c	448	13	4.4	305	19	T11257	Microsatellite marker	6.75e+02
c	449	13	4.4	305	22	T28268	Survival motor neuron	6.75e+02
	450	13	4.4	308	21	T24706	Human gene signature	6.75e+02

c	451	13	4.4	308	22	T24855	Human gene signature	6.75e+02
	452	13	4.4	311	20	T23599	Human gene signature	6.75e+02
	453	13	4.4	312	19	T09787	Tomato genomic DNA, D	6.75e+02
	454	13	4.4	318	30	T67683	H. pylori transmembra	6.75e+02
	455	13	4.4	320	22	T26465	Human gene signature	6.75e+02
	456	13	4.4	325	40	V09955	Partial cDNA Incyte c	6.75e+02
	457	13	4.4	325	8	Q59675	Human brain Expressed	6.75e+02
c	458	13	4.4	328	22	T25084	Human gene signature	6.75e+02
	459	13	4.4	333	21	T13697	AcNPV ORF 107, residu	6.75e+02
	460	13	4.4	343	20	T24165	Human gene signature	6.75e+02
	461	13	4.4	348	8	Q59950	Human brain Expressed	6.75e+02
	462	13	4.4	358	19	T19973	Human gene signature	6.75e+02
c	463	13	4.4	364	19	T20533	Human gene signature	6.75e+02
	464	13	4.4	379	22	T26329	Human gene signature	6.75e+02
c	465	13	4.4	382	8	Q60162	Human brain Expressed	6.75e+02
c	466	13	4.4	383	8	Q60217	Human brain Expressed	6.75e+02
	467	13	4.4	386	24	T34488	Neurite cDNA clone (C	6.75e+02
	468	13	4.4	386	20	T12193	pJG4-5-CDX-BP cDNA cl	6.75e+02
c	469	13	4.4	389	21	T24720	Human gene signature	6.75e+02
c	470	13	4.4	390	8	Q60000	Human brain Expressed	6.75e+02
	471	13	4.4	392	27	T13838	Canarypox C6 locus le	6.75e+02
	472	13	4.4	395	24	T34489	Neurite cDNA clone (C	6.75e+02
	473	13	4.4	396	37	T92842	Plea saliva protein f	6.75e+02
	474	13	4.4	400	11	Q64457	A/Okuda/57 human infl	6.75e+02
	475	13	4.4	400	12	Q72836	HA gene fragment of A	6.75e+02
c	476	13	4.4	403	8	Q60573	Human brain Expressed	6.75e+02
c	477	13	4.4	407	19	T20521	Human gene signature	6.75e+02
	478	13	4.4	409	11	Q64458	A/Adachi/2/57 human i	6.75e+02
	479	13	4.4	410	12	Q72838	HA gene fragment of A	6.75e+02
	480	13	4.4	411	8	Q60059	Human brain Expressed	6.75e+02
c	481	13	4.4	418	2	Q11836	TRFP chain 1 with lea	6.75e+02
c	482	13	4.4	422	7	Q41556	TRFP chain 1 (with Le	6.75e+02
c	483	13	4.4	422	8	Q49533	Human T cell reactive	6.75e+02
c	484	13	4.4	423	19	T22118	Human gene signature	6.75e+02
c	485	13	4.4	428	7	Q41557	TRFP chain 1 (with Le	6.75e+02
	486	13	4.4	431	19	T19913	Human gene signature	6.75e+02
c	487	13	4.4	442	12	Q72832	HA gene fragment of A	6.75e+02
c	488	13	4.4	464	3	Q20988	CDR-grafted, humanise	6.75e+02
c	489	13	4.4	464	4	Q27355	Encodes A5B7 gH-2 ant	6.75e+02
c	490	13	4.4	469	19	T20502	Human gene signature	6.75e+02
c	491	13	4.4	474	8	Q47371	Sequence encoding hae	6.75e+02
c	492	13	4.4	474	12	Q70203	Sequence encoding imm	6.75e+02
c	493	13	4.4	474	31	T43493	ATM gene exon 61.	6.75e+02
c	494	13	4.4	500	3	N50239	Sequence of the promo	6.75e+02
c	495	13	4.4	500	6	Q37972	Promoter and upstream	6.75e+02
c	496	13	4.4	500	9	Q53308	Promoter and upstream	6.75e+02
c	497	13	4.4	500	7	Q41240	Promoter and upstream	6.75e+02
	498	13	4.4	504	29	T67378	H. pylori surface mem	6.75e+02
c	499	13	4.4	521	1	Q06473	Human erythropoietin	6.75e+02
c	500	13	4.4	531	12	Q73904	Borrelia BITS antigen	6.75e+02
	501	13	4.4	549	11	Q66815	AmEPV thymidine-kinas	6.75e+02

	502	13	4.4	562	6	Q39633	Expressed Sequence Ta	6.75e+02
	503	13	4.4	562	8	Q59045	Human brain Expressed	6.75e+02
c	504	13	4.4	563	31	T68937	Mouse p32 11-cis-reti	6.75e+02
c	505	13	4.4	574	40	V15214	Human osteogenic prot	6.75e+02
c	506	13	4.4	574	18	T02603	Partial human CBMP3 D	6.75e+02
c	507	13	4.4	574	12	Q72701	Mature human CBMP3 pa	6.75e+02
c	508	13	4.4	574	9	Q53151	Partial coding sequen	6.75e+02
c	509	13	4.4	576	12	Q73857	Borrelia 28691 OspC a	6.75e+02
c	510	13	4.4	582	12	Q73879	Borrelia BITS antigen	6.75e+02
	511	13	4.4	592	24	T41695	Dimeric glycoprotein	6.75e+02
	512	13	4.4	592	2	Q10075	Engineered human alph	6.75e+02
c	513	13	4.4	630	8	Q47367	Sequence encoding H1	6.75e+02
c	514	13	4.4	630	12	Q70199	Sequence encoding imm	6.75e+02
	515	13	4.4	638	37	T80781	Staphylococcus aureus	6.75e+02
c	516	13	4.4	642	28	T58360	Tobacco EAS4 gene pro	6.75e+02
	517	13	4.4	654	30	T67668	H. pylori transmembra	6.75e+02
c	518	13	4.4	670	12	Q70206	Influenza haemaggluti	6.75e+02
c	519	13	4.4	670	8	Q47360	Sequence encoding the	6.75e+02
	520	13	4.4	693	40	V24830	H. pylori ORF hp3e103	6.75e+02
c	521	13	4.4	707	38	V06458	Osex sequence used fo	6.75e+02
c	522	13	4.4	717	8	Q47368	Sequence encoding hae	6.75e+02
	523	13	4.4	731	1	Q03843	Bovine alpha subunit	6.75e+02
	524	13	4.4	750	30	T67953	H. pylori transmembra	6.75e+02
c	525	13	4.4	770	14	Q85766	Operon encoding prema	6.75e+02
	526	13	4.4	774	35	T79673	BRCA2 cancer susceptI	6.75e+02
c	527	13	4.4	777	24	T42509	Humanised A5B57 Fd fr	6.75e+02
	528	13	4.4	783	6	Q38041	Synthetic gelonin gen	6.75e+02
c	529	13	4.4	789	30	T69950	FIPV matrix (M) gene	6.75e+02
c	530	13	4.4	789	17	Q89649	FIPV E1 protein codin	6.75e+02
	531	13	4.4	797	13	Q80718	A. australis cuticle-	6.75e+02
c	532	13	4.4	801	18	T04703	p2-60-HIV.3 fragment	6.75e+02
c	533	13	4.4	810	12	Q70198	Sequence encoding imm	6.75e+02
	534	13	4.4	870	17	Q98473	Carboxy terminus of M	6.75e+02
	535	13	4.4	877	29	T62114	Listeria monocytogene	6.75e+02
c	536	13	4.4	879	32	T79940	Chitinase coding sequ	6.75e+02
c	537	13	4.4	880	7	Q43675	Sequence of intron 25	6.75e+02
c	538	13	4.4	885	6	Q37679	Encodes chinese cucum	6.75e+02
c	539	13	4.4	886	28	T61376	Chinese cucumber alph	6.75e+02
	540	13	4.4	894	26	T40849	Serine protease nfSP7	6.75e+02
c	541	13	4.4	912	8	Q47370	Sequence encoding hae	6.75e+02
c	542	13	4.4	924	1	Q04072	Coding region for C13	6.75e+02
c	543	13	4.4	924	8	Q47364	Sequence encoding C13	6.75e+02
c	544	13	4.4	936	34	T69159	Trypanosoma cruzi ant	6.75e+02
	545	13	4.4	946	21	T10417	Rhizopus niveus lipas	6.75e+02
c	546	13	4.4	966	8	Q49561	Beet Cyst Nematode Re	6.75e+02
	547	13	4.4	982	34	T65039	Staphylococcus aureus	6.75e+02
c	548	13	4.4	992	7	Q40491	Clone FT36 containing	6.75e+02
c	549	13	4.4	993	4	Q24207	Transglutaminase (exp	6.75e+02
c	550	13	4.4	997	2	N70951	Sequence of enhancer	6.75e+02
c	551	13	4.4	998	32	T79941	Full length chitinase	6.75e+02
	552	13	4.4	1011	1	Q05916	Promoter sequence and	6.75e+02

c	553	13	4.4	1011	17	Q94312	Tyrosine phosphatase 6.75e+02
c	554	13	4.4	1029	10	Q55186	MS-PO23 Vector. 6.75e+02
c	555	13	4.4	1059	23	T29076	Tyrosinase gene fragm 6.75e+02
c	556	13	4.4	1059	23	T29077	Tyrosinase gene fragm 6.75e+02
c	557	13	4.4	1076	40	V20649	Sweet potato latent v 6.75e+02
c	558	13	4.4	1098	13	Q75405	Ovalbumin oval-f gene 6.75e+02
c	559	13	4.4	1104	12	Q72616	Insecticidal patatin 6.75e+02
c	560	13	4.4	1106	12	Q72615	Insecticidal patatin 6.75e+02
c	561	13	4.4	1106	12	Q72612	Insecticidal patatin 6.75e+02
c	562	13	4.4	1106	12	Q72610	Insecticidal patatin 6.75e+02
c	563	13	4.4	1109	3	Q20728	Murine macrophage inf 6.75e+02
c	564	13	4.4	1109	3	Q20612	Murine macrophage inf 6.75e+02
c	565	13	4.4	1109	1	Q03852	Sequence encoding mac 6.75e+02
	566	13	4.4	1135	12	Q72806	DNA encoding stem reg 6.75e+02
c	567	13	4.4	1136	2	N70435	Sequence encoding ins 6.75e+02
c	568	13	4.4	1158	13	Q75406	Ovalbumin Yoval-f gen 6.75e+02
c	569	13	4.4	1171	12	Q72606	Insecticidal patatin 6.75e+02
c	570	13	4.4	1172	12	Q72621	Insecticidal patatin 6.75e+02
c	571	13	4.4	1172	12	Q72617	Insecticidal patatin 6.75e+02
c	572	13	4.4	1172	12	Q72611	Insecticidal patatin 6.75e+02
c	573	13	4.4	1172	12	Q72620	Insecticidal patatin 6.75e+02
c	574	13	4.4	1175	12	Q72618	Insecticidal patatin 6.75e+02
c	575	13	4.4	1183	11	Q63613	T-cell transmembrane 6.75e+02
c	576	13	4.4	1183	3	Q14534	Lov gene (cDNA 19.5). 6.75e+02
	577	13	4.4	1191	24	T35208	Sequence encoding cyt 6.75e+02
c	578	13	4.4	1200	9	Q52267	Sequence encoding 5S 6.75e+02
	579	13	4.4	1204	1	Q06200	PR-2 cDNA cloned into 6.75e+02
c	580	13	4.4	1206	17	T09352	Aureobasidin resistan 6.75e+02
c	581	13	4.4	1206	17	T09351	Aureobasidin resistan 6.75e+02
c	582	13	4.4	1215	28	T58339	Tobacco epi-5-aristol 6.75e+02
c	583	13	4.4	1227	5	Q31826	Recombinant TCS seque 6.75e+02
c	584	13	4.4	1227	1	Q06343	Encodes Trichosanthin 6.75e+02
c	585	13	4.4	1228	4	Q26499	Trichosanthin coding 6.75e+02
	586	13	4.4	1229	3	Q25012	KL-S1_d gene. 6.75e+02
c	587	13	4.4	1230	10	Q55984	BTG-contg. sequence. 6.75e+02
c	588	13	4.4	1230	1	N90223	Malaria-specific DNA 6.75e+02
	589	13	4.4	1248	10	Q44396	Farnesyltransferase b 6.75e+02
	590	13	4.4	1248	30	T38711	Human farnesyl transf 6.75e+02
	591	13	4.4	1248	16	Q94413	Human farnesyl protei 6.75e+02
	592	13	4.4	1260	3	Q25011	KL-2 gene. 6.75e+02
	593	13	4.4	1278	30	T67906	H. pylori surface or 6.75e+02
c	594	13	4.4	1297	3	Q21645	3' coding sequence of 6.75e+02
	595	13	4.4	1314	10	Q64888	cDNA encoding beta su 6.75e+02
	596	13	4.4	1316	30	T63621	Interleukin-12 p35 su 6.75e+02
	597	13	4.4	1316	2	Q10282	Structural gene encod 6.75e+02
	598	13	4.4	1320	40	V17564	Coding sequence for t 6.75e+02
	599	13	4.4	1321	38	T85309	Saccharomyces cerevis 6.75e+02
c	600	13	4.4	1339	12	Q68892	Patatin cDNA insert i 6.75e+02
c	601	13	4.4	1340	28	T62662	cDNA encoding murine 6.75e+02
	602	13	4.4	1344	3	Q22340	KL-1 gene. 6.75e+02
	603	13	4.4	1344	9	Q51788	c-kit (proto-neogene) 6.75e+02

c	604	13	4.4	1356	15	Q90035	Human SMP30 gene. 6.75e+02
	605	13	4.4	1360	3	Q20842	Murine mast cell grow 6.75e+02
	606	13	4.4	1364	40	V15228	Heterodimeric human i 6.75e+02
	607	13	4.4	1364	26	T48099	Human interleukin-12 6.75e+02
c	608	13	4.4	1380	12	Q68891	Patatin cDNA insert i 6.75e+02
c	609	13	4.4	1384	23	T33136	Broccoli ACC synthase 6.75e+02
	610	13	4.4	1396	4	Q24371	Natural killer cell s 6.75e+02
c	611	13	4.4	1408	22	T28260	Survival motor neuron 6.75e+02
c	612	13	4.4	1408	22	T28258	Survival motor neuron 6.75e+02
c	613	13	4.4	1415	17	T03468	H6-promoted CHV gD ge 6.75e+02
	614	13	4.4	1418	17	Q98238	Promoter 1 of rice py 6.75e+02
	615	13	4.4	1440	19	Q79623	Morning Glory dihydro 6.75e+02
	616	13	4.4	1442	28	T39484	Human steroidogenesis 6.75e+02
	617	13	4.4	1443	3	Q14943	Xylose isomerase gene 6.75e+02
	618	13	4.4	1450	3	Q20843	Murine mast cell grow 6.75e+02
c	619	13	4.4	1452	15	Q88226	Smallpox virus strain 6.75e+02
c	620	13	4.4	1455	17	T00777	Bacillus sp. alkaline 6.75e+02
	621	13	4.4	1477	1	N81161	DNA encoding modified 6.75e+02
c	622	13	4.4	1479	22	T09696	Caenorhabditis elegan 6.75e+02
c	623	13	4.4	1481	24	T31854	Nucleic acid encoding 6.75e+02
c	624	13	4.4	1491	22	T31100	Papillomavirus E2 bin 6.75e+02
c	625	13	4.4	1493	40	V09951	cDNA encoding a novel 6.75e+02
	626	13	4.4	1494	40	V03305	Thermococcus phosphat 6.75e+02
	627	13	4.4	1510	22	T29628	DNA mismatch repair p 6.75e+02
	628	13	4.4	1511	39	V14508	AmEPV entomopoxvirus 6.75e+02
	629	13	4.4	1511	11	Q66798	AmEPV thymidine-kinas 6.75e+02
	630	13	4.4	1518	24	T37313	Aromatic acyl transfe 6.75e+02
c	631	13	4.4	1518	24	T37313	Aromatic acyl transfe 6.75e+02
c	632	13	4.4	1546	2	Q10864	Sequence encoding hum 6.75e+02
c	633	13	4.4	1558	22	T28255	Survival motor neuron 6.75e+02
c	634	13	4.4	1565	16	T02334	GABA-A receptor gamma 6.75e+02
c	635	13	4.4	1582	22	T18831	Human survival motor 6.75e+02
c	636	13	4.4	1582	22	T28259	Survival motor neuron 6.75e+02
c	637	13	4.4	1587	23	T29078	Tyrosinase gene fragm 6.75e+02
	638	13	4.4	1588	1	Q04926	Sequence encoding por 6.75e+02
	639	13	4.4	1588	5	Q28540	DNA for bovine sol. s 6.75e+02
	640	13	4.4	1590	38	V04317	Sorghum protox-1 cDNA 6.75e+02
c	641	13	4.4	1612	2	N60392	Sequence encoding the 6.75e+02
c	642	13	4.4	1614	12	Q67875	C6 insertion plasmid 6.75e+02
	643	13	4.4	1618	38	T96835	Intron 5 of human pRb 6.75e+02
c	644	13	4.4	1641	22	T27036	Arabidopsis fatty aci 6.75e+02
c	645	13	4.4	1644	1	Q03257	Luciola lateralis rec 6.75e+02
c	646	13	4.4	1644	6	Q34745	Thermostable lucifera 6.75e+02
c	647	13	4.4	1648	4	Q25532	Sequence of genomic c 6.75e+02
c	648	13	4.4	1659	21	T31800	Maize adenylosuccinat 6.75e+02
c	649	13	4.4	1664	15	Q90218	Condensing enzyme clo 6.75e+02
c	650	13	4.4	1667	4	Q33027	Exon 5 from the UGT1 6.75e+02
	651	13	4.4	1689	3	Q22205	Murine mast cell grow 6.75e+02
	652	13	4.4	1696	12	Q71015	lyam-1 gene exon X. 6.75e+02
	653	13	4.4	1698	30	T68198	H. pylori inner membr 6.75e+02
c	654	13	4.4	1704	29	T63268	Firefly luciferase co 6.75e+02

	655	13	4.4	1722	3	Q14716	Nitrosomonas SPN 366. 6.75e+02
c	656	13	4.4	1758	1	Q06094	Reverse transcriptase 6.75e+02
c	657	13	4.4	1761	17	T01407	H6-promoted CHV gC ge 6.75e+02
c	658	13	4.4	1770	6	Q37684	P.falciparum antigeni 6.75e+02
c	659	13	4.4	1770	6	Q37124	P.falciparum antigen. 6.75e+02
	660	13	4.4	1776	14	Q86521	Heliothis virescens r 6.75e+02
c	661	13	4.4	1777	2	Q06842	RNA segment 4 coding 6.75e+02
c	662	13	4.4	1780	40	V04269	Rat FRAG1 coding sequ 6.75e+02
	663	13	4.4	1783	12	Q72849	Full length HA gene o 6.75e+02
c	664	13	4.4	1799	31	T59214	Influenza B/Panama/45 6.75e+02
c	665	13	4.4	1803	13	Q75407	Ovalbumin pl+oval-f+3 6.75e+02
	666	13	4.4	1812	3	Q20953	Fungal 1,3-specific l 6.75e+02
	667	13	4.4	1815	17	Q97809	Sequence encoding end 6.75e+02
	668	13	4.4	1818	40	V17562	Coding sequence for t 6.75e+02
c	669	13	4.4	1824	7	Q45502	Sequence encoding an 6.75e+02
c	670	13	4.4	1833	24	T40341	Native Btt gene encod 6.75e+02
	671	13	4.4	1858	7	Q46334	rolC gene. 6.75e+02
	672	13	4.4	1858	11	Q62088	rolC gene. 6.75e+02
c	673	13	4.4	1858	7	Q46334	rolC gene. 6.75e+02
c	674	13	4.4	1858	11	Q62088	rolC gene. 6.75e+02
	675	13	4.4	1860	2	Q11780	Gene encoding lipase 6.75e+02
	676	13	4.4	1869	23	T37413	Prostate tumour induc 6.75e+02
	677	13	4.4	1875	36	V01454	A. thaliana inorganic 6.75e+02
c	678	13	4.4	1876	12	Q71053	Rabbit C7-alpha-H cor 6.75e+02
c	679	13	4.4	1886	1	N80637	Encodes human tyrosin 6.75e+02
c	680	13	4.4	1894	16	T01009	Human tyrosinase acti 6.75e+02
c	681	13	4.4	1894	2	Q06552	Sequence of human tyr 6.75e+02
c	682	13	4.4	1906	12	Q66548	Human tyrosinase DNA. 6.75e+02
c	683	13	4.4	1908	29	T63269	Gene for firefly luci 6.75e+02
c	684	13	4.4	1910	22	T35901	Tyrosinase melanoma a 6.75e+02
	685	13	4.4	1920	40	V17563	Coding sequence for t 6.75e+02
c	686	13	4.4	1920	36	T95733	Normal human tyrosina 6.75e+02
c	687	13	4.4	1932	8	Q47288	Delta endotoxin gene. 6.75e+02
	688	13	4.4	1954	7	Q45598	Sequence of the mouse 6.75e+02
c	689	13	4.4	1962	15	Q90252	Babesia merozoite p58 6.75e+02
	690	13	4.4	1975	1	Q05057	DraI/DraI fragment of 6.75e+02
c	691	13	4.4	1975	1	Q05057	DraI/DraI fragment of 6.75e+02
	692	13	4.4	1981	1	Q04267	Probe containing codi 6.75e+02
c	693	13	4.4	1981	1	Q04267	Probe containing codi 6.75e+02
c	694	13	4.4	1990	6	Q36662	Full length GPOl cDNA 6.75e+02
c	695	13	4.4	2019	23	T33850	DNA encoding streptav 6.75e+02
	696	13	4.4	2020	36	T89953	Nicotiana sp. extrace 6.75e+02
c	697	13	4.4	2022	21	T16016	DNA encoding ubiquiti 6.75e+02
	698	13	4.4	2028	8	Q51230	Human adrenocorticotr 6.75e+02
c	699	13	4.4	2055	23	T33851	DNA encoding mutant s 6.75e+02
c	700	13	4.4	2056	20	T12180	Partial pJG4-5-CDK-BP 6.75e+02
c	701	13	4.4	2075	18	T04226	Connective tissue gro 6.75e+02
c	702	13	4.4	2075	28	T45360	Human connective tiss 6.75e+02
	703	13	4.4	2088	22	T36895	Candida albicans leuk 6.75e+02
	704	13	4.4	2092	1	Q04091	2.1 kb XbaI-BamHI fra 6.75e+02
	705	13	4.4	2101	21	T31451	Drosophila Ndr serine 6.75e+02

	706	13	4.4	2110	7	Q41223	Clone PC3 encoding ma 6.75e+02
c	707	13	4.4	2121	35	T89348	Bacillus popilliae cr 6.75e+02
c	708	13	4.4	2127	21	T12414	Ubi7 ubiquitin-lytic 6.75e+02
	709	13	4.4	2128	23	T37405	Prostate tumour induc 6.75e+02
c	710	13	4.4	2133	3	Q14966	Varicella Zoster Viru 6.75e+02
c	711	13	4.4	2155	12	Q72269	Rat REC17 serotonin r 6.75e+02
c	712	13	4.4	2156	1	N82202	Sequence of HaeII rep 6.75e+02
c	713	13	4.4	2173	13	Q76108	Cucumber mosaic virus 6.75e+02
c	714	13	4.4	2176	40	V17698	Recombinant avipox vi 6.75e+02
c	715	13	4.4	2198	39	V05124	Fusion protein gene o 6.75e+02
	716	13	4.4	2263	15	Q91108	Hamster ileal/renal b 6.75e+02
c	717	13	4.4	2307	6	Q34672	P39-alpha and P39-bet 6.75e+02
c	718	13	4.4	2337	11	Q68669	PMGA1.2 Mycoplasma ga 6.75e+02
c	719	13	4.4	2337	1	N91483	NP cDNA to genomic ju 6.75e+02
c	720	13	4.4	2351	4	Q27369	HUG-Br1. 6.75e+02
c	721	13	4.4	2384	12	Q72871	Human tyrosinase gene 6.75e+02
c	722	13	4.4	2400	7	Q42496	Full-length human pse 6.75e+02
c	723	13	4.4	2412	5	Q30940	Encodes BT toxin 63B. 6.75e+02
c	724	13	4.4	2412	29	T60072	63B toxin coding sequ 6.75e+02
c	725	13	4.4	2412	5	Q31413	Bacillus thuringiensi 6.75e+02
c	726	13	4.4	2412	6	Q32164	BT toxin 63B 6.75e+02
c	727	13	4.4	2423	2	Q11727	Sequence encoding vir 6.75e+02
c	728	13	4.4	2433	33	T84441	Staphylococcal MHC cl 6.75e+02
c	729	13	4.4	2445	7	Q48393	Full-length foetal hu 6.75e+02
c	730	13	4.4	2478	1	N91246	Sequence of B. thurin 6.75e+02
	731	13	4.4	2483	24	T41947	ACC oxidase gene prom 6.75e+02
c	732	13	4.4	2483	30	T63319	Yeast mitochondria CO 6.75e+02
c	733	13	4.4	2504	15	Q90443	Murine oviduct specif 6.75e+02
c	734	13	4.4	2517	2	Q10403	Encodes M- and N-prot 6.75e+02
c	735	13	4.4	2520	2	Q11495	Tobacco SURA-C3 mutan 6.75e+02
c	736	13	4.4	2520	23	T33353	Tobacco herbicide-res 6.75e+02
c	737	13	4.4	2520	13	Q81182	ALS SURA-C3 (P197E) m 6.75e+02
c	738	13	4.4	2520	1	N81459	C3 mutant gene encodi 6.75e+02
c	739	13	4.4	2520	4	Q28388	Gene from the ALS C3 6.75e+02
	740	13	4.4	2521	25	T35921	Newcastle disease vir 6.75e+02
	741	13	4.4	2521	28	T39330	Newcastle disease vir 6.75e+02
	742	13	4.4	2521	26	T35875	Newcastle disease vir 6.75e+02
	743	13	4.4	2537	2	N70917	Sequence encoding glu 6.75e+02
c	744	13	4.4	2568	1	Q06237	Sequence encoding cel 6.75e+02
c	745	13	4.4	2581	2	Q13077	extA gene with gene p 6.75e+02
c	746	13	4.4	2615	23	T33187	Bacillus thuringiensi 6.75e+02
	747	13	4.4	2625	28	T61925	Full length human cad 6.75e+02
c	748	13	4.4	2629	1	N81338	Protein A gene with n 6.75e+02
c	749	13	4.4	2631	31	T73696	Alanyl-tRNA synthetas 6.75e+02
c	750	13	4.4	2636	32	T74133	Zea mays root prefere 6.75e+02
c	751	13	4.4	2640	17	T04005	Truncated cystic fibr 6.75e+02
	752	13	4.4	2646	11	Q61607	Mutated GAL4 gene cod 6.75e+02
c	753	13	4.4	2653	25	T44331	B. burgdorferi decori 6.75e+02
	754	13	4.4	2666	17	Q98236	Partial rice pyruvate 6.75e+02
	755	13	4.4	2675	1	Q06957	Genomic sequence enco 6.75e+02
c	756	13	4.4	2675	1	Q06957	Genomic sequence enco 6.75e+02

	757	13	4.4	2692	28	T58899	M-Delta-1 gene. 6.75e+02
c	758	13	4.4	2692	28	T58899	M-Delta-1 gene. 6.75e+02
c	759	13	4.4	2745	3	Q14808	B.thuringiensis toxin 6.75e+02
c	760	13	4.4	2757	30	T43936	Sequence flanking mar 6.75e+02
c	761	13	4.4	2809	15	Q91287	ORC2 gene encoding a 6.75e+02
c	762	13	4.4	2809	28	T51531	Mycobacterium gallise 6.75e+02
c	763	13	4.4	2861	19	T13334	Epidermal-specific B1 6.75e+02
c	764	13	4.4	2871	3	Q14807	B.thuringiensis toxin 6.75e+02
c	765	13	4.4	2872	18	T14849	Polycistronic maize U 6.75e+02
c	766	13	4.4	2872	18	T08745	Polycistronic maize U 6.75e+02
c	767	13	4.4	2881	11	Q68668	PMGA 1.2 and partial 6.75e+02
	768	13	4.4	2901	23	T32961	Mouse developmental k 6.75e+02
c	769	13	4.4	2951	17	T01406	I3L-promoted CHV gB g 6.75e+02
	770	13	4.4	2960	29	T41616	Adzuki bean endo-xylo 6.75e+02
	771	13	4.4	2967	31	T52460	STE2-GPA1 fusion gene 6.75e+02
c	772	13	4.4	2975	39	V15529	Human lymphoid-specif 6.75e+02
c	773	13	4.4	2980	1	N91071	Nucleotide sequence i 6.75e+02
c	774	13	4.4	2983	12	Q71026	CryIIIA gene which en 6.75e+02
	775	13	4.4	2992	31	T71087	Yeast alanyl-tRNA syn 6.75e+02
c	776	13	4.4	3006	9	Q52491	Helminth aminopeptida 6.75e+02
	777	13	4.4	3012	3	Q21188	Sequence encoding a 3 6.75e+02
	778	13	4.4	3012	11	Q58997	cDNA coding human cal 6.75e+02
c	779	13	4.4	3063	1	N91438	DNA encoding glycine 6.75e+02
c	780	13	4.4	3063	12	Q67874	Canarypox insert in p 6.75e+02
c	781	13	4.4	3100	40	V20419	Human discs large 1 g 6.75e+02
c	782	13	4.4	3111	15	Q87260	Flea sodium pump alph 6.75e+02
	783	13	4.4	3156	14	Q82976	Genomic clone encodin 6.75e+02
c	784	13	4.4	3180	21	T12676	F. brownii cold resis 6.75e+02
c	785	13	4.4	3196	40	V20420	Human discs large 1 g 6.75e+02
c	786	13	4.4	3219	6	Q35948	rfa-2 gene. 6.75e+02
c	787	13	4.4	3240	39	V15530	Human lymphoid-specif 6.75e+02
c	788	13	4.4	3273	22	T18868	Human survival motor 6.75e+02
	789	13	4.4	3279	18	T13919	Algal alpha-1,4-gluca 6.75e+02
c	790	13	4.4	3303	3	Q14806	B.thuringiensis toxin 6.75e+02
	791	13	4.4	3315	19	T18535	Receptor tyrosine kin 6.75e+02
c	792	13	4.4	3320	15	Q91200	H-NUC retinoblastoma 6.75e+02
c	793	13	4.4	3335	1	N91154	Turkey Herpes Virus B 6.75e+02
	794	13	4.4	3336	25	T35912	Herpesvirus of turkey 6.75e+02
	795	13	4.4	3350	5	Q29361	Ocean pout DNA sequen 6.75e+02
	796	13	4.4	3390	11	Q62283	cDNA coding human pla 6.75e+02
	797	13	4.4	3399	16	Q91095	Human herpesvirus-6B 6.75e+02
c	798	13	4.4	3414	7	Q41744	DNA encoding B.t. tox 6.75e+02
c	799	13	4.4	3414	9	Q56921	Bacillus thuringiensi 6.75e+02
	800	13	4.4	3416	11	Q62282	cDNA coding human ova 6.75e+02
	801	13	4.4	3448	39	V06055	Singapore horseshoe c 6.75e+02
c	802	13	4.4	3467	2	Q06550	Fibronectin-binding p 6.75e+02
c	803	13	4.4	3471	4	Q27167	Delta-endotoxin gene. 6.75e+02
c	804	13	4.4	3471	6	Q38653	Bt isolate Ps50C codi 6.75e+02
c	805	13	4.4	3471	5	Q28940	B.thuringiensis PS50C 6.75e+02
c	806	13	4.4	3507	24	T43221	Antiscarab pest toxin 6.75e+02
	807	13	4.4	3521	7	Q41226	Clone GP3 encoding ma 6.75e+02

	808	13	4.4	3521	7	Q41230	Gene encoding major s 6.75e+02
c	809	13	4.4	3521	33	T47554	H6 promoted cytomegal 6.75e+02
	810	13	4.4	3534	29	T67410	H. pylori cytoplasmic 6.75e+02
c	811	13	4.4	3537	30	T69981	FIPV spike (S) gene C 6.75e+02
	812	13	4.4	3549	30	T68137	H. pylori cytoplasmic 6.75e+02
	813	13	4.4	3551	35	T91041	Yeast checkpoint cont 6.75e+02
c	814	13	4.4	3600	19	T12326	DNA encoding Sulfolob 6.75e+02
c	815	13	4.4	3607	37	T45420	3' UTR of the ataxia- 6.75e+02
c	816	13	4.4	3635	13	Q80004	Sucrose-phosphate-syn 6.75e+02
	817	13	4.4	3637	35	T91056	BPV-1 replication ori 6.75e+02
c	818	13	4.4	3661	20	T05902	Human IL-2 gene (p2IL 6.75e+02
c	819	13	4.4	3704	29	T59938	Phage resistance gene 6.75e+02
c	820	13	4.4	3706	30	T69995	Canarypox virus DNA c 6.75e+02
	821	13	4.4	3712	10	Q44392	Sequence of human OSF 6.75e+02
	822	13	4.4	3763	1	Q06328	Sequence encoding mer 6.75e+02
	823	13	4.4	3763	33	T80072	Merozoite apical-end 6.75e+02
c	824	13	4.4	3825	29	T33961	Beta-amylase coding s 6.75e+02
c	825	13	4.4	3825	29	T33962	Mutated beta-amylase 6.75e+02
c	826	13	4.4	3881	19	T13335	Blec4 gene and promot 6.75e+02
	827	13	4.4	3914	10	Q44393	Sequence of human OSF 6.75e+02
c	828	13	4.4	3932	11	Q64112	cryET5 gene. 6.75e+02
c	829	13	4.4	3934	35	T95051	DNA encoding Bacillus 6.75e+02
c	830	13	4.4	3957	18	T14077	Human dihydropyrimidi 6.75e+02
	831	13	4.4	4002	12	Q74463	Endogenous plasmid pA 6.75e+02
c	832	13	4.4	4032	6	Q40416	DNA clone of lambdaUC 6.75e+02
c	833	13	4.4	4032	4	Q28506	Tomato HCD promoter (6.75e+02
	834	13	4.4	4098	24	T43682	Medium chain-specific 6.75e+02
	835	13	4.4	4133	15	Q87268	B. sphaericus mtx gen 6.75e+02
	836	13	4.4	4133	8	Q48715	Insecticidal protein 6.75e+02
	837	13	4.4	4133	3	Q14257	Mosquitotoxin gene. 6.75e+02
c	838	13	4.4	4133	8	Q48715	Insecticidal protein 6.75e+02
c	839	13	4.4	4133	15	Q87268	B. sphaericus mtx gen 6.75e+02
	840	13	4.4	4182	39	V06054	Singapore horseshoe c 6.75e+02
	841	13	4.4	4226	28	T39485	Human steroidogenesis 6.75e+02
c	842	13	4.4	4253	28	T58361	Tobacco epi-5-aristol 6.75e+02
c	843	13	4.4	4291	18	T04702	ALVAC C3 flanking arm 6.75e+02
c	844	13	4.4	4304	23	T32960	Mouse developmental k 6.75e+02
	845	13	4.4	4328	5	Q28758	Partial sequence of t 6.75e+02
c	846	13	4.4	4337	14	Q82833	Human stromalin-1 DNA 6.75e+02
c	847	13	4.4	4368	18	T03133	Human dihydropyrimidi 6.75e+02
c	848	13	4.4	4369	18	T03143	Human dihydropyrimidi 6.75e+02
c	849	13	4.4	4464	2	N71244	Genomic sequence of a 6.75e+02
c	850	13	4.4	4465	5	Q28697	Glutamine synthetase. 6.75e+02
c	851	13	4.4	4495	24	T32143	Helicobacter-specific 6.75e+02
	852	13	4.4	4526	8	Q49780	Prosystemin. 6.75e+02
	853	13	4.4	4590	2	N60472	Sequence encoding the 6.75e+02
	854	13	4.4	4604	18	T00523	pC5LCDVHAF1 coding se 6.75e+02
	855	13	4.4	4698	1	N90844	Tobacco basic beta-1, 6.75e+02
	856	13	4.4	4699	1	Q03254	Sequence of clone BSG 6.75e+02
	857	13	4.4	4706	17	T10157	Vector plasmid pOnMTG 6.75e+02
c	858	13	4.4	4711	24	T41419	HCMV Towne strain nov 6.75e+02

	859	13	4.4	4744	17	T10155	Sockeye salmon growth 6.75e+02
c	860	13	4.4	4785	8	Q48796	B.ohbensis Ser(188) c 6.75e+02
c	861	13	4.4	4785	3	Q14072	Cyclodextrin glucanot 6.75e+02
c	862	13	4.4	4785	8	Q48788	B.ohbensis mutant cyc 6.75e+02
c	863	13	4.4	4785	8	Q48797	B.ohbensis Trp(188) c 6.75e+02
	864	13	4.4	4821	21	T14051	Helicobacter pylori T 6.75e+02
	865	13	4.4	4821	14	Q86728	H. pylori tagA antige 6.75e+02
	866	13	4.4	4857	30	T69962	FIPV nucleocapsid (N) 6.75e+02
	867	13	4.4	4864	36	V01455	A. thaliana inorganic 6.75e+02
	868	13	4.4	4864	37	T91325	Arabidopsis thaliana 6.75e+02
c	869	13	4.4	4894	2	Q13605	Cystic fibrosis trans 6.75e+02
c	870	13	4.4	4909	33	T47558	42K promoted cytomega 6.75e+02
	871	13	4.4	4940	31	T73279	S. cerevisiae origin 6.75e+02
	872	13	4.4	4940	15	Q91286	ORC1 gene encoding a 6.75e+02
c	873	13	4.4	5056	18	T17738	Human C3 cDNA. 6.75e+02
c	874	13	4.4	5130	1	Q05123	Sequence encoding fel 6.75e+02
c	875	13	4.4	5255	3	Q25448	Marek Disease Virus g 6.75e+02
c	876	13	4.4	5350	37	T89254	cDNA for rat CASPR/pl 6.75e+02
	877	13	4.4	5426	25	T18210	Turkey Herpes Virus p 6.75e+02
	878	13	4.4	5495	35	T75590	Canine herpesvirus ge 6.75e+02
	879	13	4.4	5501	11	Q64896	HER4. 6.75e+02
c	880	13	4.4	5534	6	Q35988	Tomato hsp80 genomic 6.75e+02
c	881	13	4.4	5737	16	Q74053	Human interleukin-2. 6.75e+02
c	882	13	4.4	5760	3	N50530	Sequence encoding the 6.75e+02
c	883	13	4.4	5775	40	V20437	Human c-K-ras oncogen 6.75e+02
c	884	13	4.4	5852	2	Q11710	Dictyostelium plasmid 6.75e+02
	885	13	4.4	5873	10	Q57913	Arabidopsis thaliana 6.75e+02
	886	13	4.4	5890	10	Q57912	Arabidopsis thaliana 6.75e+02
	887	13	4.4	5890	10	Q57915	Arabidopsis thaliana 6.75e+02
	888	13	4.4	5890	10	Q57914	Arabidopsis thaliana 6.75e+02
	889	13	4.4	5892	35	T79680	BRCA2 cancer suscepti 6.75e+02
	890	13	4.4	5925	8	Q48733	CAI gene. 6.75e+02
c	891	13	4.4	5931	7	Q43701	Sequence of the human 6.75e+02
c	892	13	4.4	6045	32	T74665	Utrophin minigene. 6.75e+02
c	893	13	4.4	6057	10	Q54012	Alpha-agglutinin codi 6.75e+02
	894	13	4.4	6124	1	Q03568	Sequence encoding the 6.75e+02
c	895	13	4.4	6127	2	Q11371	Mutant cystic fibrosi 6.75e+02
c	896	13	4.4	6128	2	Q13072	CFTR 3659 del C. 6.75e+02
c	897	13	4.4	6128	2	Q13068	CFTR 556 del A. 6.75e+02
c	898	13	4.4	6129	14	Q84724	DNA encoding an antib 6.75e+02
c	899	13	4.4	6129	2	Q13059	CFTR G542X. 6.75e+02
c	900	13	4.4	6129	2	Q13071	CFTR 1717 -1G -> A. 6.75e+02
c	901	13	4.4	6129	2	Q13061	CFTR G551D. 6.75e+02
c	902	13	4.4	6129	2	Q13056	CFTR G178R. 6.75e+02
c	903	13	4.4	6129	2	Q13065	CFTR L1077P. 6.75e+02
c	904	13	4.4	6129	2	Q13057	CFTR A455E. 6.75e+02
c	905	13	4.4	6129	2	Q13058	CFTR Q493X. 6.75e+02
c	906	13	4.4	6129	2	Q13062	CFTR R560T. 6.75e+02
c	907	13	4.4	6129	14	Q87795	Cystic fibrosis trans 6.75e+02
c	908	13	4.4	6129	15	Q94067	DNA encoding the CFTC 6.75e+02
c	909	13	4.4	6129	2	Q13060	CFTR S549R. 6.75e+02

c	910	13	4.4	6129	2	Q13063	CFTR Y563N. 6.75e+02
c	911	13	4.4	6130	4	Q24280	Cystic fibrosis gene. 6.75e+02
	912	13	4.4	6136	14	Q90478	Chinook salmon growth 6.75e+02
c	913	13	4.4	6144	30	T69976	FIPV spike (S) gene d 6.75e+02
c	914	13	4.4	6146	4	Q23660	Cystic Fibrosis trans 6.75e+02
	915	13	4.4	6266	13	Q92411	Rice mitochondrial DN 6.75e+02
	916	13	4.4	6295	30	T67128	Constitutive triple r 6.75e+02
	917	13	4.4	6295	15	Q98818	Arabidopsis CTR1-2 mu 6.75e+02
	918	13	4.4	6312	30	T67127	Constitutive triple r 6.75e+02
	919	13	4.4	6312	15	Q98819	Arabidopsis CTR1-3 mu 6.75e+02
c	920	13	4.4	6384	13	Q71546	Component of transcri 6.75e+02
	921	13	4.4	6404	7	Q42388	6.5 kb BanHI fragmant 6.75e+02
c	922	13	4.4	6452	22	T30193	Peripheral nervous sy 6.75e+02
c	923	13	4.4	6749	32	T47564	Cytomegalovirus pp65 6.75e+02
c	924	13	4.4	6749	33	T47561	Cytomegalovirus pp65 6.75e+02
	925	13	4.4	6877	10	Q53548	Cinnamyl alcohol dehy 6.75e+02
	926	13	4.4	6922	14	Q83735	Calcium channel alpha 6.75e+02
c	927	13	4.4	6922	14	Q83735	Calcium channel alpha 6.75e+02
	928	13	4.4	7032	15	Q84662	Human neuronal calciu 6.75e+02
	929	13	4.4	7089	15	Q84663	Human neuronal calciu 6.75e+02
c	930	13	4.4	7089	15	Q84663	Human neuronal calciu 6.75e+02
	931	13	4.4	7091	33	T47530	H6 promoted cytomegal 6.75e+02
c	932	13	4.4	7156	39	V15586	Lactobacillus bulgari 6.75e+02
	933	13	4.4	7241	3	Q15140	Genomic clone LE-ACC2 6.75e+02
	934	13	4.4	7244	40	V15705	Tomato ACC synthase L 6.75e+02
	935	13	4.4	7350	30	T69982	Canarypox virus DNA c 6.75e+02
c	936	13	4.4	7350	30	T69982	Canarypox virus DNA c 6.75e+02
	937	13	4.4	7351	3	Q22770	Canarypox virus BglII 6.75e+02
c	938	13	4.4	7351	3	Q22770	Canarypox virus BglII 6.75e+02
	939	13	4.4	7452	13	Q85594	Human glycoprotein V 6.75e+02
c	940	13	4.4	7475	3	Q15312	Plasmid pHPB12 encodi 6.75e+02
	941	13	4.4	7920	20	T05899	Human IL-14 gene. 6.75e+02
c	942	13	4.4	8351	35	T91039	Yeast checkpoint cont 6.75e+02
	943	13	4.4	8371	30	T66909	Human neuropeptide Y- 6.75e+02
	944	13	4.4	8409	9	Q62611	Humanised anti-CD18 A 6.75e+02
c	945	13	4.4	8654	14	Q55138	Staphylococcus epider 6.75e+02
c	946	13	4.4	8710	17	Q89650	pSC11 FIPV E1. 6.75e+02
c	947	13	4.4	8797	27	T50856	DNA cassette for amph 6.75e+02
c	948	13	4.4	8878	22	T03852	Human fibrinogen B-be 6.75e+02
c	949	13	4.4	9071	3	Q23370	U266-Lambda gene and 6.75e+02
c	950	13	4.4	9071	3	Q22491	Human U266 lambda gen 6.75e+02
	951	13	4.4	9229	1	N80437	Entire sequence of LA 6.75e+02
	952	13	4.4	9236	1	N80436	Entire sequence of LA 6.75e+02
c	953	13	4.4	9606	13	Q70633	Adenomatous polyposis 6.75e+02
c	954	13	4.4	9714	39	T73135	Porcine transmissible 6.75e+02
	955	13	4.4	10095	13	Q81762	Japanese Black Pine c 6.75e+02
	956	13	4.4	10190	28	T33183	Fragment of human OCI 6.75e+02
c	957	13	4.4	10320	32	T74666	Full-length utrophin 6.75e+02
c	958	13	4.4	10437	12	Q72297	Adenomatous polyposis 6.75e+02
	959	13	4.4	10475	1	N80616	Genomic clone encodin 6.75e+02
	960	13	4.4	10475	2	N80643	Sequence of human pul 6.75e+02

```
c  961     13    4.4  10475  2  N80643      Sequence of human pul  6.75e+02
c  962     13    4.4  10475  1  N80616      Genomic clone encodin  6.75e+02
c  963     13    4.4  10607  6  Q37894      Beta-casein coding se  6.75e+02
c  964     13    4.4  10684 24  T33758      Control region isolat  6.75e+02
c  965     13    4.4  10704  7  Q43846      Plasmid pAH4625.       6.75e+02
c  966     13    4.4  10844  7  Q43848      Plasmid pAH4808.       6.75e+02
   967     13    4.4  11283 35  T84841      Human breast and ovar  6.75e+02
c  968     13    4.4  11336  7  Q43497      Approximate nucleotid  6.75e+02
   969     13    4.4  11385 36  T69707      Human breast cancer s  6.75e+02
c  970     13    4.4  11529  7  Q43844      Plasmid pAH4602.       6.75e+02
   971     13    4.4  12284 20  T29591      Hog cholera virus.     6.75e+02
   972     13    4.4  12284 12  Q71308      Sequence encoding 55   6.75e+02
c  973     13    4.4  12492  1  N70543      Complete sequence of   6.75e+02
c  974     13    4.4  12778 31  T68784      ATM mutant G9170C cod  6.75e+02
   975     13    4.4  13011 35  T96631      cDNA encoding rat cyt  6.75e+02
c  976     13    4.4  13807 32  T62328      Exopolysaccharide ope  6.75e+02
   977     13    4.4  19012  6  Q36952      HSA gene.              6.75e+02
   978     13    4.4  19124 32  T72882      Plasmodium var-7 gene  6.75e+02
   979     13    4.4  20303 31  T71699      Human deoxycytidylate  6.75e+02
   980     13    4.4  21126  2  N60970      TL-DNA region of A.rh  6.75e+02
c  981     13    4.4  21126  2  N60970      TL-DNA region of A.rh  6.75e+02
   982     13    4.4  24025 21  T17455      Mutated BRCA1 genomic  6.75e+02
   983     13    4.4  24026 21  T17512      Mutated BRCA1 genomic  6.75e+02
   984     13    4.4  24026 21  T17513      Mutated BRCA1 genomic  6.75e+02
   985     13    4.4  24026 21  T17527      Mutated BRCA1 genomic  6.75e+02
   986     13    4.4  24026 21  T17528      Mutated BRCA1 genomic  6.75e+02
   987     13    4.4  24026 21  T17521      Mutated BRCA1 genomic  6.75e+02
   988     13    4.4  24026 23  T32612      BRCA1, human breast a  6.75e+02
   989     13    4.4  24026 21  T17517      Mutated BRCA1 genomic  6.75e+02
   990     13    4.4  24026 21  T17519      Mutated BRCA1 genomic  6.75e+02
   991     13    4.4  24026 21  T17529      Mutated BRCA1 genomic  6.75e+02
   992     13    4.4  24026 21  T17526      Mutated BRCA1 genomic  6.75e+02
   993     13    4.4  24026 21  T17524      Mutated BRCA1 genomic  6.75e+02
   994     13    4.4  24029 21  T17520      Mutated BRCA1 genomic  6.75e+02
   995     13    4.4  24031 21  T17525      Mutated BRCA1 genomic  6.75e+02
   996     13    4.4  26764 31  T71696      Human deoxycytidylate  6.75e+02
c  997     13    4.4  28994 39  V15826.     Genomic DNA for inter  6.75e+02
c  998     13    4.4  36335 16  Q68003      Ad2/-ORF6/PGK-CFTR nu  6.75e+02
   999     13    4.4  40352 40  V02032      MAGE-B cluster DNA se  6.75e+02
c 1000     13    4.4 134525  1  Q04525      Total base sequence o  6.75e+02
```

ALIGNMENTS

```
RESULT    1
ID   T19563 standard; cDNA to mRNA; 296 BP.
AC   T19563;
DT   27-JUN-1996  (first entry)
DE   Human gene signature HUMGS00628.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
```



```
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 418; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   296 BP;    102 A;    42 C;    38 G;    114 T;

  Query Match          100.0%;  Score 296;  DB 18;  Length 296;
  Best Local Similarity 100.0%;  Pred. No. 4.08e-241;
  Matches   296;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db       1 gatctgccataagaaaatctagttcaactctaatttttatgtagtaaataaattggcaggt 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy       1 GATCTGCCATAAGAAAATCTAGTTCAACTCTAATTTTATGTAGTAAATAAATTGGCAGGT 60

Db      61 aattgtttttacaaagaatccaccctgacttcccctaatgcattaaaaatatttttattta 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      61 AATTGTTTTTACAAAGAATCCACCTGACTTCCCCTAATGCATTAAAAATATTTTTATTTA 120

Db     121 aataactttatttataacttttagaaacatgtagtattgtttaaacatcatttgttcttc 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     121 AATAACTTTATTTATAACTTTTAGAAACATGTAGTATTGTTTAAACATCATTTGTTCTTC 180

Db     181 agtatttttcatttggaagtccaatagggcaaattgaatgaagtattattatctgtctct 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     181 AGTATTTTTCATTTGGAAGTCCAATAGGGCAAATTGAATGAAGTATTATTATCTGTCTCT 240
```



```
Db     241 tgtagtacaatgtatccaacagacactcaataaacttttttggttgttaaaactgaaa 296
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     241 TGTAGTACAATGTATCCAACAGACACTCAATAAACTTTTTGGTTGTTAAACTGAAA 296


RESULT    2
ID   Q78184 standard; DNA; 399 BP.
AC   Q78184;
DT   02-JUN-1995  (first entry)
DE   Maize associated region MAR3 (SAR(L))
KW   Genetic stabilising element; matrix associated region; MAR; ss.
OS   Zea mays
FH   Key             Location/Qualifiers
FT   misc_feature    43..69
FT                   /*tag=  a
FT                   /label= ARS(11/12) A BOX(9/10) ARS(10/12)
FT   misc_feature    99..109
FT                   /*tag=  b
FT                   /label= T BOX (8/10)
FT   misc_feature    116..138
FT                   /*tag=  c
FT                   /label= AT BOX T BOX(9/10) ARS(10/12) T BOX(9/10)
FT   misc_feature    216..245
FT                   /*tag=  d
FT                   /label= ARS(10/12) ARS(10/12) ARS(10/12)
FT   misc_feature    259..271
FT                   /*tag=  e
FT                   /label= T BOX(9/10)
FT   misc_feature    308..331
FT                   /*tag=  f
FT                   /label= AT BOX ARS(10/12) T BOX(9/10)
FT   misc_feature    334..348
FT                   /*tag=  g
FT                   /label= TOPO II (15/15)
FT   misc_feature    379..389
FT                   /*tag=  h
FT                   /label= T BOX (9/10)
PN   WO9424293-A.
PD   27-OCT-1994.
PF   18-APR-1994; E01193.
PR   19-APR-1993; US-049564.
PA   (SANO ) SANDOZ LTD.
PA   (SANO ) SANDOZ PATENT GMBH.
PA   (SANO ) SANDOZ-ERFINDUNGEN VERW GES MBH.
PI   De HAAS JM,  Dietrich PS,  Gielen JJL,  Van DRIEL R;
PI   Van GRINSVEN MQJM;
DR   WPI; 94-341872/42.
PT   Stabilised exogenous genes for plant cell transformation - with
PT   at least one stabilising DNA segment in 3'- or 5'-flanking region
PS   Example; Page 36; 53pp; English.
```



```
CC   The term MAR refers to DNA segments isolated from nuclear scaffold
CC   or nuclear matrix prepns. after endonuclease treatment. Two
CC   fragments, MAR1 (maize 0.8 kb AT rich region) and MAR2 (maize 1.25
CC   kb regions with ARS3), found withing a 5 kb maize EcoR1 fragment
CC   have nuclear matrix binding activity. The sequence of the MAR1
CC   fragment is given in Q78182 and that portion of MAR2 not previously
CC   sequenced is given in Q78183 along with the published portion of
CC   MAR2 named ARS3 (Berlani et al. 1988 Plant Molecular Biology 11:
CC   173-182). Additionally the sequence from the SAR(L) a region from a
CC   small heat shock gene (MAR3) (soybean HSP17.6 0.4 kb SARL) is shown
CC   in Q78184 (Schoffl et al. Transgenic Res. 2 93-100 (1993)). MAR1,
CC   MAR2 and MAR3 were cloned to construct plasmids pZ01927, pZ01929
CC   and pSVB20-SAR(L) for gene activity.
SQ   Sequence   399 BP;    166 A;    34 C;    39 G;    160 T;

  Query Match            5.7%;  Score 17;  DB 13;  Length 399;
  Best Local Similarity 100.0%;  Pred. No. 3.99e+00;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     113 taaaaatatttttattt 129
           |||||||||||||||||
Qy     103 TAAAAATATTTTTATTT 119


RESULT    3
ID   Q98473 standard; cDNA; 870 BP.
AC   Q98473;
DT   01-MAY-1996  (first entry)
DE   Carboxy terminus of MiSP2-containing plasmid pMISS2.
KW   Spider silk; repeat unit; consensus; minor ampullate silk protein;
KW   spidroin; MiSP; orb web spider; dragline; ds.
OS   Nephila clavipes.
FH   Key             Location/Qualifiers
FT   cds             1..756
FT                   /*tag=  a
PN   WO9525165-A1.
PD   21-SEP-1995.
PF   14-MAR-1995; U03139.
PR   14-MAR-1994; US-209747.
PA   (UYWY-) UNIV WYOMING.
PI   Colgin M,  Lewis RV;
DR   WPI; 95-336970/43.
DR   P-PSDB; R80192.
PT   Polypeptide(s) comprising repeated unit amino acid sequences, also
PT   cDNAs - derived from minor ampullate spider silk proteins and used
PT   to form spider silk fibres
PS   Disclosure; Fig 2C-D; 86pp; English.
CC   Several regions of the cDNA clone, pMISS2, encoding the orb web spider
CC   minor ampullate silk protein MiSP2 have been identified and sequenced.
CC   This sequence encodes the carboxy terminal of MiSP2. Repeat unit peptides
CC   make up spider silk proteins (spidroins) which in turn aggregate to
```

```
CC   form the silk fibres. Spider silk fibres have high tensile strength
CC   and significant elasticity. An isolated cDNA clone of a silk protein
CC   encoding sequence is of use to produce the protein at high yields
CC   using recombinant DNA technology.
SQ   Sequence   870 BP;    209 A;    160 C;    235 G;    266 T;

  Query Match            5.7%;  Score 17;  DB 17;  Length 870;
  Best Local Similarity 100.0%;  Pred. No. 3.99e+00;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    835 taaataaaaatattttt 851
          !!!!!!!!!!!!!!!!!
Cp    120 TAAATAAAAATATTTTT 104


RESULT    4
ID   T87922 standard; cDNA; 3441 BP.
AC   T87922;
DT   18-DEC-1997  (first entry)
DE   Rat cerebellum derived growth factor 1 cDNA.
KW   Rat; cerebellum derived growth factor; CDGF1; screening; binding;
KW   modulation; erbB type receptor; identification; indication; risk;
KW   proliferation; differentiation; induction; neuron; hyperplasia;
KW   stem cell culture; intracerebral graft; alleviation; repair;
KW   behavioural defect; nervous system; central; peripheral; nerve;
KW   prothesis; damage; entubulation; cell survival; treatment;
KW   injury; trauma; ischaemia; ischemia; stroke; infection; disorder;
KW   inflammation; neurodegeneration; disease; Parkinson's;
KW   Huntingdon's; amylotrophic lateral sclerosis; sensory; retina;
KW   spinocerebellar degeneration; multiple sclerosis; neoplasia;
KW   amalignant glioma; medulloblastoma; neuroectodermal tumour; ds.
OS   Rattus rattus.
FH   Key             Location/Qualifiers
FT   cds             180..2444
FT                   /*tag=  a
FT   signal_peptide  180..248
FT                   /*tag=  b
FT   mat_peptide     249..2441
FT                   /*tag=  c
FT                   /product= cerebellum_derived_growth_factor
PN   WO9709425-A1.
PD   13-MAR-1997.
PF   09-SEP-1996; U14484.
PR   08-SEP-1995; US-525864.
PA   (HARD ) HARVARD COLLEGE.
PA   (STRD ) UNIV LELAND STANFORD JUNIOR.
PA   (STRD ) UNIV LELAND S STANFORD.
PI   Chang H;
DR   WPI; 97-192900/17.
DR   P-PSDB; W27536.
PT   Rat and human cerebellum-derived growth factors - used in the
```



```
PT   treatment of neuronal injury and proliferative disorders
PS   Claim 17; Pages 63-66; 94pp; English.
CC   The present sequence encodes rat cerebellum derived growth factor 1
CC   (CDGF1), which can be used to screen for modulators of CDGF
CC   binding to erbB type receptors. Identification of a modification or
CC   mutation in a CDGF gene, or aberrant expression of a CDGF gene or
CC   levels of soluble CDGF may be used to indicate the risk of unwanted
CC   cell proliferation or differentiation.
CC   CDGF may be used to induce neuronal differentiation in stem cell
CC   culture, and maintain the integrity of a terminally differentiated
CC   neuronal cell culture, e.g. useful for intracerebral grafting to
CC   alleviate behavioural defects. CDGF may also be used in nerve
CC   protheses to repair central and peripheral nerve damage, especially
CC   where a crushed or severed axon is entubulated by a prosthetic.
CC   CDGF may also be used to enhance neuronal cell survival in the
CC   central or peripheral nervous system, to treat neurological
CC   conditions associated with nervous system injury, e.g. traumatic,
CC   chemical or vasal injury and deficits such as ischaemia resulting
CC   from stroke, infectious/inflammatory and tumour induced injury,
CC   chronic neurodegenerative disease including Parkinson's and
CC   Huntingdon's, amylotrophic lateral sclerosis, spinocerebellar
CC   degeneration, chronic immunological disease of the nervous system
CC   including multiple sclerosis, disorders of the sensory neurons and
CC   degenerative diseases of the retina. CDGF may also be used to treat
CC   neoplastic or hyperplastic transformations, particularly of the
CC   central nervous system, e.g. amalignant gliomas, medulloblastomas
CC   and neuroectodermal tumours.
SQ   Sequence   3441 BP;    777 A;    1057 C;    1015 G;    592 T;

  Query Match            5.7%;  Score 17;  DB 34;  Length 3441;
  Best Local Similarity 100.0%;  Pred. No. 3.99e+00;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2527 taaaaatatttttattt 2543
          !!!!!!!!!!!!!!!!!
Qy    103 TAAAAATATTTTTATTT 119


RESULT    5
ID   T87922 standard; cDNA; 3441 BP.
AC   T87922;
DT   18-DEC-1997  (first entry)
DE   Rat cerebellum derived growth factor 1 cDNA.
KW   Rat; cerebellum derived growth factor; CDGF1; screening; binding;
KW   modulation; erbB type receptor; identification; indication; risk;
KW   proliferation; differentiation; induction; neuron; hyperplasia;
KW   stem cell culture; intracerebral graft; alleviation; repair;
KW   behavioural defect; nervous system; central; peripheral; nerve;
KW   prothesis; damage; entubulation; cell survival; treatment;
KW   injury; trauma; ischaemia; ischemia; stroke; infection; disorder;
KW   inflammation; neurodegeneration; disease; Parkinson's;
```



```
KW   Huntingdon's; amylotrophic lateral sclerosis; sensory; retina;
KW   spinocerebellar degeneration; multiple sclerosis; neoplasia;
KW   amalignant glioma; medulloblastoma; neuroectodermal tumour; ds.
OS   Rattus rattus.
FH   Key             Location/Qualifiers
FT   cds             180..2444
FT                   /*tag=  a
FT   signal_peptide  180..248
FT                   /*tag=  b
FT   mat_peptide     249..2441
FT                   /*tag=  c
FT                   /product= cerebellum_derived_growth_factor
PN   WO9709425-A1.
PD   13-MAR-1997.
PF   09-SEP-1996; U14484.
PR   08-SEP-1995; US-525864.
PA   (HARD ) HARVARD COLLEGE.
PA   (STRD ) UNIV LELAND STANFORD JUNIOR.
PA   (STRD ) UNIV LELAND S STANFORD.
PI   Chang H;
DR   WPI; 97-192900/17.
DR   P-PSDB; W27536.
PT   Rat and human cerebellum-derived growth factors - used in the
PT   treatment of neuronal injury and proliferative disorders
PS   Claim 17; Pages 63-66; 94pp; English.
CC   The present sequence encodes rat cerebellum derived growth factor 1
CC   (CDGF1), which can be used to screen for modulators of CDGF
CC   binding to erbB type receptors. Identification of a modification or
CC   mutation in a CDGF gene, or aberrant expression of a CDGF gene or
CC   levels of soluble CDGF may be used to indicate the risk of unwanted
CC   cell proliferation or differentiation.
CC   CDGF may be used to induce neuronal differentiation in stem cell
CC   culture, and maintain the integrity of a terminally differentiated
CC   neuronal cell culture, e.g. useful for intracerebral grafting to
CC   alleviate behavioural defects. CDGF may also be used in nerve
CC   protheses to repair central and peripheral nerve damage, especially
CC   where a crushed or severed axon is entubulated by a prosthetic.
CC   CDGF may also be used to enhance neuronal cell survival in the
CC   central or peripheral nervous system, to treat neurological
CC   conditions associated with nervous system injury, e.g. traumatic,
CC   chemical or vasal injury and deficits such as ischaemia resulting
CC   from stroke, infectious/inflammatory and tumour induced injury,
CC   chronic neurodegenerative disease including Parkinson's and
CC   Huntingdon's, amylotrophic lateral sclerosis, spinocerebellar
CC   degeneration, chronic immunological disease of the nervous system
CC   including multiple sclerosis, disorders of the sensory neurons and
CC   degenerative diseases of the retina. CDGF may also be used to treat
CC   neoplastic or hyperplastic transformations, particularly of the
CC   central nervous system, e.g. amalignant gliomas, medulloblastomas
CC   and neuroectodermal tumours.
SQ   Sequence   3441 BP;    777 A;    1057 C;    1015 G;    592 T;
```



```
  Query Match            5.7%;  Score 17;  DB 34;  Length 3441;
  Best Local Similarity 100.0%;  Pred. No. 3.99e+00;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2524 aaataaaaatatttttta 2540
          !!!!!!!!!!!!!!!!!
Cp    119 AAATAAAAATATTTTTA 103


RESULT    6
ID   T58840 standard; DNA; 580073 BP.
AC   T58840;
DT   27-MAR-1997  (first entry)
DE   Mycoplasma genitalium genome.
KW   M. genitalium; DNAA; DNA gyrase; origin of replication;
KW   megabase shotgun sequencing method; open reading frame; ORF; ss.
OS   Mycoplasma genitalium.
FH   Key             Location/Qualifiers
FT   cds             8552..9184
FT                   /*tag=  a
FT                   /label= MG006
FT                   /note= "Previously identified as MORF-20076, the
FT                   encoded protein shows 27.59 percentage
FT                   identity to thymidylate kinase (CDC8)
FT                   from Saccharomyces cerevisiae"
FT   cds             11252..12040
FT                   /*tag=  b
FT                   /label= MG009
FT                   /note= "Previously identified as MORF-20078, the
FT                   encoded protein shows 35.43 percentage
FT                   identity to the Bacillus subtilis hypothetical
FT                   protein covered in accession number
FT                   GB:D26185_102"
FT   cds             12069..12725
FT                   /*tag=  c
FT                   /label= MG010
FT                   /note= "Previously identified as MORF-20079, the
FT                   encoded protein shows 25.73 percentage
FT                   identity to DNA primase (dnaE) from
FT                   Clostridium acetobutylicum"
FT   cds             complement (13570..14247)
FT                   /*tag=  d
FT                   /label= MG012
FT                   /note= "Previously identified as MORF-20080, the
FT                   encoded protein shows 31.50 percentage
FT                   identity to the ribosomal protein S6
FT                   modification protein (rimK) from Escherichia
FT                   coli"
FT   cds             complement (14396..15217)
FT                   /*tag=  e
```

```
FT                   /label= MG013
FT                   /note= "Previously identified as MORF-19823, MORF-20080
FT                   and MORF-20081, the encoded protein shows 33.04
FT                   percentage identity to 5,10-methylene-tetra-
FT                   hydrofolate dehydrogenase (folD) from E. coli"
FT   cds             17474..19243
FT                   /*tag= f
FT                   /label= MG015
FT                   /note= "Previously identified as MORF-20084, the
FT                   encoded protein shows 32.23 percentage
FT                   identity to transport ATP-binding protein
FT                   (msbA) from E. coli"
FT   cds             26478..27344
FT                   /*tag= g
FT                   /label= MG023
FT                   /note= "Previously identified as MORF-20092, the
FT                   encoded protein shows 45.96 percentage
FT                   identity to fructose-bisphosphate aldolase
FT                   (tsr) from B. subtilis"
FT   cds             27345..28448
FT                   /*tag= h
FT                   /label= MG024
FT                   /note= "Previously identified as MORF-19826 and
FT                   MORF-20093, the encoded protein shows 46.84
FT                   percentage identity to GTP-binding protein
FT                   from E. coli"
FT   cds             36987..38978
FT                   /*tag= i
FT                   /label= MG032
FT                   /note= "Previously identified as MORF-20099, the
FT                   encoded protein shows 26.82 percentage
FT                   identity to ATP-dependent nuclease (addA)
FT                   from B. subtilis"
FT   cds             39242..39904
FT                   /*tag= j
FT                   /label= MG033
FT                   /note= "Previously identified as MORF-20100, the
FT                   encoded protein shows 35.90 percentage
FT                   identity to glycerol uptake facilitator
FT                   (glpF) from B. subtilis"
FT   cds             complement (39873..40514)
FT                   /*tag= k
FT                   /label= MG034
FT                   /note= "Previously identified as MORF-20101, the
FT                   encoded protein shows 48.13 percentage
FT                   identity to thymidylate kinase (tdk)
FT                   from B. subtilis"
FT   cds             40543..41787
FT                   /*tag= l
FT                   /label= MG035
FT                   /note= "Previously identified as MORF-20102, the
```



```
FT                   encoded protein shows 30.71 percentage
FT                   identity to histidyl-tRNA synthetase (hisS)
FT                   from Mycobacterium leprae"
FT   cds             complement (44751..46277)
FT                   /*tag= m
FT                   /label= MG038
FT                   /note= "Previously identified as MORF-20105, the
FT                   encoded protein shows 46.83 percentage
FT                   identity to glycerol kinase (glpK)
FT                   from E. coli"
FT   cds             complement (46268..47422)
FT                   /*tag= n
FT                   /label= MG039
FT                   /note= "Previously identified as MORF-19831 and
FT                   MORF-20106, the encoded protein shows 43.20
FT                   percentage identity to glycerol-3-phosphate
FT                   dehydrogenase (GUT2) from S. cerevisiae"
FT   cds             49377..49643
FT                   /*tag= o
FT                   /label= MG041
FT                   /note= "The encoded protein shows 48.86 percentage
FT                   identity to phosphohistidinoprotein-hexose
FT                   phosphotransferase (ptsH) from Mycoplasma
FT                   capricolum"
FT   cds             50060..51520
FT                   /*tag= p
FT                   /label= MG042
FT                   /note= "Previously identified as MORF-19832 and
FT                   MORF-20108, the encoded protein shows 41.92
FT                   percentage identity to spermidine/
FT                   putrescine transport ATP-binding protein
FT                   (potA) from E. coli"
FT   cds             51525..52382
FT                   /*tag= q
FT                   /label= MG043
FT                   /note= "Previously identified as MORF-20110, the
FT                   encoded protein shows 26.51 percentage
FT                   identity to spermidine/putrescine transport
FT                   system permease protein (potB) from E. coli"
FT   cds             52366..53220
FT                   /*tag= r
FT                   /label= MG044
FT                   /note= "Previously identified as MORF-20111, the
FT                   encoded protein shows 29.45 percentage
FT                   identity to spermidine/putrescine transport
FT                   system permease protein C (potC) from E. coli"
FT   cds             54658..55605
FT                   /*tag= s
FT                   /label= MG046
FT                   /note= "Previously identified as MORF-20112, the
FT                   encoded protein shows 36.60 percentage
```



```
FT                   identity to sialoglycoprotease (gcp)
FT                   from Pasteurella haemolytica"
FT   cds             complement (56970..58310)
FT                   /*tag= t
FT                   /label= MG048
FT                   /note= "Previously identified as MORF-19834,
FT                   MORF-20114 and MORF-20115, the encoded protein
FT                   shows 43.02 percentage identity to signal
FT                   recognition particle protein (ffh) from B.
FT                   subtilis"
FT   cds             58117..59079
FT                   /*tag= u
FT                   /label= MG049
FT                   /note= "Previously identified as MORF-20114 and
FT                   MORF-20115, the encoded protein shows 44.78
FT                   percentage identity to purine-nucleoside
FT                   phosphorylase (deoD) from E. coli"
FT   cds             59083..59754
FT                   /*tag= v
FT                   /label= MG050
FT                   /note= "Previously identified as MORF-20117, the
FT                   encoded protein shows 83.03 percentage
FT                   identity to deoxyribose-phosphate aldolase
FT                   (deoC) from Mycoplasma pneumoniae"
FT   cds             complement (64898..65731)
FT                   /*tag= w
FT                   /label= MG056
FT                   /note= "Previously identified as MORF-20122, the
FT                   encoded protein shows 30.25 percent
FT                   identity to the protein disclosed in
FT                   GB:D26185_99 from B. subtilis"
FT   cds             complement (65713..66249)
FT                   /*tag= x
FT                   /label= MG057
FT                   /note= "Previously identified as MORF-20123, the
FT                   encoded protein shows 38.90 percentage
FT                   identity to the protein disclosed in
FT                   GB:D26185_104 from B. subtilis"
FT   cds             81047..82597
FT                   /*tag= y
FT                   /label= MG067
FT                   /note= "Previously identified as MORF-19845, the
FT                   encoded protein shows 28.84 percentage
FT                   identity to glutamic acid specific protease
FT                   (SPase) from Staphylococcus aureus"
FT   cds             91065..91919
FT                   /*tag= z
FT                   /label= MG070
FT                   /note= "Previously identified as MORF-20136, the
FT                   encoded protein shows 34.8 percentage
FT                   identity to ribosomal protein S2 (rpS2)
```



```
FT                   from Spirulina plantensis"
FT   cds             103104..104324
FT                   /*tag= aa
FT                   /label= MG077
FT                   /note= "Previously identified as MORF-20140, the
FT                   encoded protein shows 28.05 percentage
FT                   identity to oligopeptide transport system
...
Note: remainder of annotations omitted.

  Query Match           5.7%;  Score 17;  DB 27;  Length 580073;
  Best Local Similarity 100.0%;  Pred. No. 3.99e+00;
  Matches   17;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    856 atttaaataaaaatatt 872
          |||||||||||||||||
Cp    123 ATTTAAATAAAAATATT 107


RESULT    7
ID   T26241 standard; cDNA to mRNA; 265 BP.
AC   T26241;
DT   13-NOV-1996  (first entry)
DE   Human gene signature HUMGS08480.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; JO1916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 2036; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
```

```
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   265 BP;    79 A;    42 C;    56 G;    86 T;

  Query Match            5.4%;  Score 16;  DB 23;  Length 265;
  Best Local Similarity 100.0%;  Pred. No. 1.55e+01;
  Matches    16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    153 acttttagaaacatgt 168
          ||||||||||||||||
Qy    137 ACTTTTAGAAACATGT 152


RESULT    8
ID   V04278 standard; cDNA; 296 BP.
AC   V04278;
DT   22-JUN-1998  (first entry)
DE   Secreted protein C195_4 cDNA 3' end.
KW   Secreted protein; C195_4; peripheral blood mononuclear cell; PBMC;
KW   protein factor; human; ds.
OS   Homo sapiens.
PN   WO9748800-A1.
PD   24-DEC-1997.
PF   16-JUN-1997; U10500.
PR   25-OCT-1996; US-738367.
PR   17-JUN-1996; US-664596.
PR   27-SEP-1996; US-721926.
PA   (GEMY ) GENETICS INST INC.
PI   Jacobs K, Lavallie ER, McCoy JM, Merberg D, Racie LA,
PI   Spaulding V, Treacy M;
DR   WPI; 98-063141/06.
PT   Nucleic acid encoding secreted protein from human peripheral blood
PT   mononuclear cells - useful, e.g. as immunomodulators, antitumour
PT   agents, promoters of tissue growth, haemostatic and thrombolytic
PT   agents etc.
PS   Disclosure; Page 37; 52pp; English.
CC   This cDNA sequence comprises the 3' end, including the polyA tail,
CC   of cDNA clone C195_4 (see V04274) that codes for a novel secreted
CC   protein (see W41601) of human peripheral blood mononuclear cells
CC   (PBMC).  Clone C195_4 (ATCC 98079) was isolated from a PBMC cDNA
CC   library using methods which are selective for cDNAs encoding
CC   secreted proteins.  C195_4 can be used to express recombinant
CC   secreted proteins; as a tissue/molecular weight marker; for
CC   chromosome identification; to identify possible genetic disorders;
CC   to isolate new related DNA; as a source of PCR primers; to generate
CC   anti-protein or anti-DNA antibodies; in interaction trap assays to
CC   identify sequences that encode interacting proteins etc.  The
CC   C195_4 secreted protein may have a number of biological activities.
```



```
SQ   Sequence   296 BP;    93 A;    50 C;    48 G;    100 T;

  Query Match            5.4%;  Score 16;  DB 40;  Length 296;
  Best Local Similarity 100.0%;  Pred. No. 1.55e+01;
  Matches    16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    260 cattaaaaaatatttttt 275
          ||||||||||||||||
Qy    100 CATTAAAAATATTTTT 115


RESULT    9
ID   T20533 standard; cDNA to mRNA; 364 BP.
AC   T20533;
DT   31-JUL-1996  (first entry)
DE   Human gene signature HUMGS01746.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 662; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   364 BP;    108 A;    60 C;    73 G;    122 T;

  Query Match            5.4%;  Score 16;  DB 19;  Length 364;
```



```
  Best Local Similarity 100.0%;  Pred. No. 1.55e+01;
  Matches    16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    338 cattaaaaaatatttttt 353
          ||||||||||||||||
Qy    100 CATTAAAAATATTTTT 115


RESULT   10
ID   Q78184 standard; DNA; 399 BP.
AC   Q78184;
DT   02-JUN-1995  (first entry)
DE   Maize associated region MAR3 (SAR(L))
KW   Genetic stabilising element; matrix associated region; MAR; ss.
OS   Zea mays
FH   Key             Location/Qualifiers
FT   misc_feature    43..69
FT                   /*tag=  a
FT                   /label= ARS(11/12) A BOX(9/10) ARS(10/12)
FT   misc_feature    99..109
FT                   /*tag=  b
FT                   /label= T BOX (8/10)
FT   misc_feature    116..138
FT                   /*tag=  c
FT                   /label= AT BOX T BOX(9/10) ARS(10/12) T BOX(9/10)
FT   misc_feature    216..245
FT                   /*tag=  d
FT                   /label= ARS(10/12) ARS(10/12) ARS(10/12)
FT   misc_feature    259..271
FT                   /*tag=  e
FT                   /label= T BOX(9/10)
FT   misc_feature    308..331
FT                   /*tag=  f
FT                   /label= AT BOX ARS(10/12) T BOX(9/10)
FT   misc_feature    334..348
FT                   /*tag=  g
FT                   /label= TOPO II (15/15)
FT   misc_feature    379..389
FT                   /*tag=  h
FT                   /label= T BOX (9/10)
PN   WO9424293-A.
PD   27-OCT-1994.
PF   18-APR-1994; E01193.
PR   19-APR-1993; US-049564.
PA   (SANO ) SANDOZ LTD.
PA   (SANO ) SANDOZ PATENT GMBH.
PA   (SANO ) SANDOZ-ERFINDUNGEN VERW GES MBH.
PI   De HAAS JM,  Dietrich PS,  Gielen JJL,  Van DRIEL R;
PI   Van GRINSVEN MQJM;
DR   WPI; 94-341872/42.
PT   Stabilised exogenous genes for plant cell transformation - with
```



```
PT   at least one stabilising DNA segment in 3'- or 5'-flanking region
PS   Example; Page 36; 53pp; English.
CC   The term MAR refers to DNA segments isolated from nuclear scaffold
CC   or nuclear matrix prepns. after endonuclease treatment. Two
CC   fragments, MAR1 (maize 0.8 kb AT rich region) and MAR2 (maize 1.25
CC   kb regions with ARS3), found withing a 5 kb maize EcoR1 fragment
CC   have nuclear matrix binding activity. The sequence of the MAR1
CC   fragment is given in Q78182 and that portion of MAR2 not previously
CC   sequenced is given in Q78183 along with the published portion of
CC   MAR2 named ARS3 (Berlani et al. 1988 Plant Molecular Biology 11:
CC   173-182). Additionally the sequence from the SAR(L) a region from a
CC   small heat shock gene (MAR3) (soybean HSP17.6 0.4 kb SARL) is shown
CC   in Q78184 (Schoffl et al. Transgenic Res. 2 93-100 (1993)). MAR1,
CC   MAR2 and MAR3 were cloned to construct plasmids pZ01927, pZ01929
CC   and pSVB20-SAR(L) for gene activity.
SQ   Sequence   399 BP;    166 A;    34 C;    39 G;    160 T;

  Query Match            5.4%;  Score 16;  DB 13;  Length 399;
  Best Local Similarity 100.0%;  Pred. No. 1.55e+01;
  Matches    16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    111 aataaaaaatattttta 126
          ||||||||||||||||
Cp    118 AATAAAAATATTTTTA 103


RESULT   11
ID   Q45345 standard; cDNA to mRNA; 1218 BP.
AC   Q45345;
DT   09-NOV-1994  (first entry)
DE   Rat 20 alpha-hydroxysteroid dehydrogenase gene.
KW   20-alpha-HSD; hydroxysteroid dehydrogenase; hormone production;
KW   cancer; tumour; diagnosis; ds.
OS   Rattus sp.
FH   Key             Location/Qualifiers
FT   5'utr           1..30
FT                   /*tag=  a
FT   cds             31..1002
FT                   /*tag=  b
FT                   /EC_number= 1.1.1.149
FT   3'utr           1003..1207
FT                   /*tag=  c
FT   polya_signal    1179..1184
FT                   /*tag=  d
FT   misc_feature    1208
FT                   /*tag=  e
FT                   /standard_name= polyA_site
PN   J06062863-A.
PD   08-MAR-1994.
PF   11-AUG-1992; 213944.
PR   11-AUG-1992; JP-213944.
```

```
PA   (SAKA ) OTSUKA PHARM CO LTD.
DR   WPI; 94-121125/15.
DR   P-PSDB; R52981.
PT   Rat 20 alpha-HSD gene - used for diagnosis of tumour and disease
PT   due to abnormal hormone production
PS   Claim 2; Page 12-14; 18pp; Japanese.
CC   The rat 20-alpha-HSD gene was isolated from a rat ovary lambda ZAP
CC   cDNA library. The gene can be used as a probe for the human 20 alpha-
CC   HSD gene and is useful for diagnosis of tumours and diseases due to
CC   abnormal hormone production.
SQ   Sequence   1218 BP;    354 A;    270 C;    274 G;    320 T;

  Query Match           5.4%;  Score 16;  DB 10;  Length 1218;
  Best Local Similarity 100.0%;  Pred. No. 1.55e+01;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1181 ttaaataaaaatatttt 1196
          ||||||||||||||||
Cp    120 TAAATAAAAATATTTT 105


RESULT   12
ID   N60225 standard; DNA; 1393 BP.
AC   N60225;
DT   08-JUN-1991  (first entry)
DE   Sequence of autonomous replication sequence (ARS) on pCYG 97.
KW   Yeast expression vector; ds.
OS   Acremonium chrysogenum.
PN   EP-187514-A.
PD   16-JUL-1986.
PF   20-DEC-1985; 309365.
PR   02-JAN-1985; GB-000040.
PR   17-OCT-1985; GB-025667.
PA   (FUJI ) FUJISAWA PHARM KK.
PI   Isogai T, Yoshida M;
DR   WPI; 86-184452/29.
PT   Hybrid plasmid(s) comprising DNA fragments - functioning as
PT   autonomous replication sequences in Saccharomyces cerevisiae and
PT   Acremonium chrysogenum
PS   Claim 10; Fig 8; 40pp; English.
CC   The mol. size of the ARS is 1.39 kbp. pCYG is useful as a vector in
CC   recombinant DNA work in which Acremonium chrysogenum or
CC   Saccharomyces cerevisiae is used as a host.
SQ   Sequence   1393 BP;    461 A;    185 C;    167 G;    579 T;

  Query Match           5.4%;  Score 16;  DB 2;  Length 1393;
  Best Local Similarity 100.0%;  Pred. No. 1.55e+01;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1007 atatttttatttaaat 1022
          ||||||||||||||||
```



```
Qy    108 ATATTTTTATTTAAAT 123


RESULT   13
ID   T15232 standard; cDNA; 1694 BP.
AC   T15232;
DT   09-DEC-1996  (first entry)
DE   Tumor necrosis factor receptor 1 death domain ligand cDNA (clone 1TU).
KW   TNF-R1-DD; tumour necrosis factor receptor 1 death domain; inhibitor;
KW   P55; anti-inflammatory; autoimmune disease; graft versus host reaction;
KW   osteoporosis; cachexia; diabetes; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             2..934
FT                   /*tag=  a
PN   WO9612735-A1.
PD   02-MAY-1996.
PF   12-OCT-1995; U12724.
PR   19-OCT-1994; US-327514.
PR   19-JUN-1995; US-494440.
PR   26-SEP-1995; US-533901.
PA   (GEMY ) GENETICS INST INC.
PI   Chen J,  Graham J,  Lin L,  Schievella AR;
DR   WPI; 96-230551/23.
DR   P-PSDB; R95330.
PT   TNF receptor death domain ligand proteins and inhibitors of ligand
PT   binding  - for prevention and treatment of pref. anti-inflammatory
PT   conditions, e.g. auto-immune disease, graft versus host reaction
PT   osteoporosis, etc.
PS   Claim 29; Page 43-45; 83pp; English.
CC   The present sequence, clone 1TU (ATCC 69848), encodes a tumour necrosis
CC   factor (TNF) receptor 1 (R1) death domain (DD) ligand (see R95330). A
CC   yeast genetic selection method, the "interaction trap", was used to
CC   screen W138 cell cDNA libraries for proteins that interact with the DD
CC   of the P55 type 1 TNF-R. Clone 1TU is believed to be a partial cDNA
CC   clone of a longer full length coding sequence (not given in the
CC   specification), however the protein encoded by 1TU does bind the DD of
CC   TNF-R1. TNF-R1-DD ligands and their inhibitors are useful in the
CC   prevention and treatment of anti-inflammatory conditions and other
CC   conditions such as cachexia, autoimmune disease, graft versus host
CC   reaction, osteoporosis, diabetes, etc.
SQ   Sequence   1694 BP;    365 A;    488 C;    455 G;    386 T;

  Query Match           5.4%;  Score 16;  DB 23;  Length 1694;
  Best Local Similarity 100.0%;  Pred. No. 1.55e+01;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1652 ttaaataactttattt 1667
          ||||||||||||||||
Qy    118 TTAAATAACTTTATTT 133
```



```
RESULT   14
ID   T94635 standard; cDNA; 1694 BP.
AC   T94635;
DT   19-MAR-1998  (first entry)
DE   TNF-R1-DD ligand protein clone 1TU coding sequence.
KW   Tumour necrosis factor receptor P55 type; TNF-R1-DD ligand protein;
KW   death domain; TNF-R1; inhibitor identification; TNF-induced condition;
KW   insulin-like growth factor binding protein-5; inflammatory condition;
KW   IGFBP-5; therapy; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             2..934
FT                   /*tag=  a
PN   WO9730084-A1.
PD   21-AUG-1997.
PF   11-FEB-1997; U02146.
PR   15-AUG-1996; US-698551.
PR   15-FEB-1996; US-602228.
PA   (GEMY ) GENETICS INST INC.
PI   Chen J, Graham J, Lin L, Schievella AR;
DR   WPI; 97-424976/39.
DR   P-PSDB; W35573.
PT   Tumour necrosis factor receptor P55 type death domain ligand
PT   proteins - useful for preventing or ameliorating inflammatory
PT   conditions
PS   Claim 29; Page 46-48; 103pp; English.
CC   This sequence represents the coding sequence of the invention. This
CC   sequence was isolated from clone 20DD and encodes a tumour necrosis
CC   factor receptor P55 type (TNF-R1) death domain (DD) ligand protein. A
CC   host cell containing this sequence is used for the recombinant production
CC   of TNF-R1-DD. The TNF-R1-DD ligand protein can be used in a method to
CC   identify inhibitors of TNF-R DD binding. The TNF-R1-DD ligand protein,
CC   insulin-like growth factor binding protein-5 (IGFBP-5, has TNF-R1-DD
CC   ligand activity) or inhibitors of TNF-R1-DD ligand protein are capable of
CC   preventing or ameliorating an inflammatory condition, preferably by
CC   inhibiting TNF-R DD binding. Identification and isolation of ligands
CC   allows their effects upon TNF-R signal transduction and use as
CC   therapeutic agents for treatment of TNF-induced conditions to be
CC   examined.
SQ   Sequence   1694 BP;    365 A;    488 C;    455 G;    386 T;

  Query Match           5.4%;  Score 16;  DB 36;  Length 1694;
  Best Local Similarity 100.0%;  Pred. No. 1.55e+01;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1652 ttaaataactttattt 1667
          ||||||||||||||||
Qy    118 TTAAATAACTTTATTT 133
```



```
RESULT   15
ID   Q14669 standard; DNA; 3543 BP.
AC   Q14669;
DT   04-FEB-1992  (first entry)
DE   Dipteran active toxin gene.
KW   Insecticide; B.t.; crystal; delta endotoxin; cryIVA; ss.
OS   Bacillus thurigiensis var. morrisoni (PS71M3).
FH   Key             Location/Qualifiers
FT   cds             1..3543
FT                   /*tag=  a
PN   EP-457498-A.
PD   21-NOV-1991.
PF   09-MAY-1991; 304180.
PR   15-MAY-1990; US-524255.
PR   01-OCT-1990; US-590903.
PA   (MYCO-) MYCOGEN CORP.
PI   Sick AJ;
DR   WPI; 91-341902/47.
DR   P-PSDB; R14373.
PT   Bacillus thuringiensis genes encoding diptera-active toxins - and
PT   transformed microbes used to control insects in various
PT   environments.
PS   Claim 1; Page 10; 20pp; English.
CC   The sequence was obtd. from plasmid pMYC1625 which was isolated
CC   from a genomic library prepd. from DNA from B.t. PS71M3 (from
CC   B.t. PS71M3-69 (NRRL B-18515)].  It is related to the cryIVA
CC   family of genes, the 140 kD endotoxin gene and the type II gene
CC   from B.t. var. israelensis.  The gene encodes a 130 kD protein.
CC   Microorganisms transformed with the DNA may be administered to
CC   dipteran insects or their environments, the expressed toxins
CC   acting as an insecticide.
CC   See also Q14670-Q14672.
SQ   Sequence   3543 BP;    1278 A;    567 C;    612 G;    1086 T;

  Query Match           5.4%;  Score 16;  DB 3;  Length 3543;
  Best Local Similarity 100.0%;  Pred. No. 1.55e+01;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1406 taaataaaaatatttt 1421
          ||||||||||||||||
Cp    120 TAAATAAAAATATTT 105


RESULT   16
ID   Q81178 standard; DNA; 3543 BP.
AC   Q81178;
DT   12-AUG-1995  (first entry)
DE   B.t. toxin PS71M3 gene.
KW   Delta-endotoxin; crystal protein; biological control agent;
KW   Calliphorid; screw-worm; sheep blowfly; Lucilla; Phormia;
KW   Calliphora; insecticide; pesticide; B.t.; ss.
```

```
OS   Bacillus thuringiensis.
PN   WO9502694-A.
PD   26-JAN-1995.
PF   13-JUL-1994; U07902.
PR   15-JUL-1993; US-093199.
PA   (MYCO ) MYCOGEN CORP.
PI   Hickle LA, Payne J;
DR   WPI; 95-067338/09.
DR   P-PSDB; R63078.
PT   Method for controlling Calliphoridae pests - specifically
PT   utilises Bacillus thuringiensis isolates or toxins.
PS   Disclosure; Page 36-38; 50pp; English.
CC   A library was constructed from Bacillus thuringiensis PS71M3
CC   total cellular DNA in lambda Gem-11.  Plasmid pMYC1625, selected
CC   in Escherichia coli, contained a 8.0 kb insert expressing a
CC   beta-endotoxin gene.  This was sequenced (Q81178).  A cured,
CC   acrystalliferous B.t. host carrying pMYC1625 produced a 130 kDa
CC   crystal protein (R63078) toxic to calliphorids.
SQ   Sequence   3543 BP;    1278 A;    566 C;    612 G;    1087 T;

  Query Match           5.4%;  Score 16;  DB 13;  Length 3543;
  Best Local Similarity 100.0%;  Pred. No. 1.55e+01;
  Matches    16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1406 taaataaaaaatatttt 1421
          ||||||||||||||||
Cp    120 TAAATAAAAATATTTT 105


RESULT   17
ID   N93054 standard; DNA; 3940 BP.
AC   N93054;
DT   30-MAY-1990  (first entry)
DE   Delta-endotoxin crystal protein gene.
KW   Delta-endotoxin; crystal protein; insecticide; ss; pCC130;
KW   biological control agent.
OS   Bacillus thurigiensis subsp israelensis.
FH   Key             Location/Qualifiers
FT   cds             891
FT                   /*tag=  a
FT                   /product=delta-endotoxin
FT   rbs             879..884
FT                   /*tag=  b
FT                   /label=Shine-Delgarno sequence
PN   EP-296870-A.
PD   28-DEC-1988.
PF   24-JUN-1988; 305772.
PR   26-JUN-1987; US-067653.
PA   (DUPO) Du Pont de Nemours Co.
PI   Ellar DJ, Ward ES;
DR   WPI; 89-001322/01.
```



```
DR   P-PSDB; P94035.
PT   DNA fragment encoding insecticidal protein - obtained from
PT   Bacillus thuringiensis subsp. israelensis, and used in
PT   microorganisms and plant cells.
PS   Disclosure; fig 5; 26pp; English.
CC   The nucleotide sequence is an insert in plasmid pCC130. The delta
CC   endotoxin protein is insecticidal and can be used to control insect pests
CC   esp. mosquitos.
CC   See also N93059.
SQ   Sequence   3940 BP;    1457 A;    603 C;    659 G;    1221 T;

  Query Match           5.4%;  Score 16;  DB 1;  Length 3940;
  Best Local Similarity 100.0%;  Pred. No. 1.55e+01;
  Matches    16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2296 taaataaaaaatatttt 2311
          ||||||||||||||||
Cp    120 TAAATAAAAATATTTT 105


RESULT   18
ID   N93059 standard; DNA; 4571 BP.
AC   N93059;
DT   30-MAY-1990  (first entry)
DE   Delta-endotoxin crystal protein gene.
KW   Delta-endotoxin; crystal protein; insecticide; pCH130; ss;
KW   biological control agent.
OS   Bacillus thurigiensis subsp israelensis.
FH   Key             Location/Qualifiers
FT   misc_feature    891..4430
FT                   /*tag=  a
FT                   /product=delta-endotoxin crystal protein
FT                   RBS                 879..884
FT                   /*tag=  b
FT                   /label= Shine-Dalgarno sequence.
PN   EP-296870-A.
PD   28-DEC-1988.
PF   24-JUN-1988; 305772.
PR   26-JUN-1987; US-067653.
PA   (DUPO) Du Pont de Nemours Co.
PI   Ellar DJ, Ward ES;
DR   WPI; 89-001322/01.
DR   P-PSDB; P93715.
PT   DNA fragment encoding insecticidal protein - obtained from
PT   Bacillus thuringiensis subsp. israelensis, and used in
PT   microorganisms and plant cells.
PS   Disclosure; fig 7; 26pp; English.
CC   The sequence encodes the 130 kDa delta-endotoxin gene from
CC   B. thuringiensis subsp. israelensisis  and is a 4.46 kb insert of pCH130.
CC   The delta endotoxin protein is insecticidal and can be used to control
CC   insect pests esp. mosquitos.
```



```
CC   See also N93054.
SQ   Sequence   4571 BP;    1678 A;    685 C;    817 G;    1391 T;

  Query Match           5.4%;  Score 16;  DB 1;  Length 4571;
  Best Local Similarity 100.0%;  Pred. No. 1.55e+01;
  Matches    16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2296 taaataaaaaatatttt 2311
          ||||||||||||||||
Cp    120 TAAATAAAAATATTTT 105


RESULT   19
ID   Q05264 standard; DNA; 4704 BP.
AC   Q05264;
DT   23-NOV-1990  (first entry)
DE   Sequence encoding bean chitinase gene fragment.
KW   Bean chitinase; phytopathogenic fungi; agrobacterium; ds.
OS   Phaseolus vulgaris var. SAXA.
FH   Key             Location/Qualifiers
FT   cds             2055..3035
FT                   /*tag=  a
PN   WO9007001-A.
PD   28-JUN-1990.
PF   13-DEC-1989; U05501.
PR   16-DEC-1988; US-285252.
PA   (DUPO ) DU PONT DE NEMOURS CO.
PI   BROGLIE KE, BROGLIE RM;
DR   WPI; 90-224529/29.
DR   P-PSDB; R05931.
PT   Recombinant DNA contg. high level promoter and plant chitinase
PT   gene - for transforming plants to over:express chitinase,
PT   imparting resistance to phytopathogenic fungi.
PS   Disclosure; pp; English.
CC   Sequence encoding chitinase may be used to transform mono-
CC   or dicotyledons to show increased resistance to fungal attack.
CC   A high level promoter is used from CMV  35s or 19s complement,
CC   NOS or OCS promoters of Agrobacterium opine synthetase gene,
CC   the RUBISCO small unit or the chlorophyl A/B binding protein.
SQ   Sequence   4704 BP;    1528 A;    783 C;    872 G;    1521 T;

  Query Match           5.4%;  Score 16;  DB 1;  Length 4704;
  Best Local Similarity 100.0%;  Pred. No. 1.55e+01;
  Matches    16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   3295 ttaaaaatatttttat 3310
          ||||||||||||||||
Qy    102 TTAAAAATATTTTTAT 117


RESULT   20
```



```
ID   N81490 standard; DNA; 4934 BP.
AC   N81490;
DT   05-NOV-1990  (first entry)
DE   Insecticidal (Diptera) protein gene.
KW   Insecticidal protein gene; Diptera; Aedes; ss.
OS   Bacillus thuringiensis israelensis.
FH   Key             Location/Qualifiers
FT   cds             1..4908
FT                   /*tag=  a
FT                   /product=insecticidal protein
FT   intron          620..1368
FT                   /*tag=  b
PN   J63230090-A.
PD   26-SEP-1988.
PF   19-MAR-1987; 066844.
PR   19-MAR-1987; JP-066844.
PA   (SUMO) Sumitomo Chem Ind KK.
DR   WPI; 88-311968/44.
DR   P-PSDB; P81034, P82589.
PT   New insecticidal protein of Bacillus thuringiensis israelensis strain
PT   - prepd. by forming gene library from plasmid contg. insecticidal protein
PT   gene, and transforming host cell.
PS   Disclosure; ; 9pp; Japanese.
CC   A plasmid contg. the gene is isolated by forming a gene library
CC   from plasmid DNA of B.thuringiensis israelensis (HD 522 strain, USA
CC   Goldberg ONR60) followed by screening with anti-israelensis insecticidal
CC   protein IG.
CC   The insecticidal protein is highly effective against Diptera, esp. Aedes.
SQ   Sequence   4934 BP;    1797 A;    749 C;    877 G;    1511 T;

  Query Match           5.4%;  Score 16;  DB 1;  Length 4934;
  Best Local Similarity 100.0%;  Pred. No. 1.55e+01;
  Matches    16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   2774 taaataaaaaatatttt 2789
          ||||||||||||||||
Cp    120 TAAATAAAAATATTTT 105


RESULT   21
ID   Q55753 standard; DNA; 4954 BP.
AC   Q55753;
DT   25-SEP-1995  (first entry)
DE   Escherichia coli genomic probe EC-34.
KW   Probe; S.aureus; S.epidermis; E.faecalis; P.aeruginosa; E.coli;
KW   K.pneumoniae; E.cloacae; clinical sample; ds.
OS   Escherichia coli.
PN   WO9401583-A.
PD   20-JAN-1994.
PF   07-JUL-1993; J00936.
PR   07-JUL-1992; JP-179719.
```

```
PA   (FUSO ) FUSO PHARM IND LTD.
PA   (OHNO/) OHNO T.
PI   Eda S,  Matsuhisa A,  Ohno T,  Uehara H;
DR   WPI; 94-035086/04.
PT   Probe for identifying bacteria causing infectious disease -
PT   consists of a DNA fragment obtained by HindIII cleavage of the
PT   pathogenic bacterial genomic DNA
PS   Claim 7; Page 64-67; 100pp; Japanese.
CC   The nucleotide sequence of a 4954 bp probe obtained by digestion of
CC   Escherichia coli genomic DNA with the restriction enzyme HindIII.
CC   The probes (Q55133-48 and Q55752-58) represent other probes derived,
CC   by HindIII digestion of the genomes of Staphylococcus aureus,
CC   S.epidermis, Enterococcus faecalis, Pseudomonas aeruginosa, E.coli,
CC   Klebsiella pneumoniae or Enterobacter cloacae.  The probes can be used
CC   to detect their respective microorganisms in clinical samples.
SQ   Sequence   4954 BP;    1491 A;    1051 C;    1007 G;    1405 T;

  Query Match           5.4%;  Score 16;  DB 14;  Length 4954;
  Best Local Similarity 100.0%;  Pred. No. 1.55e+01;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  4480 tatttttaatgcatta 4495
         ||||||||||||||||
Cp   110 TATTTTTAATGCATTA 95


RESULT   22
ID   T34137 standard; DNA; 37895 BP.
AC   T34137;
DT   30-JAN-1997  (first entry)
DE   Gene for biosynthetic enzyme group of icosapentaenoic acid synthase.
KW   Icosapentaenoic acid synthase; EPA; drugs; agrochemicals;
KW   foodstuffs; animal feed; lipid balance correction; antihypertensive;
KW   antiinflammatory; anticancer agent; ss.
OS   Shewanella putrefaciens.
FH   Key             Location/Qualifiers
FT   cds             6121..8103
FT                   /*tag=  a
FT                   /product= Icosapentaenoic acid biosynthetic enzyme
FT   cds             complement (9016..8186)
FT                   /*tag=  b
FT                   /product= Icosapentaenoic acid biosynthetic enzyme
FT   cds             9681..12590
FT                   /*tag=  c
FT                   /product= Icosapentaenoic acid biosynthetic enzyme
FT   cds             13040..13903
FT                   /*tag=  d
FT                   /product= Icosapentaenoic acid biosynthetic enzyme
FT   cds             13906..22173
FT                   /*tag=  e
FT                   /product= Icosapentaenoic acid biosynthetic enzyme
```



```
FT   cds             22176..24515
FT                   /*tag=  f
FT                   /product= Icosapentaenoic acid biosynthetic enzyme
FT   cds             24518..30529
FT                   /*tag=  g
FT                   /product= Icosapentaenoic acid biosynthetic enzyme
FT   cds             30730..32358
FT                   /*tag=  h
FT                   /product= Icosapentaenoic acid biosynthetic enzyme
FT   cds             32753..34327
FT                   /*tag=  i
FT                   /product= Icosapentaenoic acid biosynthetic enzyme
PN   WO9621735-A1.
PD   18-JUL-1996.
PF   12-JAN-1996; J00030.
PR   13-JAN-1995; JP-004299.
PA   (SAGA ) SAGAMI CHEM RES CENTRE.
PI   Kato S,  Kondo K,  Yamada A,  Yazawa K;
DR   WPI; 96-342288/34.
DR   P-PSDB; R99458, R99459, R99460, R99461, R99462, R99463, R99464,
DR   R99465, R99466.
PT   Production of icosapentaenoic acid using transformed E. coli  - uses
PT   DNA coding for icosapentaenoic acid synthase derived from Shewanella
PT   strain
PS   Claim 1; Page 19-40; 145pp; English.
CC   The DNA sequence encoding the biosynthetic enzymes of
CC   icosapentaenoic acid (EPA) can be used to transform Escherichia coli.
CC   The DNA sequence allows efficient microbial production of EPA, which
CC   is a raw material for drugs, agrochemicals, foods and animal
CC   feedstuffs. EPA is also useful for lipid balance correction and as
CC   an antihypertensive, antiinflammatory and anticancer agent.
SQ   Sequence   37895 BP;    11266 A;    8624 C;    8032 G;    9972 T;

  Query Match           5.4%;  Score 16;  DB 24;  Length 37895;
  Best Local Similarity 100.0%;  Pred. No. 1.55e+01;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  12810 tttactacataaaatt 12825
          ||||||||||||||||
Cp     47 TTTACTACATAAAATT 32


RESULT   23
ID   Q51128 standard; DNA; 37913 BP.
AC   Q51128;
DT   27-MAY-1994  (first entry)
DE   Eicosapentaenoic acid (EPA) synthesis operon.
KW   EPA; eicosapentaenoic acid synthetase; drug; anticoagulant;
KW   hypolipemic; hypoglycemic; antihypertensive; anticancer; pesticide;
KW   foodstuff; additive; ss.
OS   Shewanella putrefaciens.
```



```
FH   Key             Location/Qualifiers
FT   cds             6121..8103
FT                   /*tag=  a
FT                   /product= Enzyme used in EPA biosynthetic pathway.
FT   cds             9681..12590
FT                   /*tag=  b
FT                   /product= Enzyme used in EPA biosynthetic pathway.
FT   cds             13905..18806
FT                   /*tag=  c
FT                   /product= Enzyme used in EPA biosynthetic pathway.
FT   cds             18997..22170
FT                   /*tag=  d
FT                   /product= Enzyme used in EPA biosynthetic pathway.
FT   cds             22173..24512
FT                   /*tag=  e
FT                   /product= Enzyme used in EPA biosynthetic pathway.
FT   cds             24515..30151
FT                   /*tag=  f
FT                   /product= Enzyme used in EPA biosynthetic pathway.
FT   cds             30727..32355
FT                   /*tag=  g
FT                   /product= Enzyme used in EPA biosynthetic pathway.
FT   cds             32750..34345
FT                   /*tag=  h
FT                   /product= Enzyme used in EPA biosynthetic pathway.
PN   WO9323545-A.
PD   25-NOV-1993.
PF   14-MAY-1993; J00641.
PR   15-MAY-1992; JP-147945.
PA   (SAGA ) SAGAMI CHEM RES CENTRE.
PI   Kato S,  Kondo K,  Yamada A,  Yazawa K;
DR   WPI; 93-386577/48.
DR   P-PSDB; R42450-57.
PT   Gene coding for eicosa-penta:enoic acid synthetase - is isolated
PT   from Pseudomonas, Alteromonas or Shewanella and used for
PT   recombinant prodn. of eicosa-penta:enoic acid
PS   Example 3; Page 16-39; 106pp; Japanese.
CC   EPA is useful as a drug, having anticoagulant, hypolipemic,
CC   hypoglycemic, antihypertensive and anticancer activity.  It is also
CC   a pesticide and is useful as a nutritional foodstuff and animal feed
CC   additive.
SQ   Sequence   37913 BP;    11275 A;    8689 C;    7966 G;    9980 T;

  Query Match           5.4%;  Score 16;  DB 10;  Length 37913;
  Best Local Similarity 100.0%;  Pred. No. 1.55e+01;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  12810 tttactacataaaatt 12825
          ||||||||||||||||
Cp     47 TTTACTACATAAAATT 32
```



```
RESULT   24
ID   T13635 standard; DNA; 133894 BP.
AC   T13635;
DT   03-SEP-1996  (first entry)
DE   AcNPV genomic DNA clone 6.
KW   Autographa californica nuclear polyhedrosis virus clone 6;
KW   disruption; non-essential gene; heterologous protein production;
KW   expression vector; baculovirus; ss.
OS   Autographa californica nuclear polyhedrosis virus clone 6.
PN   WO9601320-A2.
PD   18-JAN-1996.
PF   30-JUN-1995; IB0578.
PR   04-JUL-1994; GB-013420.
PA   (NATU-) NATURAL ENVIRONMENT RES COUNCIL.
PI   Ayres M,  Bishop D,  Possee R;
DR   WPI; 96-087670/09.
DR   GENBANK; L22858.
PT   Autographa californica nuclear polyhedrosis virus complete genome
PT   sequence  - useful in the prodn. of vectors for enhanced
PT   heterologous protein expression, such as interleukin(s),
PT   interferon(s) and neurotoxin(s)
PS   Disclosure; Page 90-186; 122pp; English.
CC   The complete nucleotide sequence of the genome of clone 6 of the
CC   baculovirus Autographa californica nuclear polyhedrosis virus (AcNPV)
CC   has been determined.  The sequence is taken from the Genbank record
CC   L22858. The patent specification claims a polynucleotide selected from
CC   open reading frames (ORFs 13, 20, 22-26, 28-30, 32, 38, 41-46, 50-60,
CC   62-63, 66, 68-79, 81-87, 91-92, 96-98, 101-103, 106-126, 129-130,
CC   140-146, 148-150, 152 and 154 from a total of 154 ORFs identified by
CC   the patentees. See T13636-731. Expression vectors contg. the complete
CC   genomic sequence of AcNPV, with the exception that at least one non-
CC   essential ORF is disrupted or replaced are useful for the synthesis of
CC   heterologous proteins.
SQ   Sequence   133894 BP;    39195 A;    27151 C;    27347 G;    40201 T;

  Query Match           5.4%;  Score 16;  DB 17;  Length 133894;
  Best Local Similarity 100.0%;  Pred. No. 1.55e+01;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  27000 ttatttaaataaaaat 27015
          ||||||||||||||||
Cp    125 TTATTTAAATAAAAAT 110


RESULT   25
ID   Q04525 standard; DNA; 134525 BP.
AC   Q04525;
DT   01-OCT-1990  (first entry)
DE   Total base sequence of rice plant chloroplast DNA.
KW   Chloroplast; true grass; rice plant; ss.
```

```
OS   Oryza sativa.
PN   J02100682-A.
PD   12-APR-1990.
PF   07-OCT-1988; 251967.
PR   07-OCT-1988; JP-251967.
PA   (MITK) Mitsui Toatsu Chem Inc.
DR   WPI; 90-159709/21.
PT   Chloroplast DNA of true grasses - used to produce various
PT   DNA base sequences by decomposition of rice plant DNA.
PS   Claim 1; Fig 1; 20pp; Japanese.
CC   The sequence is that of the whole of rice chloroplast DNA.
SQ   Sequence   134525 BP;   41249 A;   26129 C;   26331 G;   40816 T;

  Query Match           5.4%;  Score 16;  DB 1;  Length 134525;
  Best Local Similarity 100.0%;  Pred. No. 1.55e+01;
  Matches   16;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   4143 attttatttaaataa 4158
          ||||||||||||||||
Qy    110 ATTTTTATTTAAATAA 125


RESULT   26
ID   T25856 standard; cDNA to mRNA; 66 BP.
AC   T25856;
DT   22-OCT-1996  (first entry)
DE   Human gene signature HUMGS08086.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 1944; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
```



```
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   66 BP;   22 A;   10 C;   9 G;   24 T;

  Query Match           5.1%;  Score 15;  DB 22;  Length 66;
  Best Local Similarity 100.0%;  Pred. No. 5.76e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     35 tatgtagtaaataaa 49
          |||||||||||||||
Qy     37 TATGTAGTAAATAAA 51


RESULT   27
ID   T19710 standard; cDNA to mRNA; 146 BP.
AC   T19710;
DT   05-JUL-1996  (first entry)
DE   Human gene signature HUMGS00784.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 455; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
```



```
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   146 BP;   58 A;   16 C;   20 G;   52 T;

  Query Match           5.1%;  Score 15;  DB 18;  Length 146;
  Best Local Similarity 100.0%;  Pred. No. 5.76e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     71 accaaaaagtttatt 85
          |||||||||||||||
Cp    283 ACCAAAAAGTTTATT 269


RESULT   28
ID   T24769 standard; cDNA to mRNA; 163 BP.
AC   T24769;
DT   09-OCT-1996  (first entry)
DE   Human gene signature HUMGS06844.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 1690; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
```



```
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   163 BP;   62 A;   18 C;   25 G;   57 T;

  Query Match           5.1%;  Score 15;  DB 21;  Length 163;
  Best Local Similarity 100.0%;  Pred. No. 5.76e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     53 atatttttaatgcat 67
          |||||||||||||||
Cp    111 ATATTTTTAATGCAT 97


RESULT   29
ID   N90214 standard; DNA; 205 BP.
AC   N90214;
DT   1-NOV-1989  (first entry)
DE   Malaria-specific DNA insert of clone 41-5.
KW   DNA; malaria; clone 41-5; Plasmodium falciparum; protein; vaccine.
OS   Plasmodium falciparum
FH   Key             Location/Qualifiers
FT   cds             3..203
FT                   /*tag= a
PN   EP-322712-A.
PD   05-JUL-1989.
PF   20-DEC-1988; 121299.
PR   30-DEC-1987; DE-831351.
PA   (BEHW) Behringwerke.
PI   Knapp B, Hundt E, Enders B, Kupper H;
DR   WPI; 89-194071/27.
DR   P-PSDB; P90406.
PT   New antigenic proteins from Plasmodium falciparum - new encoding nucleic
PT   acid sequences and derived antibodies, useful in vaccines, diagnosis etc.
PS   Claim 1; page 23 and Table 5; 25pp; German.
CC   The DNA is of a malaria-specific DNA insert of clone 41-5. The encoded
CC   peptide (see P90406) is useful in a protective vaccine, esp. against
CC   malaria. Produced antibodies are useful for passive immunisation, and
CC   antibodies, the protein and the DNA sequence are useful as diagnostic
CC   reagents. See also N90211-25 and P90403-19.
SQ   Sequence   205 BP;   89 A;   20 C;   25 G;   71 T;

  Query Match           5.1%;  Score 15;  DB 1;  Length 205;
  Best Local Similarity 100.0%;  Pred. No. 5.76e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    161 aatgaaaaatactga 175
          |||||||||||||||
Cp    193 AATGAAAAAATACTGA 179


RESULT   30
```

```
ID   T25769 standard; cDNA to mRNA; 246 BP.
AC   T25769;
DT   10-OCT-1996  (first entry)
DE   Human gene signature HUMGS07997.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 1924; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   246 BP;    72 A;    45 C;    66 G;    57 T;

  Query Match           5.1%;  Score 15;  DB 21;  Length 246;
  Best Local Similarity  88.2%;  Pred. No. 5.76e+01;
  Matches    15;  Conservative    0;  Mismatches   2;  Indels   0;  Gaps   0;

Db    222 antaaaaatattttnat 238
          | ||||||||||||| ||
Qy    101 ATTAAAAATATTTTTAT 117


RESULT   31
ID   Q50336 standard; cDNA; 271 BP.
AC   Q50336;
DT   24-NOV-1993  (first entry)
```



```
DE   Serglycin - 3' end.
KW   Haematopoietic cell secretory granule proteoglycan; HL-60;
KW   leukemia cell line; peptide core; ss.
OS   Homo sapiens.
PN   WO9313119-A.
PD   08-JUL-1993.
PF   23-DEC-1992; U11194.
PR   03-JAN-1992; US-816289.
PR   02-JUL-1992; US-906871.
PA   (BGHM ) BRIGHAM & WOMENS HOSPITAL.
PI   Avraham S,  Stevens RL;
DR   WPI; 93-227261/28.
PT   Transcription regulatory elements of ser-glycin gene - specific
PT   for haematopoietic cells, also trans-acting transcriptional
PT   binding factors
PS   Example 1; Page 43; 112pp; English.
CC   Individual HL-60 cell-derived cDNAs and their subcloned fragments
CC   were inserted into M13mp18 and M13mp19.  The consensus nucleotide
CC   sequence of the HL-60 derived secretory granule proteoglycan
CC   peptide core cDNA is represented in Q44277.  A 249 bp EcoRI-EcoRI
CC   fragment was isolated from a EcoRI digest of cDNA-H19.
CC   The nucleotide sequence of this fragment (Q50336) contains the
CC   sequence expected for a polyadenylation site and the poly(A)+ tail.
CC   This fragment hybridises to a genomic fragment that encodes the gene
CC   for this proteoglycan peptide core, and thus probalby represents the
CC   next 249 bp of the transcript.
SQ   Sequence   271 BP;    98 A;    36 C;    36 G;    101 T;

  Query Match           5.1%;  Score 15;  DB 7;  Length 271;
  Best Local Similarity 100.0%;  Pred. No. 5.76e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    172 aaattggcaggtaat 186
          ||||||||||||||| 
Qy     49 AAATTGGCAGGTAAT 63


RESULT   32
ID   T26536 standard; cDNA to mRNA; 304 BP.
AC   T26536;
DT   10-OCT-1996  (first entry)
DE   Human gene signature HUMGS08780.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
```



```
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 2110; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   304 BP;    90 A;    84 C;    63 G;    66 T;

  Query Match           5.1%;  Score 15;  DB 21;  Length 304;
  Best Local Similarity 100.0%;  Pred. No. 5.76e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    277 ttatttaaataaaaa 291
          |||||||||||||||
Cp    125 TTATTTAAATAAAAA 111


RESULT   33
ID   T19975 standard; cDNA to mRNA; 351 BP.
AC   T19975;
DT   17-JUL-1996  (first entry)
DE   Human gene signature HUMGS01107.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
```



```
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 525; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   351 BP;    117 A;    44 C;    44 G;    141 T;

  Query Match           5.1%;  Score 15;  DB 19;  Length 351;
  Best Local Similarity  83.3%;  Pred. No. 5.76e+01;
  Matches    15;  Conservative    0;  Mismatches   3;  Indels   0;  Gaps   0;

Db     32 aannanaaatatttttaa 49
          ||  | ||||||||||| 
Cp    119 AAATAAAAATATTTTTAA 102


RESULT   34
ID   T22717 standard; cDNA to mRNA; 390 BP.
AC   T22717;
DT   14-AUG-1996  (first entry)
DE   Human gene signature HUMGS04375.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
```

```
PT   tissues
PS   Claim 1; Page 1198; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   390 BP;    98 A;    58 C;    70 G;    146 T;

  Query Match           5.1%;  Score 15;  DB 19;  Length 390;
  Best Local Similarity 100.0%;  Pred. No. 5.76e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    352 tttttggttgttaaa 366
          |||||||||||||||
Qy    276 TTTTTGGTTGTTAAA 290


RESULT   35
ID   T20524 standard; cDNA to mRNA; 391 BP.
AC   T20524;
DT   22-JUL-1996  (first entry)
DE   Human gene signature HUMGS01735.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 660; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
```



```
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   391 BP;    141 A;    61 C;    75 G;    113 T;

  Query Match           5.1%;  Score 15;  DB 19;  Length 391;
  Best Local Similarity 100.0%;  Pred. No. 5.76e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    221 tttagaaacatgtag 235
          |||||||||||||||
Qy    140 TTTAGAAACATGTAG 154


RESULT   36
ID   Q85373 standard; cDNA; 593 BP.
AC   Q85373;
DT   29-AUG-1995  (first entry)
DE   Growth factor-activated fic gene.
KW   Growth factor; fic gene; heparanase; heparin; heparan sulfate;
KW   arthritis; restenosis; cancer; wound healing; ss.
OS   Homo sapiens.
PN   WO9504158-A.
PD   09-FEB-1995.
PF   26-JUL-1994; U08207.
PR   29-JUL-1993; US-099866.
PR   13-OCT-1993; US-136117.
PA   (UPJO ) UPJOHN CO.
PI   Hoogwerf AJ,  Ledbetter SR;
DR   WPI; 95-082239/11.
DR   P-PSDB; R70803.
PT   Screening for cpds. with anti-heparanase activity - by detecting
PT   inhibition of heparin or heparan sulphate degradation,
PT   potentially useful for treating arthritis, restenosis, cancer.
PS   Disclosure; Page 53; 60pp; English.
CC   Purified heparanases, prepared under reducing conditions and
CC   activated with transglutaminase, are given in R70786-805.  Most
CC   are prepared by reverse transcription of mRNA from activated human
CC   leukocytes, then cloning of the cDNA into pVL1392 baculovirus
CC   vector, and expression in Sf9 cells in the presence of reduced
```



```
CC   glutathione and dithiothreitol.
SQ   Sequence   593 BP;    170 A;    136 C;    135 G;    152 T;

  Query Match           5.1%;  Score 15;  DB 14;  Length 593;
  Best Local Similarity 100.0%;  Pred. No. 5.76e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    497 aatatttttatttaa 511
          |||||||||||||||
Qy    107 AATATTTTTATTTAA 121


RESULT   37
ID   T48043 standard; DNA; 609 BP.
AC   T48043;
DT   11-JUN-1997  (first entry)
DE   Yeast chromosome IX-derived promoter sequence J1.
KW   Transcription; promoter; yeast; chromosome; phleomycin resistance; ss;
KW   pulsed field gel electrophoresis; reversion; E.coli; vector; expression;
KW   heterologous; cytokine; growth factor; receptor; anti-coagulant; enzyme;
KW   inhibitor; antibody; immunogenic polypeptide.
OS   Saccharomyces cerevisiae.
FH   Key             Location/Qualifiers
FT   promoter        5..523
FT                   /*tag= a
FT                   /note= "claimed fragment"
PN   WO9631611-A1.
PD   10-OCT-1996.
PF   05-APR-1996; F00523.
PR   07-APR-1995; FR-004171.
PA   (TRGE ) TRANSGENE SA.
PI   Achstetter T,  Nacken V;
DR   WPI; 96-465032/46.
PT   New yeast promoters from chromosome IX DNA fragments - useful for
PT   constructing vectors for expression of heterologous proteins e.g.
PT   cytokine(s), antibodies, etc, in yeast
PS   Claim 1; Page 31; 60pp; French.
CC   The sequences shown in T48041-3 represent novel fragments with
CC   transcriptional promoter activity which are derived from a yeast
CC   chromosome IX library.  The library was generated by isolating yeast
CC   chromosome IX separated by pulsed field gel electrophoresis.  The
CC   isolated DNA was partially digested with Sau3A and ligated into the
CC   plasmid pTG9852.  The fragments were ligated upstream of the phleomycin
CC   resistance gene ble.  The library was transformed into the yeast strain
CC   TGY74.3 by electroporation.  Fragments containing putative promoter
CC   sequences were isolated by growing the transformed yeast in the presence
CC   of phleomycin. In order to verify that the positives transformants grown
CC   were due to promoter activity and not due to reversion, the plasmids
CC   were isolated from the yeast cells and re-transformed into E.coli,
CC   followed by re-introduction into yeast strain TGY74.3.  Of 9600 original
CC   transformants, 20 positives with resistance to high concentrations of
```



```
CC   phleomycin (2 mg/ml) were isolated by the first round of screening.
CC   After the re-introduction into TGY74.3, 4 positive transformants were
CC   isolated which contained the 3 sequences defined above.  This sequence
CC   corresponds to the 3' end of the yeast COX4 gene. The 3' of this sequence
CC   is thought to contain a putative open reading frame of 25 amino acids.
CC   Vectors containing the promoter sequence can be used for expression of
CC   heterologous genes in yeast cells, esp. for prodn. of commercially or
CC   therapeutically useful polypeptides, e.g. cytokines, growth factors,
CC   receptors. anti-coagulants, enzymes or their inhibitors, antibodies,
CC   immunogenic polypeptides, structural or marker proteins.
SQ   Sequence   609 BP;    183 A;    110 C;    129 G;    187 T;

  Query Match           5.1%;  Score 15;  DB 29;  Length 609;
  Best Local Similarity 100.0%;  Pred. No. 5.76e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     21 atatttttatttaaa 35
          |||||||||||||||
Qy    108 ATATTTTTATTTAAA 122


RESULT   38
ID   Q03374 standard; DNA; 709 BP.
AC   Q03374;
DT   04-AUG-1990  (first entry)
DE   Partial sequence of gene that encodes the peptide core of human secretory
DE   proteoglycans (hSGP) depicting intron 2 and exon 3
KW   Diagnostic probe; immune effector cell monitoring;
KW   human secretoy granule proteoglycan peptide core protein.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   intron          1..38
FT                   /*tag= a
FT                   /note="intron 2"
FT   exon            39..709
FT                   /*tag= b
FT                   /note="exon 3"
FT   mat_peptide     39..288
FT                   /*tag= c
FT   misc_feature    704..709
FT                   /*tag= d
FT                   /note="polyadenylation site"
FT   misc_binding    91..144
FT                   /*tag= e
FT                   /note="serine-glycine rich glycosaminoglycan attachment
FT                   region"
FT   cds             40..288
FT                   /*tag= f
FT   misc_feature    1
FT                   /*tag= g
FT                   /note="unknown 6 kb fragment, see CC"
```

```
PN   WO9000606-A.
PD   25-JAN-1990.
PF   13-JUL-1989; U03051.
PR   13-JUL-1988; US-224035.
PA   (BRIG-) Brigham & Womens.
PI   Stevens RL; Weis JH; Nicodemus CF;
DR   WPI; 90-051710/07.
DR   P-PSDB; R05247.
PT   Human proteo-glycan peptide core protein -
PT   present in effector cells of the human immune system
PS   Disclosure; p; English.
CC   The locus of the gene encoding substantially pure human secretory
CC   granule proteoglycan peptide core protein (hSGP), is on chromosome 10 in
CC   HL-60 cells. HSGP can be synthesised in a biological system, esp. a
CC   cultured mammalian cell where it is translated from mRNA, transcribed
CC   from recombinant DNA. hSGP cDNA, a recombinant DNA molecule contg. DNA
CC   encoding it or a fragment of it, and an antibody capable of binding human
CC   secretory proteoglycan or a fragment of it, are useful as diagnostic
CC   probes for monitoring the activation and involvement of immune effector
CC   cells in health and disease, eg cancer, inflammation and allergic
CC   disorders. The 5' region of the sequence that encodes hSGP is contained
CC   in n91787 and n93373. This sequence is separated from n93373 by an
CC   unknown 6 kb fragment (see FT).
SQ   Sequence   709 BP;   221 A;   124 C;   124 G;   240 T;

  Query Match           5.1%;  Score 15;  DB 1;  Length 709;
  Best Local Similarity 100.0%;  Pred. No. 5.76e+01;
  Matches   15;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    610 aaattggcaggtaat 624
          |||||||||||||||
Qy     49 AAATTGGCAGGTAAT 63


RESULT   39
ID   Q31693 standard; DNA; 731 BP.
AC   Q31693;
DT   15-APR-1993  (first entry)
DE   RsaI restriction fragment from Dd ras genomic clone.
KW   RsaI; restriction fragment; Dd ras; SmaI; pGEM1; promoter; vector;
KW   pAV-CAT; pEDII-CS; CS; leader peptide; CAT; Discoidin I; promoter;
KW   chloramphenicol acetyl transferase; CAT; geneticin; G418; resistance;
KW   actin 15; terminator; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   promoter        182..366
FT                   /*tag= a
PN   WO9220806-A.
PD   26-NOV-1992.
PF   18-MAY-1992; E01112.
PR   17-MAY-1991; NL-000869.
```



```
PA   (RMFD-) RMF DICTAGENE SA.
PI   Fasel NJ, Reymond CD;
DR   WPI; 92-415784/50.
PT   Dictyostelium expression vector - used to produce Plasmodium
PT   circumsporozoite protein for malaria vaccines
PS   Disclosure; Fig 4; 20pp; English.
CC   The sequence given represents the RsaI restriction fragment from the
CC   genomic clone of Dd ras.  This sequence was inserted into the SmaI
CC   site of pGEM1 and sequenced.  Parts of this promoter sequence were
CC   placed in the vector pAV-CAT to generate a plasmid which resembled
CC   pEDII-CS (see also Q31692), with CS and leader peptides replaced by
CC   the chloramphenicol acetyl transferase gene (CAT), Discoidin I promoter
CC   by Dd ras promoter and the geneticin (G418) resistance gene under the
CC   control of the actin 15 promoter.  The terminator region was placed in
CC   the opposite orientation.
SQ   Sequence   731 BP;   317 A;   52 C;   42 G;   320 T;

  Query Match           5.1%;  Score 15;  DB 5;  Length 731;
  Best Local Similarity 100.0%;  Pred. No. 5.76e+01;
  Matches   15;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    568 aaataaaaaatatttt 582
          |||||||||||||||
Cp    119 AAATAAAAATATTTT 105


RESULT   40
ID   N81961 standard; cDNA; 749 BP.
AC   N81961;
DT   02-FEB-1991  (first entry)
DE   Sequence of rat reg cDNA
KW   Pancreatic islet B cell regeneration; diabetes; therapy; ss.
OS   Rat.
FH   Key          Location/Qualifiers
FT   cds                CDS   1..550
FT                   /tag=a
PN   EP-286114-A.
PD   12-DEC-1988.
PF   08-APR-1988; 105623.
PR   10-AUG-1987; JP-200514.
PA   (SHIO) Shionogi Seiyaku Kk.
PI   Okamoto H;;
DR   WPI; 88-287314/41.
DR   P-PSDB; P81513.
PT   Rat and human reg genes -
PT   used for producing proteins for regeneration of
PT   insulin-producing B cells of patients with diabetes
PS   Disclosure; fig 2; 12pp; English.
CC   The reg gene is specifically expressed in regenerating pancreatic islet
CC   B cells.  A gene hybridising to a probe corresponding to at least a part
CC   of the whole base sequence of rat reg gene or human reg gene is claimed.
```



```
CC   By mass producing the proteins encoded by the gene it may be possible
CC   to open a new dimension in the treatment of diabetes
SQ   Sequence   749 BP;   216 A;   180 C;   147 G;   206 T;

  Query Match           5.1%;  Score 15;  DB 2;  Length 749;
  Best Local Similarity 100.0%;  Pred. No. 5.76e+01;
  Matches   15;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    725 atttaaataaaaata 739
          |||||||||||||||
Cp    123 ATTTAAATAAAAATA 109


RESULT   41
ID   V06125 standard; DNA; 919 BP.
AC   V06125;
DT   05-MAY-1998  (first entry)
DE   Viral infection gene SEQ ID NO:45.
KW   Viral infection; tumour suppressor; cellular gene; rat; cancer;
KW   serum protein; inhibitor; malignant phenotype; HIV; influenza;
KW   hepatitis; retrovirus; immunodeficiency; ds.
OS   Rattus sp.
PN   WO9739119-A1.
PD   23-OCT-1997.
PF   11-APR-1997; U06067.
PR   15-APR-1996; US-015334.
PA   (UTVA-) UNIV VANDERBILT.
PI   Dubois RN, Organ EL, Rubin DH;
DR   WPI; 97-526456/48.
PT   Genes involved in viral infection and tumour suppression - used to
PT   develop products for reducing or preventing viral infection or for
PT   suppressing tumours
PS   Claim 1; Page 63; 101pp; English.
CC   The present sequence represents a viral infection gene. The present
CC   invention describes nucleic acid sequences isolated from rat. The
CC   sequences of the invention comprise 70 viral infection (VI) genes and
CC   8 tumour suppressor (TS) genes. Propagating cell cultures in the
CC   absence of the serum protein (SP) allows selective elimination of cells
CC   persistently infected with a virus from the cell culture. Inhibitors of
CC   the SP can be used for reducing or inhibiting a viral infection.
CC   Inhibitors of the TS gene products can be used to suppress a malignant
CC   phenotype (MP). The methods and inhibitors can be used with viruses
CC   such as HIV, influenza, hepatitis virus or animal retroviruses such as
CC   simian immunodeficiency virus, avian immunodeficiency virus, bovine
CC   immunodeficiency virus, feline immunodeficiency virus, equine infectious
CC   anemia virus, caprine arthritis encephalitis virus or visna virus.
CC   Because the identified genes are non-essential to cell survival, the
CC   treatment methods can be used in subjects without serious detrimental
CC   effects to the subjects.
SQ   Sequence   919 BP;   225 A;   153 C;   255 G;   207 T;
```



```
  Query Match           5.1%;  Score 15;  DB 38;  Length 919;
  Best Local Similarity 93.8%;  Pred. No. 5.76e+01;
  Matches   15;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;

Db    133 aaataaaaaanattttt 148
          ||||||||| ||||||
Cp    119 AAATAAAAATATTTTT 104


RESULT   42
ID   V24837 standard; DNA; 975 BP.
AC   V24837;
DT   16-JUN-1998  (first entry)
DE   H. pylori ORF hp3p10807orf6.
KW   Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW   identification; binding compound; bacteria; life cycle; activator;
KW   inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis; ds.
OS   Helicobacter pylori.
FH   Key             Location/Qualifiers
FT   CDS             1..975
FT                   /*tag= a
FT                   /note= "no stop codon given"
PN   WO9737044-A1.
PD   09-OCT-1997.
PF   27-MAR-1997; U05223.
PR   06-DEC-1996; US-761318.
PR   29-MAR-1996; US-625811.
PR   02-APR-1996; US-758731.
PR   25-OCT-1996; US-736905.
PR   28-OCT-1996; US-738859.
PA   (ASTR ) ASTRA AB.
PI   Alm RA, Smith D;
DR   WPI; 97-503122/46.
DR   P-PSDB; W55428.
PT   Helicobacter pylori nucleic acid sequences and encoded
PT   polypeptide(s) - useful in vaccines to treat or prevent H. pylori
PT   infection and for diagnosis of H. pylori infection
PS   Claims 5,6; Page 275; 1145pp; English.
CC   This sequence encodes a H. pylori protein of unspecified function.
CC   The protein may be used in a vaccine to prevent or treat H. pylori
CC   infection or to identify H. pylori polypeptide binding compounds,
CC   useful as potential H. pylori life cycle activators or inhibitors. The
CC   DNA and probes derived from it may be used for the identification of
CC   H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
CC   acid sequences complementary to the DNA act as antisense sequences and
CC   can be used to prevent the translation of H. pylori mRNA. Antibodies
CC   against the protein can be used in immunoassays to evaluate the abundance
CC   and distribution of H. pylori-specific antigens. The genomic sequence of
CC   H. pylori (ATCC 55679) was determined from overlapping contigs generated
CC   by mechanically shearing the bacterial DNA. The sequences were analysed
CC   for ORF of at least 180 nucleotides, and the predicted coding regions
```

```
CC   defined by computer evaluation. To identify likely H. pylori antigens for
CC   vaccine development, the amino acid sequences predicted from various ORF
CC   were analysed for significant homology to other known or exported
CC   membrane proteins. Having identified and determined the sequences of
CC   interest, particular regions can be isolated from H. pylori by PCR
CC   amplification for recombinant polypeptide production, e.g. in E. coli
CC   hosts.
SQ   Sequence   975 BP;    279 A;    174 C;    238 G;    284 T;

  Query Match          5.1%;  Score 15;  DB 40;  Length 975;
  Best Local Similarity 100.0%;  Pred. No. 5.76e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    464 tcaataaacttttg 478
          |||||||||||||||
Qy    267 TCAATAAACTTTTTG 281


RESULT   43
ID   T75679 standard; DNA; 1116 BP.
AC   T75679;
DT   05-FEB-1998  (first entry)
DE   Borrelia burgdorferi sensu stricto outer membrane protein oms28 gene.
KW   oms28; outer membrane porin protein; Oms28; spirochete;
KW   systemic illness; vaccine; Borrelia; Lyme disease; tick;
KW   B. burgdorferi sensu lato; ss.
OS   Borrelia burgdorferi sensu stricto.
FH   Key             Location/Qualifiers
FT   -35_signal      26..31
FT                   /*tag=  a
FT                   /note= "putative"
FT   -10_signal      54..59
FT                   /*tag=  b
FT                   /note= "putative"
FT   RBS             88..93
FT                   /*tag=  c
FT                   /note= "putative"
FT   CDS             102..875
FT                   /*tag=  d
FT   sig_peptide     102..173
FT                   /*tag=  e
FT   mat_peptide     174..872
FT                   /*tag=  f
FT                   /product= Oms28_outer-membrane_protein
FT   terminator      914..953
FT                   /*tag=  g
FT                   /note= "rho independent potential transcription
FT                        termination signal"
FT   unsure          300..302
FT                   /*tag=  h
FT                   /note= "encodes Thr"
```



```
FT   unsure          501..503
FT                   /*tag=  i
FT                   /note= "encodes Val"
FT   unsure          504..506
FT                   /*tag=  j
FT                   /note= "encodes Val"
FT   unsure          534..536
FT                   /*tag=  k
FT                   /note= "encodes Thr"
FT   modified_base   502
FT                   /*tag=  l
FT                   /mod_base= i
FT   modified_base   505
FT                   /*tag=  m
FT                   /mod_base= i
PN   WO9726273-A1.
PD   24-JUL-1997.
PF   21-JAN-1997; U01040.
PR   22-JAN-1996; US-010321.
PA   (SLOK ) SLOAN KETTERING INST CANCER RES.
PA   (REGC ) UNIV CALIFORNIA.
PI   Champion CI,  Kagan BL,  Lovett MA,  Miller JN;
PI   Mirzabekov TA, Shang ES,  Skare JT,  Tempst P;
DR   WPI; 97-385292/35.
DR   P-PSDB; W22093.
PT   Borrelia outer membrane proteins Oms28, Oms45 and Oms66 - used to
PT   induce immune response against Borrelia in an animal
PS   Claim 18; Fig 10; 132pp; English.
CC   The DNA sequence is of a gene, oms28, that encodes a Borrelia
CC   burgdorferi sensu stricto outer membrane porin protein, Oms28,
CC   which has a porin activity of 0.6 nS in 1 M KCL. Other Oms proteins
CC   are known, but Oms28 is the only one not expressed by other spirochetes,
CC   and appears to be specific to B. burgdorferi sensu stricto, an organism
CC   that initially causes a flu-like systemic illness that, if untreated,
CC   may develop into a disease characterised by arthritic, cardiac and
CC   neurological involvement. The Oms proteins can be used immunologically as
CC   vaccines for diseases caused by Borrelia, e.g. Lyme disease, a tick borne
CC   infection caused by B. burgdorferi sensu lato. The antibody against
CC   Oms28, or a nucleic acid probe for the coding sequence of Oms28, can be
CC   used to detect Borrelia in a sample.
SQ   Sequence   1116 BP;    425 A;    118 C;    200 G;    367 T;

  Query Match          5.1%;  Score 15;  DB 35;  Length 1116;
  Best Local Similarity 100.0%;  Pred. No. 5.76e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     19 atttttatttaaata 33
          |||||||||||||||
Qy    110 ATTTTTATTTAAATA 124
```



```
RESULT   44
ID   Q63613 standard; cDNA to mRNA; 1183 BP.
AC   Q63613;
DT   06-JAN-1995  (first entry)
DE   T-cell transmembrane protein 19.5 cDNA.
KW   transmembrane protein; T-lymphocytes; Lov gene; manipulation;
KW   T-cell development; altering tumourigenic phenotype;
KW   antibody production; identify; target drugs; ds.
OS   Mus musculus.
FH   Key             Location/Qualifiers
FT   misc_signal     52..58
FT                   /*tag=  a
FT                   /standard_name= Kozak consensus sequence
FT   polya_signal    1142..1147
FT                   /*tag=  b
FT   cds             55..876
FT                   /*tag=  c
FT                   /product= T-cell transmembrane protein
FT                   /note= "product of novel Lov gene (19.5 clone)"
PN   US5312733-A.
PD   17-MAY-1994.
PF   13-APR-1990; 509684.
PR   13-APR-1990; US-509684.
PR   11-APR-1991; US-686322.
PA   (RERE-) RES DEV FOUND.
PI   Macleod CL;
DR   WPI; 94-159119/19.
DR   P-PSDB; R53466.
PT   Expression vectors contg T-cell DNA sequences - coding for T-cell
PT   trans membrane proteins
PS   Claim 1; Fig 3A-3B; 56pp; English.
CC   Q63613 encodes a novel T-cell multiple transmembrane protein (R53466).
CC   The novel gene, 19.5, also referred to as Lov, is inducible in
CC   SL 12.4 cells (CD4 CD8 double negative phenotype, therefore
CC   resembling thymocytes at a relatively early stage of development.).
CC   The Lov gene has been mapped to murine Chromosome 16. Its gene
CC   product is developmentally regulated and plays a role in T cell
CC   development. The gene transcript appears to be highly expressed in
CC   ovaries and gut associated lymphoid tissue (GALT) as well as
CC   thymus. The T cell protein may be useful for manipulating
CC   regulation of T cell development, altering the tumourigenic
CC   phenotype and localising metastatic foci of tumours, and for
CC   producing antibodies which may be used to identify or target drugs
CC   or other agents to specific cell types expressing the proteins.
SQ   Sequence   1183 BP;    297 A;    251 C;    256 G;    379 T;

  Query Match          5.1%;  Score 15;  DB 11;  Length 1183;
  Best Local Similarity 100.0%;  Pred. No. 5.76e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1142 attaaaaatattttt 1156
```



```
            |||||||||||||||
Qy    101 ATTAAAAATATTTTT 115


RESULT   45
ID   Q14534 standard; cDNA; 1183 BP.
AC   Q14534;
DT   29-JAN-1992  (first entry)
DE   Lov gene (cDNA 19.5).
KW   Multiple membrane spanning protein; T cell; development; lymphoid;
KW   cellular; ovarian; expression; ss.
OS   Mus musculus.
FH   Key             Location/Qualifiers
FT   cds             55..873
FT                   /*tag=  a
FT   polya_signal    1142..1147
FT                   /*tag=  a
PN   WO9116430-A.
PD   31-OCT-1991.
PF   12-APR-1990; U02518.
PR   13-APR-1990; US-509684.
PA   (REDE-) RES DEV FOUND.
PI   Macleod CL;
DR   WPI; 91-339818/46.
DR   P-PSDB; R14644.
PT   New recombinant polypeptide comprising a T-cell protein - used to
PT   regulate T-cell development and tumorigenic phenotype and to
PT   block T-cell activation in auto:immune disease
PS   Disclosure; Fig 3; 98pp; English.
CC   The 19.5 gene is also referred to as Lov (Lymphoid and Ovarian
CC   Cellular expression).  It is inducible in SL 12.4 cells after co-
CC   cultivation on thymic epithelial monolayers.  The Lov gene has been
CC   mapped to murine chromosome 16.  The Lov gene product is
CC   developmentally regulated and plays a role in T cell development.
CC   The 19.5 cDNA encodes a multiple membrane spanning protein.
CC   See also Q14530-34.
SQ   Sequence   1183 BP;    297 A;    250 C;    255 G;    381 T;

  Query Match          5.1%;  Score 15;  DB 3;  Length 1183;
  Best Local Similarity 100.0%;  Pred. No. 5.76e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1142 attaaaaatattttt 1156
            |||||||||||||||
Qy    101 ATTAAAAATATTTTT 115


RESULT   46
ID   T79694 standard; DNA; 1198 BP.
AC   T79694;
DT   03-FEB-1998  (first entry)
```

```
DE   BRCA2 cancer susceptibility gene exon 25 and intron boundaries.
KW   BRCA2 cancer susceptibility gene; breast cancer; ovarian cancer;
KW   gene therapy; prostate cancer; colorectal cancer; ocular melanoma;
KW   leukaemia; human; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   intron          1..500
FT                   /*tag=  a
FT                   /number= 24
FT                   /note= "End of intron 24"
FT   exon            501..745
FT                   /*tag=  b
FT                   /number= 25
FT   intron          746..1198
FT                   /*tag=  c
FT                   /number= 25
FT                   /note= "Start of intron 25"
PN   GB2307477-A.
PD   28-MAY-1997.
PF   25-NOV-1996; 024453.
PR.  28-AUG-1996; GB-017961.
PR   23-NOV-1995; GB-023959.
PR   14-DEC-1995; GB-025555.
PA   (CANC-) CANCER RES CAMPAIGN TECHNOLOGY.
PA   (UYDU-) UNIV DUKE.
PI   Ashworth A,  Futreal PA,  Stratton MR,  Wooster RF;
DR   WPI; 97-261854/24.
DR   P-PSDB; W25063.
PT   Nucleic acid molecules comprising part or all of the BRCA2 cancer
PT   susceptibility gene - useful for diagnosis, prognosis or therapeutic
PT   treatment of cancer
PS   Claim 10; Fig 7; 124pp; English.
CC   The present sequence represents exon 25 and the intron boundaries of the
CC   BRCA2 cancer susceptibility gene. The nucleic acid molecule can be used
CC   to construct probes for screening cDNA or genomic libraries, sequencing
CC   positive clones obtained, and assembling the full length BRCA2
CC   sequence. The BRCA2 nucleic acid molecules and proteins are
CC   useful in a method of medical treatment, preferably gene therapy,
CC   especially for treating cancer, where the cancer is female or male
CC   breast cancer, ovarian, prostate or colorectal cancer, ocular melanoma
CC   or leukaemia. In particular antisense oligonucleotides capable of
CC   hybridising to the BRCA2 nucleic acid, pre-mRNA or mature mRNA are used
CC   so that the expression of the BRCA2 nucleic acid is reduced or
CC   prevented. The nucleic acid molecules are also useful in a method for
CC   diagnosing susceptibility or predisposition to cancer in a patient. The
CC   nucleic acid molecules are used to design probes or primers for PCR to
CC   determine or detect the presence of mutations in a sample of nucleic
CC   acid from a patient. The BRCA2 promoter region is useful for screening
CC   for substances which modulate the expression of nucleic acid under
CC   control of the promoter. Antibodies are used to determine the presence,
CC   amount or location in a cell of a BRCA2 polypeptide or its mutant forms.
```



```
CC   The polypeptides are used to screen for binding partners, these are
CC   useful to screen for substances which mimic the activity of BRCA2
CC   polypeptide, which can be used as cancer therapeutics.
SQ   Sequence   1198 BP;    372 A;    183 C;    227 G;    416 T;

  Query Match           5.1%;  Score 15;  DB 35;  Length 1198;
  Best Local Similarity 100.0%;  Pred. No. 5.76e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    726 aaatgaaaaaatactg 740
          ||||||||||||||| ..
Cp    194 AAATGAAAAATACTG 180


RESULT   47
ID   T61407 standard; cDNA; 1301 BP.
AC   T61407;
DT   29-OCT-1997  (first entry)
DE   MACH isoform alpha3 coding sequence.
KW   MACH; MORT-1 binding protein; mediator of receptor toxicity; cell death;
KW   antibody; FAS ligand receptor; FAS-R; death domain region; septic shock;
KW   tumour necrosis factor; tumour; HIV-infection; oligodendrocyte death;
KW   apoptosis/programmed cell death; p55-R; graft rejection; acute hepatitis;
KW   autoimmune disease; multiple sclerosis; AIDS-inhibited T-cell suicide;
KW   TNF; therapy; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             3..803
FT                   /*tag=  a
FT                   /product= MORT-1 binding protein (MACH) isoform alpha3
PN   WO9703998-A1.
PD   06-FEB-1997.
PF   14-JUN-1996; U10521.
PR   16-APR-1996; IL-117932.
PR   16-JUL-1995; IL-114615.
PR   17-AUG-1995; IL-114986.
PR   14-SEP-1995; IL-115319.
PR   27-SEP-1995; IL-116588.
PA   (WEIN/) WEINWURZEL H.
PA   (YEDA ) YEDA RES & DEV CO LTD.
PI   Boldin M,  Goltsev YV,  Goncharov T,  Wallach D;
DR   WPI; 97-132570/12.
DR   P-PSDB; W11896.
PT   New DNA encoding MACH protein that interacts with MORT-1 protein -
PT   to mediate intracellular effects of FAS or TNF receptors, partic.
PT   for regulating apoptosis in tumours, virus-infected cells etc.
PS   Example 3; Page 122; 163pp; English.
CC   T61396, and T61405-T61411 represent coding sequences for different
CC   isoforms of MACH. MACH is a binding protein for the mediator of receptor
CC   toxicity (MORT-1) protein. MORT-1 binds to the FAS ligand receptor
CC   (FAS-R) death domain region, and triggers part of the cell death
```



```
CC   signalling cascade in mammalian cells. Vectors containing these
CC   sequences, the encoded proteins, and antibodies (Ab) against them are
CC   used to modulate the effect of FAS-R ligand or TNF on cells that carry
CC   FAS-R or p55-R. This is specifically for treating tumours, HIV-infected
CC   cells or other diseased cells, by control of apoptosis/programmed cell
CC   death. The encoded proteins are mediators of the cell death pathway
CC   initiated by TNF and FAS-R binding, i.e. it mimics or enhances the effect
CC   of MORT-1 where increased cytotoxicity is required. To inhibit the effect
CC   of MORT-1, e.g. in cases of septic shock, graft rejection and acute
CC   hepatitis, sequences encoding antisense molecules or ribozymes, or Ab
CC   against the protein, are used. Compounds that inhibit MACH are
CC   potentially useful for controlling MACH activity e.g. in cases of
CC   autoimmune disease, oligodendrocyte death in multiple sclerosis or
CC   AIDS-inhibited T-cell suicide. The MACH proteins can also be used to
CC   isolate and characterise other proteins and receptors involved in
CC   signalling and for Ab production. The Ab can be used to purify the new
CC   proteins and for diagnosis of conditions involving abnormal function of
CC   FAS-R mediated cellular effects.
SQ   Sequence   1301 BP;    347 A;    297 C;    294 G;    363 T;

  Query Match           5.1%;  Score 15;  DB 33;  Length 1301;
  Best Local Similarity 100.0%;  Pred. No. 5.76e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    996 ttaaaaatattttta 1010
          ||||||||||||||| .
Qy    102 TTAAAAATATTTTTA 116


RESULT   48
ID   Q83838 standard; DNA; 1361 BP.
AC   Q83838;
DT   05-SEP-1995  (first entry)
DE   B. burgdorferi S1 gene.
KW   Lyme disease; borreliosis; immunization; vaccine; diagnostic; ds.
OS   Borrelia burgdorferi N40.
FH   Key             Location/Qualifiers
FT   cds             109..1361
FT                   /*tag=  a
PN   WO9504145-A.
PD   09-FEB-1995.
PF   26-JUL-1994; U08529.
PR   30-JUL-1993; US-099757.
PR   08-SEP-1993; US-118469.
PA   (UYYA ) UNIV YALE.
PI   Barthold SW,  Fikrig E,  Flavell RA,  Kantor FS,  Lam TT;
DR   WPI; 95-082229/11.
DR   P-PSDB; R69630.
PT   New DNA encoding new Borrelia burgdorferi polypeptide(s) - and
PT   related vectors, transformed cells, fusion or multimeric proteins
PT   and antibodies, for treatment, prevention and diagnosis of Lyme
```



```
PT   disease
PS   Claim 1; Page 80-82; 114pp; English.
CC   A cDNA library in lambda ZAPII was constructed from B. burgdorferi
CC   N40 and screened with rabbit anti-serum against N40.  6 Positive
CC   clones, not contg. OspA, OspB or flagellin genes, were selected.
CC   1 of these contained the S1 gene (Q83838) that encoded a 48.9 kDa
CC   protein (R69630). The protein may be used in the development of
CC   multimeric vaccines.
SQ   Sequence   1361 BP;    525 A;    162 C;    246 G;    428 T;

  Query Match           5.1%;  Score 15;  DB 14;  Length 1361;
  Best Local Similarity 100.0%;  Pred. No. 5.76e+01;
  Matches   15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     65 atatttttatttaaa 79
          ||||||||||||||| .
Qy    108 ATATTTTTATTTAAA 122


RESULT   49
ID   T61411 standard; cDNA; 1443 BP.
AC   T61411;
DT   29-OCT-1997  (first entry)
DE   MACH isoform beta5 coding sequence.
KW   MACH; MORT-1 binding protein; mediator of receptor toxicity; cell death;
KW   antibody; FAS ligand receptor; FAS-R; death domain region; septic shock;
KW   tumour necrosis factor; tumour; HIV-infection; oligodendrocyte death;
KW   apoptosis/programmed cell death; p55-R; graft rejection; acute hepatitis;
KW   autoimmune disease; multiple sclerosis; AIDS-inhibited T-cell suicide;
KW   TNF; therapy; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             3..278
FT                   /*tag=  a
FT                   /product= MORT-1 binding protein (MACH) isoform beta5
PN   WO9703998-A1.
PD   06-FEB-1997.
PF   14-JUN-1996; U10521.
PR   16-APR-1996; IL-117932.
PR   16-JUL-1995; IL-114615.
PR   17-AUG-1995; IL-114986.
PR   14-SEP-1995; IL-115319.
PR   27-SEP-1995; IL-116588.
PA   (WEIN/) WEINWURZEL H.
PA   (YEDA ) YEDA RES & DEV CO LTD.
PI   Boldin M,  Goltsev YV,  Goncharov T,  Wallach D;
DR   WPI; 97-132570/12.
DR   P-PSDB; W11899.
PT   New DNA encoding MACH protein that interacts with MORT-1 protein -
PT   to mediate intracellular effects of FAS or TNF receptors, partic.
PT   for regulating apoptosis in tumours, virus-infected cells etc.
```

```
PS   Example 3; Page 130; 163pp; English.
CC   T61396, and T61405-T61411 represent coding sequences for different
CC   isoforms of MACH. MACH is a binding protein for the mediator of receptor
CC   toxicity (MORT-1) protein. MORT-1 binds to the FAS ligand receptor
CC   (FAS-R) death domain region, and triggers part of the cell death
CC   signalling cascade in mammalian cells. Vectors containing these
CC   sequences, the encoded proteins, and antibodies (Ab) against them are
CC   used to modulate the effect of FAS-R ligand or TNF on cells that carry
CC   FAS-R or p55-R. This is specifically for treating tumours, HIV-infected
CC   cells or other diseased cells, by control of apoptosis/programmed cell
CC   death. The encoded proteins are mediators of the cell death pathway
CC   initiated by TNF and FAS-R binding, i.e. it mimics or enhances the effect
CC   of MORT-1 where increased cytotoxicity is required. To inhibit the effect
CC   of MORT-1, e.g. in cases of septic shock, graft rejection and acute
CC   hepatitis, sequences encoding antisense molecules or ribozymes, or Ab
CC   against the protein, are used. Compounds that inhibit MACH are
CC   potentially useful for controlling MACH activity e.g. in cases of
CC   autoimmune disease, oligodendrocyte death in multiple sclerosis or
CC   AIDS-inhibited T-cell suicide. The MACH proteins can also be used to
CC   isolate and characterise other proteins and receptors involved in
CC   signalling and for Ab production. The Ab can be used to purify the new
CC   proteins and for diagnosis of conditions involving abnormal function of
CC   FAS-R mediated cellular effects.
SQ   Sequence   1443 BP;    400 A;    326 C;    327 G;    390 T;

  Query Match            5.1%;  Score 15;  DB 33;  Length 1443;
  Best Local Similarity 100.0%;  Pred. No. 5.76e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1138 ttaaaaatattttta 1152
          ||||||||||||||| 
Qy    102 TTAAAAATATTTTTA 116


RESULT   50
ID   T88774 standard; cDNA; 1626 BP.
AC   T88774;
DT   22-FEB-1998  (first entry)
DE   Human receptor protein 2F1 cDNA clone hu2F1.
KW   2F1; receptor; prostaglandin; inflammation; arthritis; allergy;
KW   asthma; inflammatory bowel disease; fever; diabetes;
KW   dysmenorrhoea; therapy; human; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             1..1626
FT                   /*tag=  a
FT   sig_peptide     1..57
FT                   /*tag=  b
FT   mat_peptide     58..1623
FT                   /*tag=  c
FT   variation       949..951
```



```
FT                   /*tag=  d
FT                   /note= "the codon for Ala-296 is polymorphic, being
FT                          present in PBL clones and 2 KB clones, and
FT                          absent in 2 other KB clones"
PN   WO9731010-A1.
PD   28-AUG-1997.
PF   06-FEB-1997; U01697.
PR   21-FEB-1996; US-604333.
PA   (IMMV ) IMMUNEX CORP.
PI   Parnet P,  Sims JE;
DR   WPI; 97-435081/40.
DR   P-PSDB; W31907.
PT   DNA encoding 2F1 receptors that inhibit prostaglandin synthesis -
PT   for treatment of inflammatory diseases, e.g. arthritis, allergy,
PT   asthma, etc.
PS   Claim 3; Page 31-33; 49pp; English.
CC   cDNA clone hu2F1 encodes a novel human receptor protein designated
CC   2F1 (see W31907) that is a member of the interleukin-1 receptor
CC   (IL-1R) family.  It was obtained by subjecting a human artificial
CC   yeast chromosome, CO2133, containing DNA from chromosome 2q12, to
CC   PCR using primers based on IL-1R sequences.  The human cDNA was
CC   used to isolate the corresponding murine gene (see T88775).  Some
CC   human clones lacked the codon for Ala-298.  The isolated cDNA can
CC   be used for production of recombinant 2F1, especially soluble 2F1,
CC   in transformed host cells for use in claimed methods of inhibiting
CC   prostaglandin synthesis in a mammal, and of reducing inflammation
CC   in a mammal (e.g. for treating arthritis, allergy, asthma,
CC   inflammatory bowel disease, fever, diabetes, dysmenorrhoea, etc.).
CC   Since 2F1 can inhibit NF-kappaB activity, it may also be used to
CC   reduce expression of human immunodeficiency virus in infected cells
CC   (not claimed).  Fragments of DNA are useful as probes to isolate
CC   related sequences, in hybridisation assays and as sense or
CC   antisense molecules to block expression of 2F1 protein.
SQ   Sequence   1626 BP;    544 A;    276 C;    350 G;    456 T;

  Query Match            5.1%;  Score 15;  DB 35;  Length 1626;
  Best Local Similarity 100.0%;  Pred. No. 5.76e+01;
  Matches    15;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    292 ttccaaatgaaaaat 306
          ||||||||||||||| 
Cp    198 TTCCAAATGAAAAAT 184


Search completed: Tue Dec 22 19:48:12 1998
Job time : 169 secs.
```

```
**********************************************************************

 MMMM    MMMM  PPPPPPP   SSSSSS  EEEEEEE  AAAAA   RRRRRR   CCCCC  HH   HH
 MM MM  MM MM  PP   PP  SS       EE      AA   AA  RR   RR CC    C HH   HH
 MM  MMMM  MM  PPPPPPP   SSSSS   EEEEE   AAAAAAA  RRRRRR  CC      HHHHHHH
 MM   MM   MM  PP           SS   EE      AA   AA  RR  RR  CC    C HH   HH
 MM   MM   MM  PP      SSSSSS    EEEEEEE AA   AA  RR   RR  CCCCC  HH   HH

                                                                    (TM)
**********************************************************************

        Release 3.1A John F. Collins, Biocomputing Research Unit.
        Copyright (c) 1993-1998  University of Edinburgh, U.K.
                Distribution rights by Oxford Molecular Ltd

MPsrch_nn   n.a. - n.a. database search, using Smith-Waterman algorithm

Run on:             Tue Dec 22 19:23:39 1998;  MasPar time 457.99 Seconds
                                               1424.324 Million cell updates/sec
Tabular output not generated.

Title:              >US-08-530-112A-563
Description:        (1-296) from US08530112A.seq
Perfect Score:      296
N.A. Sequence:           1 GATCTGCCATAAGAAAATCT.........TTTTGGTTGTTAAACTGAAA 296
   Comp:                   CTAGACGGTATTCTTTTAGA.........AAAACCAACAATTTGACTTT

Scoring table:      TABLE jmetric
                    Gap 60

Nmatch   STD :      Dbase 0;  Query 0

Searched:           567134 seqs, 1101898692 bases x 2

Post-processing:    Minimum Match 0%
                    Listing first 1000 summaries

Database:           embl55
                      1:em_ba 2:em_fun 3:em_htg 4:em_huml 5:em_hum2 6:em_in
                      7:em_om 8:em_or 9:em_ov 10:em_pat 11:em_ph 12:em_pl
                      13:em_ro 14:em_vi
Database:           genbank107
                      15:gb_bal 16:gb_ba2 17:gb_htg 18:gb_in 19:gb_om 20:gb_ov
                      21:gb_pat 22:gb_ph 23:gb_pll 24:gb_pl2 25:gb_prl
                      26:gb_pr2 27:gb_pr3 28:gb_ro 29:gb_st 30:gb_sts 31:gb_sy
                      32:gb_un 33:gb_vi

Statistics:         Mean 8.270;  Variance 3.358;  scale 2.463
```


length of match

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
c	1	22	7.4	153617	17	HS117O3	Human DNA sequence ***	5.37e-02
	2	21	7.1	121434	17	AC005116	*** SEQUENCING IN PROG	2.62e-01
c	3	20	6.8	37172	23	SPAC22E12	S.pombe chromosome I c	1.23e+00
	4	20	6.8	81729	18	L49405	Drosophila melanogaste	1.23e+00
c	5	20	6.8	85579	26	AC004744	Homo sapiens BAC clone	1.23e+00
	6	20	6.8	116315	26	AC004130	Homo sapiens BAC clone	1.23e+00
	7	20	6.8	125590	26	AC002382	Human BAC clone RG022J	1.23e+00
c	8	20	6.8	170952	26	HS501N12	Homo sapiens DNA seque	1.23e+00
c	9	20	6.8	202233	26	AC003664	Homo sapiens chromosom	1.23e+00
	10	19	6.4	273	30	G22210	human STS WI-16602.	5.52e+00
	11	19	6.4	582	18	PARMTDI1B	paramecium species 1,1	5.52e+00
	12	19	6.4	2313	28	MMPSCO	M.musculus mRNA for pr	5.52e+00
	13	19	6.4	3225	18	DRO37DC97Z	Drosophila melanogaste	5.52e+00
c	14	19	6.4	3566	18	AF028834	Drosophila ananassae v	5.52e+00
c	15	19	6.4	3626	18	AF028835	Drosophila pallidosa v	5.52e+00
	16	19	6.4	4152	28	MUSGRIPGHS	Mouse glucocortoid-reg	5.52e+00
	17	19	6.4	11622	15	U32740	Haemophilus influenzae	5.52e+00
c	18	19	6.4	33268	18	CELF08D12	Caenorhabditis elegans	5.52e+00
	19	19	6.4	33802	18	CER07H5	Caenorhabditis elegans	5.52e+00
	20	19	6.4	37000	26	AC003987	Human PAC clone DJ0866	5.52e+00
	21	19	6.4	42361	18	CEC25D7	Caenorhabditis elegans	5.52e+00
	22	19	6.4	48158	18	AC002516	Drosophila melanogaste	5.52e+00
c	23	19	6.4	99512	26	AC004606	Homo sapiens chromosom	5.52e+00
	24	19	6.4	112984	26	HS422H11	Homo sapiens DNA seque	5.52e+00
	25	19	6.4	118021	26	HS516C23	Human DNA sequence ***	5.52e+00
	26	19	6.4	165197	26	HSAC002070	Human BAC clone 7E17 f	5.52e+00
c	27	19	6.4	167101	26	AC002094	Genomic sequence from	5.52e+00
	28	19	6.4	170342	17	AC003668	*** SEQUENCING IN PROG	5.52e+00
c	29	19	6.4	185364	17	AC005477	*** SEQUENCING IN PROG	5.52e+00
	30	19	6.4	190943	17	AC005037	*** SEQUENCING IN PROG	5.52e+00
	31	19	6.4	302988	17	CEY46G5	Caenorhabditis elegans	5.52e+00
c	32	18	6.1	250	30	G15468	human STS SHGC-16896.	2.36e+01
c	33	18	6.1	351	19	BTMSINRBP	B.taurus microsatellit	2.36e+01
	34	18	6.1	405	20	NAJRG16S	Naja naja naja 16S rib	2.36e+01
	35	18	6.1	412	20	MTNHRNA16	N.naja mitochondrial g	2.36e+01
	36	18	6.1	422	30	G21159	human STS WI-14062.	2.36e+01
c	37	18	6.1	504	23	PTAJ5292	Populus tremula x Popu	2.36e+01
c	38	18	6.1	504	30	G24158	human STS WI-13539.	2.36e+01
	39	18	6.1	905	18	TETHIS22MA	T.thermophila macronuc	2.36e+01
	40	18	6.1	905	18	TTH2B2	Tetrahymena thermophil	2.36e+01
c	41	18	6.1	1052	25	HSSCPXT16	Human sterol carrier p	2.36e+01
c	42	18	6.1	1219	25	HUMSTEAA	Human sterol carrier p	2.36e+01

c	43	18	6.1	1229	25	S52450	sterol carrier protein	2.36e+01
c	44	18	6.1	1260	15	SLU73444	Staphylococcus lugdune	2.36e+01
	45	18	6.1	1530	26	AF070566	Homo sapiens clones 24	2.36e+01
c	46	18	6.1	1678	26	HSUB5994	Human clone IMAGE:3000	2.36e+01
c	47	18	6.1	1892	25	HSUB5997	Human unknown protein	2.36e+01
	48	18	6.1	2424	18	DMU77310	Drosophila melanogaste	2.36e+01
c	49	18	6.1	2572	25	HUMSCP2A	Human sterol carrier p	2.36e+01
c	50	18	6.1	2613	26	HSL81395	Homo sapiens (subclone	2.36e+01
c	51	18	6.1	2648	26	MFU91582	Macaca fascicularis UD	2.36e+01
c	52	18	6.1	2661	19	AF051897	Oryctolagus cuniculus	2.36e+01
c	53	18	6.1	2820	23	YSCMTATPSA	Yeast mitochondrial ox	2.36e+01
	54	18	6.1	3060	27	HSU33317	Human defensin 6 (HD-6	2.36e+01
c	55	18	6.1	3141	26	HSL81826	Homo sapiens (subclone	2.36e+01
c	56	18	6.1	3735	28	MUSMHCIIAB	Mus musculus Mhc class	2.36e+01
	57	18	6.1	4032	20	XLCETS1A	Xenopus laevis c-ets-1	2.36e+01
	58	18	6.1	4192	23	AF057708	Populus balsamifera su	2.36e+01
c	59	18	6.1	4601	18	DMU11584	Drosophila melanogaste	2.36e+01
	60	18	6.1	8233	28	MMMB414A3	Mus musculus X inactiv	2.36e+01
c	61	18	6.1	10000	28	MMMH04	Mouse gene coding for	2.36e+01
c	62	18	6.1	10016	15	AE001167	Borrelia burgdorferi (	2.36e+01
c	63	18	6.1	12573	18	PFSC03006	Plasmodium falciparum	2.36e+01
	64	18	6.1	14607	33	VVA10L	Variola virus (Garcia-	2.36e+01
c	65	18	6.1	19517	18	DMU37541	Drosophila melanogaste	2.36e+01
c	66	18	6.1	21153	23	YSCMTCG13	Saccharomyces cerevisi	2.36e+01
c	67	18	6.1	25547	18	CELZK994	Caenorhabditis elegans	2.36e+01
c	68	18	6.1	27451	18	CEF53F8	Caenorhabditis elegans	2.36e+01
c	69	18	6.1	29552	18	CELF45F2	Caenorhabditis elegans	2.36e+01
	70	18	6.1	30969	18	CELT13B5	Caenorhabditis elegans	2.36e+01
c	71	18	6.1	33000	18	CER11H3	Caenorhabditis elegans	2.36e+01
	72	18	6.1	37108	18	CET20F10	Caenorhabditis elegans	2.36e+01
c	73	18	6.1	39842	17	HS329F2	Human DNA sequence ***	2.36e+01
	74	18	6.1	40380	18	CEF53H4	Caenorhabditis elegans	2.36e+01
c	75	18	6.1	41591	18	ACU12386	Acanthamoeba castellan	2.36e+01
c	76	18	6.1	42300	18	CELC16C8	Caenorhabditis elegans	2.36e+01
	77	18	6.1	43795	26	AC000022	Genomic sequence from	2.36e+01
	78	18	6.1	44422	18	L32205	Drosophila melanogaste	2.36e+01
c	79	18	6.1	60791	26	AC004749	Homo sapiens chromosom	2.36e+01
	80	18	6.1	80514	26	AC004454	Homo sapiens PAC clone	2.36e+01
c	81	18	6.1	84246	18	AC002512	Drosophila melanogaste	2.36e+01
c	82	18	6.1	86188	18	AC004318	Drosophila melanogaste	2.36e+01
c	83	18	6.1	86686	24	ATAC004786	Arabidopsis thaliana c	2.36e+01
c	84	18	6.1	86884	18	AC004366	Drosophila melanogaste	2.36e+01
	85	18	6.1	87841	23	AB005247	Arabidopsis thaliana g	2.36e+01
c	86	18	6.1	87841	23	AB005247	Arabidopsis thaliana g	2.36e+01
c	87	18	6.1	94038	24	ATAC004401	Arabidopsis thaliana c	2.36e+01
c	88	18	6.1	94459	28	MMTSXDNA	M.musculus 94kb genomi	2.36e+01
c	89	18	6.1	97006	24	ATF16G20	Arabidopsis thaliana D	2.36e+01
c	90	18	6.1	100997	26	HS262D12	Homo sapiens DNA seque	2.36e+01
	91	18	6.1	104913	26	AC004242	Homo sapiens 12q13 PAC	2.36e+01
c	92	18	6.1	109160	24	AC002328	Genomic sequence for A	2.36e+01
	93	18	6.1	112575	17	HS715N11	Human DNA sequence ***	2.36e+01

	94	18	6.1	128379	17	AC004889	*** SEQUENCING IN PROG	2.36e+01
c	95	18	6.1	129577	26	AC004659	Homo sapiens chromosom	2.36e+01
c	96	18	6.1	135599	23	CPU30821	Cyanophora paradoxa cy	2.36e+01
	97	18	6.1	136150	27	AC002081	Homo sapiens BAC clone	2.36e+01
c	98	18	6.1	136755	26	AC002529	Human BAC clone RG161N	2.36e+01
c	99	18	6.1	138187	28	MMMH461	Mus musculus Major His	2.36e+01
c	100	18	6.1	139259	18	CEY51A2D	Caenorhabditis elegans	2.36e+01
	101	18	6.1	148400	26	AC004768	Homo sapiens chromosom	2.36e+01
c	102	18	6.1	149322	17	HS1119D9	Human DNA sequence ***	2.36e+01
c	103	18	6.1	167575	17	HS531H16	Human DNA sequence ***	2.36e+01
	104	18	6.1	170101	33	ASU18466	African swine fever vi	2.36e+01
	105	18	6.1	172106	17	CEY67D11	Caenorhabditis elegans	2.36e+01
	106	18	6.1	175825	26	HS106H8	Human DNA sequence fro	2.36e+01
	107	18	6.1	177923	33	U94848	Vaccinia virus strain	2.36e+01
	108	18	6.1	185578	33	VVCGAA	Variola virus DNA comp	2.36e+01
	109	18	6.1	186103	33	VARCG	Variola major virus (s	2.36e+01
c	110	18	6.1	189694	17	AC005234	*** SEQUENCING IN PROG	2.36e+01
	111	18	6.1	191737	33	VACCG	Vaccinia virus, comple	2.36e+01
c	112	18	6.1	210216	17	AC004888	*** SEQUENCING IN PROG	2.36e+01
c	113	18	6.1	235922	17	CEY7A5	Caenorhabditis elegans	2.36e+01
c	114	18	6.1	251124	26	HUAE000660	Homo sapiens T-cell re	2.36e+01
	115	18	6.1	286715	17	CEY45F3	Caenorhabditis elegans	2.36e+01
c	116	18	6.1	328808	17	CEY44A6	Caenorhabditis elegans	2.36e+01
	117	18	6.1	338234	18	DMU31961	Drosophila melanogaste	2.36e+01
c	118	17	5.7	195	28	MUSSPDOT12	Mus caroli Sp100 gene,	9.52e+01
	119	17	5.7	255	30	G38578	SHGC-64311 Human Homo	9.52e+01
c	120	17	5.7	428	18	AF003512	Chelonus sp. 16S ribos	9.52e+01
c	121	17	5.7	480	18	DEUMTRGDA	Physodeutera fairmaire	9.52e+01
	122	17	5.7	551	23	SPTRX2	Schizosaccharomyces po	9.52e+01
c	123	17	5.7	624	18	SCMCBP	Schistosoma mansoni ca	9.52e+01
c	124	17	5.7	855	18	AF027736	Nephila clavipes minor	9.52e+01
c	125	17	5.7	1076	15	MMTRNA2	Mycoplasma mycoides tr	9.52e+01
	126	17	5.7	1109	18	SMU30261	Schistosoma mansoni G	9.52e+01
	127	17	5.7	1358	20	QULMF1A	Coturnix coturnix japo	9.52e+01
c	128	17	5.7	1380	33	AMUB0056	Amsacta moorei entomop	9.52e+01
	129	17	5.7	1443	33	NDVPP	Newcastle disease viru	9.52e+01
c	130	17	5.7	1462	23	CHCEPETD	Chlamydomonas eugameto	9.52e+01
	131	17	5.7	1711	23	SOYHSP176	Soybean (Glycine max)	9.52e+01
	132	17	5.7	1836	18	PBDNAGARP	P.berghei glutamine al	9.52e+01
c	133	17	5.7	1855	23	OSU66041	Oryza sativa ADP-gluco	9.52e+01
	134	17	5.7	1998	15	CLOENGB	Clostridium cellulovor	9.52e+01
c	135	17	5.7	2047	24	S77461	CYP52A3 (CYP52A3-b)=n-	9.52e+01
c	136	17	5.7	2047	23	YSACY52A3	Candida maltosa CYP52A	9.52e+01
	137	17	5.7	2053	23	CSXCPNDH	Celsia arturus chlorop	9.52e+01
c	138	17	5.7	2068	28	L1BG	Mus musculus beige gen	9.52e+01
	139	17	5.7	2111	23	HHU62180	Hemiboea henryi putati	9.52e+01
	140	17	5.7	2111	23	LPU62182	Lysionotus pauciflorus	9.52e+01
	141	17	5.7	2111	23	OPU62183	Opithandra primuloides	9.52e+01
	142	17	5.7	2188	28	AB001576	Rattus rattus mRNA for	9.52e+01
c	143	17	5.7	2188	28	AB001576	Rattus rattus mRNA for	9.52e+01
c	144	17	5.7	2323	24	YSRPER3G	Pichia pastoris PER3 g	9.52e+01

c	145	17	5.7	2340	33	NDVNDVF	Newcastle disease viru 9.52e+01
	146	17	5.7	2512	20	XLXFIA	X.laevis mRNA XE1-b fo 9.52e+01
	147	17	5.7	2645	15	AB000510	Phytoplasma sp. plasmi 9.52e+01
c	148	17	5.7	2707	28	MMU18773	Mus musculus GPI-ancho 9.52e+01
	149	17	5.7	2947	28	D89995	Rattus rattus mRNA for 9.52e+01
c	150	17	5.7	2947	28	D89995	Rattus rattus mRNA for 9.52e+01
	151	17	5.7	3020	26	AB005060	Homo sapiens mRNA for 9.52e+01
c	152	17	5.7	3020	26	AB005060	Homo sapiens mRNA for 9.52e+01
	153	17	5.7	3077	28	D89996	Rattus rattus mRNA for 9.52e+01
c	154	17	5.7	3077	28	D89996	Rattus rattus mRNA for 9.52e+01
c	155	17	5.7	3108	28	AF011424	Mus musculus putative 9.52e+01
	156	17	5.7	3333	18	TBRAB	T.brucei rab7 gene and 9.52e+01
c	157	17	5.7	3689	28	AF011425	Mus musculus putative 9.52e+01
c	158	17	5.7	3844	18	TFU66073	Tritrichomonas foetus 9.52e+01
c	159	17	5.7	3859	23	RCHLEU1	Rhizopus niveus leul D 9.52e+01
c	160	17	5.7	3917	18	DMCOLLIER	D.melanogaster mRNA fo 9.52e+01
c	161	17	5.7	4021	28	MMU27838	Mus musculus glycosyl- 9.52e+01
	162	17	5.7	4126	28	MUSIGKWNP	Mouse germline Ig kapp 9.52e+01
	163	17	5.7	4844	23	ATRADAIE	Arabidopsis thaliana R 9.52e+01
	164	17	5.7	4980	18	PFSC03073	Plasmodium falciparum 9.52e+01
	165	17	5.7	4984	23	YSPPMA1A	S.pombe H+-ATPase gene 9.52e+01
c	166	17	5.7	5371	28	MMU91967	Mus musculus platelet 9.52e+01
c	167	17	5.7	5565	20	GGCOLVIA	Chicken colVIA2 gene f 9.52e+01
c	168	17	5.7	6018	25	S44029	red visual pigment (5' 9.52e+01
	169	17	5.7	6325	18	CEF52D4	Caenorhabditis elegans 9.52e+01
c	170	17	5.7	9118	15	BHVMP7G	B.hermsii vmp7 gene en 9.52e+01
	171	17	5.7	9284	18	CEK11B4	Caenorhabditis elegans 9.52e+01
c	172	17	5.7	9457	15	BORVMP21G	B.hermsii vmp21 gene e 9.52e+01
	173	17	5.7	10025	15	U67607	Methanococcus jannasch 9.52e+01
c	174	17	5.7	10109	15	MPAE000048	Mycoplasma pneumoniae 9.52e+01
	175	17	5.7	10328	15	MPAE000013	Mycoplasma pneumoniae 9.52e+01
c	176	17	5.7	10535	18	PFSC03008	Plasmodium falciparum 9.52e+01
	177	17	5.7	13133	23	ALRIBP	A.longa plastid genes 9.52e+01
	178	17	5.7	13133	23	ALRBCL	A.longa rbcL, rpl5, rp 9.52e+01
	179	17	5.7	13928	15	U67510	Methanococcus jannasch 9.52e+01
c	180	17	5.7	14345	15	U39679	Mycoplasma genitalium 9.52e+01
c	181	17	5.7	15421	18	PPCOMPIRA	P.falciparum complete 9.52e+01
c	182	17	5.7	16571	15	U67532	Methanococcus jannasch 9.52e+01
	183	17	5.7	16840	20	XMU17896	Xiphophorus maculatus 9.52e+01
c	184	17	5.7	18036	15	U67579	Methanococcus jannasch 9.52e+01
c	185	17	5.7	19639	23	OLU72723	Oryza longistaminata r 9.52e+01
c	186	17	5.7	20217	15	MPAE000047	Mycoplasma pneumoniae 9.52e+01
	187	17	5.7	20303	18	CEF58H10	Caenorhabditis elegans 9.52e+01
	188	17	5.7	20445	24	AF034411	Lycopersicon esculentu 9.52e+01
c	189	17	5.7	21480	25	HSQCBB	Human DNA sequence fro 9.52e+01
	190	17	5.7	24711	18	CELE03D2	Caenorhabditis elegans 9.52e+01
	191	17	5.7	25195	18	CELC24A3	Caenorhabditis elegans 9.52e+01
	192	17	5.7	25400	18	CEF59A2	Caenorhabditis elegans 9.52e+01
	193	17	5.7	26542	18	CEM18	Caenorhabditis elegans 9.52e+01
	194	17	5.7	27019	18	CEF47C8	Caenorhabditis elegans 9.52e+01
	195	17	5.7	29093	18	CELC49D10	Caenorhabditis elegans 9.52e+01

c	196	17	5.7	29186	18	CELT25F10	Caenorhabditis elegans 9.52e+01
c	197	17	5.7	29214	18	CELC50F2	Caenorhabditis elegans 9.52e+01
	198	17	5.7	29766	15	AE000786	Borrelia burgdorferi p 9.52e+01
c	199	17	5.7	29792	23	SPAC7D4	S.pombe chromosome I c 9.52e+01
	200	17	5.7	29795	18	CELF31A3	Caenorhabditis elegans 9.52e+01
c	201	17	5.7	30016	17	CEH26P12	Caenorhabditis elegans 9.52e+01
c	202	17	5.7	30275	18	CEEK792	Caenorhabditis elegans 9.52e+01
c	203	17	5.7	31252	25	HSJ256K2B	Human DNA sequence fro 9.52e+01
c	204	17	5.7	33367	26	AC004400	Homo sapiens chromosom 9.52e+01
c	205	17	5.7	33399	18	CEC09H6	Caenorhabditis elegans 9.52e+01
c	206	17	5.7	35557	18	CEY106G6D	Caenorhabditis elegans 9.52e+01
c	207	17	5.7	36306	18	CEM142	Caenorhabditis elegans 9.52e+01
	208	17	5.7	36849	15	AE000788	Borrelia burgdorferi p 9.52e+01
	209	17	5.7	36922	18	PFSC03026	Plasmodium falciparum 9.52e+01
	210	17	5.7	40433	26	HSU22376	Human (c-myb) gene, co 9.52e+01
	211	17	5.7	40897	18	CELC39D10	Caenorhabditis elegans 9.52e+01
c	212	17	5.7	41039	18	CELC52B11	Caenorhabditis elegans 9.52e+01
c	213	17	5.7	41087	18	CELF09G2	Caenorhabditis elegans 9.52e+01
c	214	17	5.7	41322	18	CEC50B6	Caenorhabditis elegans 9.52e+01
c	215	17	5.7	41367	25	HSD163D10	Human DNA sequence fro 9.52e+01
c	216	17	5.7	42245	26	AC004476	Homo sapiens chromosom 9.52e+01
	217	17	5.7	42398	26	D84401	Homo sapiens genomic D 9.52e+01
	218	17	5.7	42472	18	CEC26C6	Caenorhabditis elegans 9.52e+01
c	219	17	5.7	42879	25	HSAC000123	Human Cosmid g0771a106 9.52e+01
c	220	17	5.7	44175	26	AC000049	Homo sapiens Chromosom 9.52e+01
c	221	17	5.7	44676	25	HSU92G6	Human DNA sequence fro 9.52e+01
	222	17	5.7	47334	18	AC003054	Drosophila melanogaste 9.52e+01
c	223	17	5.7	49457	17	CEC47A4	Caenorhabditis elegans 9.52e+01
	224	17	5.7	53561	15	AE000790	Borrelia burgdorferi p 9.52e+01
c	225	17	5.7	57688	26	AP000048	Homo sapiens genomic D 9.52e+01
c	226	17	5.7	58645	27	AC002068	Homo sapiens BAC clone 9.52e+01
	227	17	5.7	62262	26	HSTCRBK17	Homo sapiens T cell re 9.52e+01
	228	17	5.7	63916	26	HS69E11	Homo sapiens DNA seque 9.52e+01
	229	17	5.7	68700	26	AC002533	Human BAC RG024C13 fro 9.52e+01
c	230	17	5.7	69034	18	AF007261	Reclinomonas americana 9.52e+01
	231	17	5.7	78844	23	AB006708	Arabidopsis thaliana g 9.52e+01
	232	17	5.7	80536	17	AC005452	*** SEQUENCING IN PROG 9.52e+01
c	233	17	5.7	81522	18	AC004298	Drosophila melanogaste 9.52e+01
	234	17	5.7	87459	24	AB015477	Arabidopsis thaliana g 9.52e+01
	235	17	5.7	89328	26	HS398C22	Homo sapiens DNA seque 9.52e+01
c	236	17	5.7	89779	23	AB005234	Arabidopsis thaliana g 9.52e+01
	237	17	5.7	92802	17	AC004710	*** SEQUENCING IN PROG 9.52e+01
c	238	17	5.7	95565	26	AC000379	Human Chromosome 15 pa 9.52e+01
c	239	17	5.7	96180	17	AC005065	*** SEQUENCING IN PROG 9.52e+01
c	240	17	5.7	96642	23	AC002291	Arabidopsis thaliana c 9.52e+01
c	241	17	5.7	97916	27	HS408N23	Human DNA sequence fro 9.52e+01
c	242	17	5.7	99384	24	AF076243	Arabidopsis thaliana B 9.52e+01
c	243	17	5.7	100783	17	HS344127	Human DNA sequence *** 9.52e+01
	244	17	5.7	102204	26	AC003103	Homo sapiens chromosom 9.52e+01
c	245	17	5.7	103567	26	HS124C6	Homo sapiens DNA seque 9.52e+01
c	246	17	5.7	105127	26	AC004548	Homo sapiens PAC clone 9.52e+01

c	247	17	5.7	106184	23	TM018A10	Arabidopsis thaliana B 9.52e+01
c	248	17	5.7	107897	26	AP000057	Homo sapiens genomic D 9.52e+01
c	249	17	5.7	110000	17	CEY57A10	Caenorhabditis elegans 9.52e+01
c	250	17	5.7	110000	17	HSY214H10	Human DNA sequence *** 9.52e+01
c	251	17	5.7	112184	26	AC004054	Homo sapiens chromosom 9.52e+01
c	252	17	5.7	115838	17	CEY24F12	Caenorhabditis elegans 9.52e+01
c	253	17	5.7	116884	17	PFMAL1P1	Plasmodium falciparum 9.52e+01
c	254	17	5.7	117140	17	AC004880	*** SEQUENCING IN PROG 9.52e+01
c	255	17	5.7	117752	26	AC004743	Homo sapiens PAC clone 9.52e+01
	256	17	5.7	118451	27	AC004930	Homo sapiens PAC clone 9.52e+01
	257	17	5.7	118521	26	HUAC004514	Homo sapiens Chromosom 9.52e+01
	258	17	5.7	118907	17	HS799N4	Human DNA sequence *** 9.52e+01
c	259	17	5.7	128779	26	HS29K1	Human DNA sequence fro 9.52e+01
	260	17	5.7	129169	17	AC005506	*** SEQUENCING IN PROG 9.52e+01
c	261	17	5.7	129169	17	AC005506	*** SEQUENCING IN PROG 9.52e+01
c	262	17	5.7	131398	26	HS445C9	Human DNA sequence fro 9.52e+01
	263	17	5.7	135556	27	HSU85198	Homo sapiens BAC810, c 9.52e+01
	264	17	5.7	136299	26	HS359N14	Human DNA sequence fro 9.52e+01
	265	17	5.7	136746	17	AC004928	*** SEQUENCING IN PROG 9.52e+01
	266	17	5.7	137271	26	AC004552	Homo sapiens Xp22 BAC 9.52e+01
c	267	17	5.7	141851	26	HS75N13	Human DNA sequence fro 9.52e+01
c	268	17	5.7	143612	17	HS936P19	Human DNA sequence *** 9.52e+01
	269	17	5.7	144136	17	AC004987	*** SEQUENCING IN PROG 9.52e+01
c	270	17	5.7	144545	17	AC005287	*** SEQUENCING IN PROG 9.52e+01
	271	17	5.7	144763	17	CEY40D7	Caenorhabditis elegans 9.52e+01
c	272	17	5.7	144861	33	HHU43400	Human herpesvirus-7 (H 9.52e+01
	273	17	5.7	144868	27	HS80I19	Human DNA sequence fro 9.52e+01
	274	17	5.7	151519	26	HS231L4	Human DNA sequence *** 9.52e+01
	275	17	5.7	152184	17	CEY70G10	Caenorhabditis elegans 9.52e+01
c	276	17	5.7	153080	33	AF037218	Human herpesvirus 7 st 9.52e+01
c	277	17	5.7	154028	26	HSAC002087	Human BAC clone GS113D 9.52e+01
c	278	17	5.7	156994	17	HS1018K9	Human DNA sequence *** 9.52e+01
c	279	17	5.7	157248	26	AC002312	Human Chromosome 15 pa 9.52e+01
c	280	17	5.7	168056	26	AC004458	Homo sapiens BAC clone 9.52e+01
c	281	17	5.7	168238	17	AC005014	*** SEQUENCING IN PROG 9.52e+01
	282	17	5.7	169693	26	HS57A13	Human DNA sequence fro 9.52e+01
c	283	17	5.7	170101	33	ASU18466	African swine fever vi 9.52e+01
c	284	17	5.7	171681	26	HUAC002299	Homo sapiens Chromosom 9.52e+01
	285	17	5.7	173420	17	HS800F24	Human DNA sequence *** 9.52e+01
	286	17	5.7	175737	27	HS1033B10	Human DNA sequence *** 9.52e+01
	287	17	5.7	177380	26	AC002457	Human BAC clone RG060P 9.52e+01
c	288	17	5.7	185035	26	AC004770	Homo sapiens chromosom 9.52e+01
	289	17	5.7	185893	26	AC005191	Homo sapiens PAC clone 9.52e+01
	290	17	5.7	186454	17	AC005678	*** SEQUENCING IN PROG 9.52e+01
c	291	17	5.7	189959	17	AC005023	*** SEQUENCING IN PROG 9.52e+01
	292	17	5.7	190000	17	AC005150	*** SEQUENCING IN PROG 9.52e+01
	293	17	5.7	190482	26	AC004106	Homo sapiens Xp22 BAC 9.52e+01
c	294	17	5.7	190482	26	AC004106	Homo sapiens Xp22 BAC 9.52e+01
	295	17	5.7	190665	17	HS495O10	Human DNA sequence *** 9.52e+01
	296	17	5.7	191565	26	HUAC004531	Homo sapiens Chromosom 9.52e+01
c	297	17	5.7	192936	17	CEY95D11	Caenorhabditis elegans 9.52e+01

	298	17	5.7	196672	18	AC004767	Drosophila melanogaste 9.52e+01
	299	17	5.7	198060	17	AC004831	*** SEQUENCING IN PROG 9.52e+01
	300	17	5.7	200167	26	AC002480	Human BAC clone GS055K 9.52e+01
	301	17	5.7	201175	17	AC004676	*** SEQUENCING IN PROG 9.52e+01
	302	17	5.7	202004	26	HUAF001549	Human Chromosome 16 BA 9.52e+01
c	303	17	5.7	207957	26	AC004470	Homo sapiens Xp22 BAC 9.52e+01
c	304	17	5.7	208438	17	AC005034	*** SEQUENCING IN PROG 9.52e+01
	305	17	5.7	224664	17	AC004153	*** SEQUENCING IN PROG 9.52e+01
c	306	17	5.7	233941	17	CEY52B11	Caenorhabditis elegans 9.52e+01
	307	17	5.7	243041	26	AC004388	Homo Sapiens Chromosom 9.52e+01
	308	17	5.7	244133	17	HS32B1	Human DNA sequence *** 9.52e+01
c	309	17	5.7	245265	17	HS581F12	Human DNA sequence *** 9.52e+01
	310	17	5.7	247877	26	HUAE000661	Homo sapiens T-cell re 9.52e+01
c	311	17	5.7	251969	17	CEY79H2	Caenorhabditis elegans 9.52e+01
	312	17	5.7	255479	17	HS503N11	Human DNA sequence *** 9.52e+01
c	313	17	5.7	255479	17	HS503N11	Human DNA sequence *** 9.52e+01
	314	17	5.7	267156	26	U66059	Human germline T-cell 9.52e+01
	315	17	5.7	275138	17	CEY54E5	Caenorhabditis elegans 9.52e+01
	316	17	5.7	281376	17	CEY39C12	Caenorhabditis elegans 9.52e+01
c	317	17	5.7	323900	17	HS402G11	Human DNA sequence *** 9.52e+01
c	318	17	5.7	326569	17	AC004912	*** SEQUENCING IN PROG 9.52e+01
	319	17	5.7	336638	17	CEY75B8	Caenorhabditis elegans 9.52e+01
c	320	17	5.7	336638	17	CEY75B8	Caenorhabditis elegans 9.52e+01
c	321	16	5.4	151	28	MUSB2REC	Mouse B2 repetitive se 3.62e+02
c	322	16	5.4	209	30	G05809	human STS WI-6879. 3.62e+02
c	323	16	5.4	252	18	DLU94246	Drosophila sp. 'white 3.62e+02
c	324	16	5.4	252	18	DIU94244	Drosophila iki 16S rib 3.62e+02
c	325	16	5.4	261	25	HUMUT594A	Human STS UT594, 5' pr 3.62e+02
	326	16	5.4	273	30	G06300	human STS WI-6903. 3.62e+02
c	327	16	5.4	300	32	ANRNAP	A.nidulans RNA compone 3.62e+02
	328	16	5.4	325	18	SNNMTCYTBT	Sanctanus balli mitoch 3.62e+02
	329	16	5.4	331	30	DM98G5T	D. melanogaster STS de 3.62e+02
	330	16	5.4	339	18	XMU77861	Xantholobus muticus 12 3.62e+02
	331	16	5.4	366	18	DVU82927	Drosophila virilis his 3.62e+02
	332	16	5.4	372	32	MTHS93711	Heratemis sp. 16S rRNA 3.62e+02
	333	16	5.4	387	30	HS320YC5	H.sapiens (D9S1677) DN 3.62e+02
	334	16	5.4	406	20	MTMDRNA16	M.diastema mitochondri 3.62e+02
	335	16	5.4	407	30	G36810	SHGC-54257 Human Homo 3.62e+02
	336	16	5.4	408	18	PBU68153	Pholetesor bedelliae 1 3.62e+02
	337	16	5.4	413	18	HRGAME2	Haliotis rufescens fro 3.62e+02
	338	16	5.4	420	33	NPHHR2R	Bombyx mori nuclear po 3.62e+02
	339	16	5.4	450	18	AF003479	Betylobracon waterhous 3.62e+02
	340	16	5.4	457	32	CDZ79651	C.deichmanni mitochond 3.62e+02
	341	16	5.4	475	18	AF003484	Heterospilus sp. 16S r 3.62e+02
c	342	16	5.4	482	30	HUMUT5208	Human STS UT5208, sequ 3.62e+02
c	343	16	5.4	504	18	AB009907	Lucila kuroiwae mitoc 3.62e+02
	344	16	5.4	514	15	BUHARGSD	Buchners aphidicola Ar 3.62e+02
c	345	16	5.4	515	25	HSABLINT	Human DNA for breakpoi 3.62e+02
	346	16	5.4	573	33	HIVT1U124	Human immunodeficiency 3.62e+02
	347	16	5.4	573	33	HIVT1U126	Human immunodeficiency 3.62e+02
	348	16	5.4	573	33	HIVT1U120	Human immunodeficiency 3.62e+02

	349	16	5.4	579	33	HIVT1U466	Human immunodeficiency	3.62e+02
	350	16	5.4	579	33	HIVIFA97	HIV-1 proviral vif gen	3.62e+02
c	351	16	5.4	609	23	AF017367	Oryza sativa cytochrom	3.62e+02
c	352	16	5.4	614	15	AF022357	Legionella sainthelens	3.62e+02
	353	16	5.4	618	27	HUM2D8BA01	Homo sapiens full leng	3.62e+02
c	354	16	5.4	648	23	AF047971	Phoebanthus tenuifoliu	3.62e+02
c	355	16	5.4	650	23	AF047972	Phoebanthus grandiflor	3.62e+02
c	356	16	5.4	654	23	RICYCYTC	Rice mRNA for cytochro	3.62e+02
	357	16	5.4	663	24	AF051241	Picea mariana phosphog	3.62e+02
c	358	16	5.4	665	26	AG000645	Homo sapiens genomic D	3.62e+02
c	359	16	5.4	670	26	AG000649	Homo sapiens genomic D	3.62e+02
c	360	16	5.4	751	18	TVU26965	Trichomonas vaginalis	3.62e+02
c	361	16	5.4	761	18	ZAPMTRG16S	Zaprionus tuberculatus	3.62e+02
	362	16	5.4	784	18	DVUB8330	Drosophila virilis his	3.62e+02
	363	16	5.4	828	22	CP5RSIT2	Bacteriophage Cp-5 (S.	3.62e+02
c	364	16	5.4	828	22	CP5RSIT2	Bacteriophage Cp-5 (S.	3.62e+02
	365	16	5.4	844	18	RBU39423	Rhagoletis basiola 16S	3.62e+02
c	366	16	5.4	844	16	AF000958	Legionella longbeachae	3.62e+02
	367	16	5.4	859	19	OCGPPTKI	O.cuniculus mRNA for g	3.62e+02
c	368	16	5.4	881	16	LNU92213	Legionella nautarum ma	3.62e+02
c	369	16	5.4	948	16	LOU92214	Legionella oakridgensi	3.62e+02
	370	16	5.4	1001	18	DVU88331	Drosophila virilis his	3.62e+02
	371	16	5.4	1022	22	CP5RSIT1	Bacteriophage Cp-5 (S.	3.62e+02
	372	16	5.4	1022	24	AF079512	Mesembryanthemum cryst	3.62e+02
c	373	16	5.4	1022	22	CP5RSIT1	Bacteriophage Cp-5 (S.	3.62e+02
c	374	16	5.4	1052	30	G05833	human STS WI-7945.	3.62e+02
c	375	16	5.4	1058	15	ECOCYSQ	Escherichia coli ammon	3.62e+02
c	376	16	5.4	1067	33	RDVS11	Rice dwarf phytoreovir	3.62e+02
c	377	16	5.4	1155	18	PPAF001849	Paramecium primaurelia	3.62e+02
	378	16	5.4	1163	18	AF044861	Eurema nicippe large s	3.62e+02
c	379	16	5.4	1164	20	U71227	Desmognathus ochrophae	3.62e+02
c	380	16	5.4	1164	20	U71228	Desmognathus ochrophae	3.62e+02
c	381	16	5.4	1207	28	RATHSDIA	Rat mRNA for 20-alpha-	3.62e+02
c	382	16	5.4	1215	28	RAT20AHYDE	Rat 20 alpha-hydroxyst	3.62e+02
	383	16	5.4	1218	23	POU28355	Pythium oligandrum ATP	3.62e+02
c	384	16	5.4	1218	21	E06822	cDNA encoding 20 alpha	3.62e+02
c	385	16	5.4	1218	23	POU28355	Pythium oligandrum ATP	3.62e+02
c	386	16	5.4	1224	24	AB011546	Pleurotus ostreatus mn	3.62e+02
	387	16	5.4	1230	15	ECONEUSK92	E.coli glycosyl transf	3.62e+02
	388	16	5.4	1250	15	MC036	M.capricolum DNA for C	3.62e+02
c	389	16	5.4	1267	16	LSU92219	Legionella sainthelens	3.62e+02
c	390	16	5.4	1309	15	LLDNAMIP	L.longbeachae mip gene	3.62e+02
c	391	16	5.4	1324	18	DM166R	Drosophila mRNA to mit	3.62e+02
c	392	16	5.4	1355	29	LYARRLSUB	Mitochondrion Lymantri	3.62e+02
	393	16	5.4	1365	15	S82042	sodC=periplasmic coppe	3.62e+02
	394	16	5.4	1393	23	APEARS	Acremonium chrysogenum	3.62e+02
	395	16	5.4	1393	21	A11049	Synthetic nucleotide s	3.62e+02
	396	16	5.4	1393	21	E00963	ARS of acremonium chry	3.62e+02
c	397	16	5.4	1393	21	A11050	Synthetic nucleotide s	3.62e+02
	398	16	5.4	1432	22	BT4ORFS	Bacteriophage T4 orfs	3.62e+02
c	399	16	5.4	1433	26	AF054830	Homo sapiens interleuk	3.62e+02

c	400	16	5.4	1433	26	AF005039	Homo sapiens secretory	3.62e+02
	401	16	5.4	1447	28	MUSTRP1PR	Mus musculus tyrosinas	3.62e+02
	402	16	5.4	1459	23	ALFMSENODC	Medicago sativa leghem	3.62e+02
	403	16	5.4	1522	19	BTU19469	Bos taurus lysozyme ps	3.62e+02
	404	16	5.4	1584	20	XLHBOXH	Xenopus laevis XlHbox	3.62e+02
c	405	16	5.4	1648	18	MIDSMTDS2	D.subobscura mitochond	3.62e+02
c	406	16	5.4	1669	15	BCU94743	Bacillus cereus hemoly	3.62e+02
	407	16	5.4	1694	21	I82814	Sequence 9 from patent	3.62e+02
c	408	16	5.4	1711	23	SOYHSP176	Soybean (Glycine max)	3.62e+02
c	409	16	5.4	1723	18	ACU02684	Aplysia californica Y-	3.62e+02
	410	16	5.4	1757	28	RNU57552	Rattus norvegicus gluc	3.62e+02
	411	16	5.4	1883	26	AF055022	Homo sapiens clone 246	3.62e+02
c	412	16	5.4	1941	24	AF079307	Schizosaccharomyces po	3.62e+02
	413	16	5.4	1952	23	SCYKL117W	S.cerevisiae chromosom	3.62e+02
c	414	16	5.4	1952	23	SCYKL117W	S.cerevisiae chromosom	3.62e+02
c	415	16	5.4	2012	18	DROORFB	D.virilis homeobox pro	3.62e+02
	416	16	5.4	2013	18	DDCMFRNA	Dictyostelium discoide	3.62e+02
c	417	16	5.4	2042	19	SSGEMGBP	Sus scrofa mRNA for GT	3.62e+02
	418	16	5.4	2052	28	MMSIG41	M.musculus mRNA for SI	3.62e+02
c	419	16	5.4	2107	23	GMLBA5	Soybean leghaemoglobin	3.62e+02
c	420	16	5.4	2115	18	TTSNRU31	T.thermophila gene for	3.62e+02
c	421	16	5.4	2121	15	ECOAMTA	Escherichia coli ammon	3.62e+02
c	422	16	5.4	2156	25	HSU10550	Human Gem GTPase (gem)	3.62e+02
c	423	16	5.4	2220	15	LACAB2AA	L.lactis abi829 gene,	3.62e+02
c	424	16	5.4	2237	20	XELCP450	Xenopus laevis mRNA fo	3.62e+02
	425	16	5.4	2292	20	SAU80000	Sphyraena argentea lac	3.62e+02
c	426	16	5.4	2293	15	CLOALPTOX	Clostridium septicum g	3.62e+02
	427	16	5.4	2377	23	SCYKL116C	S.cerevisiae chromosom	3.62e+02
c	428	16	5.4	2377	23	SCYKL116C	S.cerevisiae chromosom	3.62e+02
c	429	16	5.4	2408	23	RICCYTC	Rice cytochrome c (OsC	3.62e+02
c	430	16	5.4	2424	18	DMU77310	Drosophila melanogaste	3.62e+02
c	431	16	5.4	2515	28	CGU49841	Cricetulus griseus Gol	3.62e+02
c	432	16	5.4	2539	23	POPPALG4B	Populus kitakamiensis	3.62e+02
c	433	16	5.4	2576	33	BDMSRLLA	Bean dwarf mosaic viru	3.62e+02
c	434	16	5.4	2598	23	SCYLL007C	S.cerevisiae chromosom	3.62e+02
	435	16	5.4	2599	16	CJU38524	Campylobacter jejuni p	3.62e+02
	436	16	5.4	2632	25	HSANTENII	H.sapiens gene for ang	3.62e+02
	437	16	5.4	2675	23	YSCCOT1A	Saccharomyces cerevisi	3.62e+02
c	438	16	5.4	2746	15	MVGLNA	Methanococcus voltae g	3.62e+02
c	439	16	5.4	2786	22	AF090369	Lactococcus phage P008	3.62e+02
c	440	16	5.4	2861	28	MUSMHABD	Mouse MHC class II H2-	3.62e+02
c	441	16	5.4	2887	23	SCYLL008W	S.cerevisiae chromosom	3.62e+02
c	442	16	5.4	2900	28	PLMTS1	Peromyscus leucopus re	3.62e+02
	443	16	5.4	3121	22	BAT4LNH	Bacteriophage T4 hoc g	3.62e+02
c	444	16	5.4	3125	28	AF011415	Mus musculus putative	3.62e+02
	445	16	5.4	3160	18	AF031904	Plasmodium berghei nuc	3.62e+02
c	446	16	5.4	3171	18	DDU68754	Dictyostelium discoide	3.62e+02
c	447	16	5.4	3361	15	SAPDHDNA	S.aureus pdhB, pdhC an	3.62e+02
	448	16	5.4	3380	25	HSFAN	H.sapiens mRNA for FAN	3.62e+02
c	449	16	5.4	3425	28	RNGCRY	Rat gene for lens gamm	3.62e+02
c	450	16	5.4	3542	23	SCYOR316C	S.cerevisiae chromosom	3.62e+02

c	451	16	5.4	3543	15	BTTOXD1	Bacillus thuringiensis	3.62e+02
	452	16	5.4	3625	23	PEAAPXIA	Pisum sativum ascorbat	3.62e+02
	453	16	5.4	3853	22	PT4NSP18X	Bacteriophage T4 nsp1-	3.62e+02
c	454	16	5.4	3925	15	PSEPLAOP	Pseudomonas aeruginosa	3.62e+02
c	455	16	5.4	4025	23	SCYDR014W	S.cerevisiae chromosom	3.62e+02
	456	16	5.4	4047	23	CHHCCLPC	H.carterae Chloroplast	3.62e+02
c	457	16	5.4	4100	23	CHCRPSBC	Chlamydomonas rheinhar	3.62e+02
c	458	16	5.4	4253	15	BACISRH4	Bacillus thuringiensis	3.62e+02
c	459	16	5.4	4451	21	I08083	Sequence 1 from Patent	3.62e+02
c	460	16	5.4	4451	21	I09103	Sequence 1 from Patent	3.62e+02
c	461	16	5.4	4456	15	SSU33768	Sulfolobus solfataricu	3.62e+02
	462	16	5.4	4459	18	AF063402	Anopheles gambiae defe	3.62e+02
c	463	16	5.4	4535	18	DDY10159	D.discoideum racGAP ge	3.62e+02
	464	16	5.4	4704	23	S43926	CHSB=chitinase [Phaseo	3.62e+02
c	465	16	5.4	4831	33	PXRCRMBX	Cowpox virus secreted	3.62e+02
c	466	16	5.4	4900	26	HSP2X7EX1	Homo sapiens P2X7 gene	3.62e+02
c	467	16	5.4	4915	28	MUSTHBS11	Mouse thrombospondin (	3.62e+02
c	468	16	5.4	4934	21	E01676	DNA sequence encoding	3.62e+02
	469	16	5.4	4941	23	YSPRTPTFL	S.pombe retrotransposo	3.62e+02
c	470	16	5.4	4948	28	MUSMHABE2	Mouse MHC class II H2-	3.62e+02
	471	16	5.4	5009	26	D80002	Human mRNA for KIAA018	3.62e+02
c	472	16	5.4	5010	15	STRFRUA	Streptococcus mutans b	3.62e+02
	473	16	5.4	5143	20	AF067418	Xenopus laevis focus f	3.62e+02
	474	16	5.4	5235	15	LMMESOHEY	L.mesenteroides mesC,	3.62e+02
c	475	16	5.4	5236	27	HSU67156	Human mitogen-activate	3.62e+02
c	476	16	5.4	5378	25	HUMGAPDHG	Human glyceraldehyde-3	3.62e+02
c	477	16	5.4	5847	20	FRCRABPI1	Fugu rubripes CRABP-I	3.62e+02
c	478	16	5.4	6418	20	GGU25125	Gallus gallus preproga	3.62e+02
	479	16	5.4	6550	33	PUULRNA	Puumala virus CG1820 R	3.62e+02
	480	16	5.4	6591	23	YSCMTCG06	Saccharomyces cerevisi	3.62e+02
c	481	16	5.4	7277	18	PFSC03086	Plasmodium falciparum	3.62e+02
c	482	16	5.4	7391	15	BORVMPPG	Borrelia hermsii rearr	3.62e+02
	483	16	5.4	7768	16	AF041817	Buchnera aphidicola pl	3.62e+02
	484	16	5.4	8351	18	DRONTM2A	Drosophila melanogaste	3.62e+02
c	485	16	5.4	8416	23	OLRKLP2	Oryza longistaminata r	3.62e+02
	486	16	5.4	8421	18	DVU39746	Drosophila virilis gla	3.62e+02
	487	16	5.4	8910	18	TETCNJBA	Tetrahymena thermophil	3.62e+02
c	488	16	5.4	8958	15	LLU17233	Lactococcus lactis pla	3.62e+02
	489	16	5.4	8987	33	U88826	HIV-1 isolate 92NG083	3.62e+02
	490	16	5.4	9132	28	D45201	Rat mRNA for neurofibr	3.62e+02
	491	16	5.4	9248	23	SCAF000227	Secale cereale omega s	3.62e+02
	492	16	5.4	9276	18	CELF40E12	Caenorhabditis elegans	3.62e+02
c	493	16	5.4	9416	25	HUMSEQX	Human microtubule-asso	3.62e+02
c	494	16	5.4	9968	18	CELT23A7	Caenorhabditis elegans	3.62e+02
c	495	16	5.4	9985	18	PCU43145	Plasmodium chabaudi re	3.62e+02
	496	16	5.4	10003	15	U67544	Methanococcus jannasch	3.62e+02
c	497	16	5.4	10111	18	CELC05H8	Caenorhabditis elegans	3.62e+02
	498	16	5.4	10118	15	BSUEROP	B.subtilis urease oper	3.62e+02
c	499	16	5.4	10159	15	U32770	Haemophilus influenzae	3.62e+02
	500	16	5.4	10171	15	U32832	Haemophilus influenzae	3.62e+02
c	501	16	5.4	10181	15	ECAE000492	Escherichia coli K-12	3.62e+02

c	502	16	5.4	10388	15	AE001118	Borrelia burgdorferi (	3.62e+02
	503	16	5.4	10402	15	AE001183	Borrelia burgdorferi (	3.62e+02
	504	16	5.4	10660	23	S93804	BAP1/ABF1/YRL50S=ARS-b	3.62e+02
c	505	16	5.4	10660	23	S93804	BAP1/ABF1/YRL50S=ARS-b	3.62e+02
	506	16	5.4	10702	15	ECAE000247	Escherichia coli K-12	3.62e+02
c	507	16	5.4	10819	15	ECAE000493	Escherichia coli K-12	3.62e+02
	508	16	5.4	10929	15	U67554	Methanococcus jannasch	3.62e+02
	509	16	5.4	10976	15	AE001127	Borrelia burgdorferi (	3.62e+02
c	510	16	5.4	10976	15	AE001127	Borrelia burgdorferi (	3.62e+02
	511	16	5.4	11514	18	CELK04E7	Caenorhabditis elegans	3.62e+02
c	512	16	5.4	11779	18	DDIMTRP	Cellullar slime mold m	3.62e+02
c	513	16	5.4	12229	18	CELF49E7	Caenorhabditis elegans	3.62e+02
c	514	16	5.4	12806	15	U67469	Methanococcus jannasch	3.62e+02
c	515	16	5.4	12971	15	MCRPCLUS	Mycoplasma capricolum	3.62e+02
	516	16	5.4	13407	28	MMPAFALIS3	Mus musculus Pafah/Li	3.62e+02
c	517	16	5.4	14466	18	CEY43C5B	Caenorhabditis elegans	3.62e+02
c	518	16	5.4	14563	15	SAU10927	Staphylococcus aureus	3.62e+02
	519	16	5.4	15359	15	D90792	E.coli genomic DNA, Ko	3.62e+02
c	520	16	5.4	15859	18	CEF07D3	Caenorhabditis elegans	3.62e+02
c	521	16	5.4	15923	26	HUMTCRBVDL	Homo sapiens (clone HV	3.62e+02
	522	16	5.4	16463	26	HS140H19	Human DNA sequence fro	3.62e+02
c	523	16	5.4	16741	18	PFSC03095	Plasmodium falciparum	3.62e+02
c	524	16	5.4	17492	15	U67539	Methanococcus jannasch	3.62e+02
c	525	16	5.4	17758	18	CELT11F4	Caenorhabditis elegans	3.62e+02
	526	16	5.4	17879	15	AE001161	Borrelia burgdorferi (	3.62e+02
	527	16	5.4	18124	18	CEU26736	Caenorhabditis elegans	3.62e+02
c	528	16	5.4	19014	18	PFSC03009	Plasmodium falciparum	3.62e+02
	529	16	5.4	19517	18	DMU37541	Drosophila melanogaste	3.62e+02
c	530	16	5.4	19692	18	CEF02C12	Caenorhabditis elegans	3.62e+02
	531	16	5.4	19911	26	AC004507	Homo sapiens chromosom	3.62e+02
	532	16	5.4	19929	15	D90791	E.coli genomic DNA, Ko	3.62e+02
	533	16	5.4	20343	20	GGVITTIG	Chicken vitellogenin I	3.62e+02
c	534	16	5.4	20495	18	CELC02C2	C. elegans cosmid C02C	3.62e+02
	535	16	5.4	20535	15	BACSRPAB	B.subtilis srfAA and s	3.62e+02
	536	16	5.4	20856	18	CEK09A11	Caenorhabditis elegans	3.62e+02
	537	16	5.4	21262	18	PFSC03028	Plasmodium falciparum	3.62e+02
	538	16	5.4	21598	18	CEM04G12	Caenorhabditis elegans	3.62e+02
c	539	16	5.4	21948	18	CEF58B3	Caenorhabditis elegans	3.62e+02
c	540	16	5.4	22729	18	CELF53F10	Caenorhabditis elegans	3.62e+02
c	541	16	5.4	23536	18	CEF57B7	Caenorhabditis elegans	3.62e+02
c	542	16	5.4	23895	18	CELT27B7	Caenorhabditis elegans	3.62e+02
c	543	16	5.4	24643	23	SPAC6C3	S.pombe chromosome I c	3.62e+02
	544	16	5.4	25003	18	CELC55B6	Caenorhabditis elegans	3.62e+02
	545	16	5.4	26258	18	CELK816	Caenorhabditis elegans	3.62e+02
c	546	16	5.4	26568	18	CELF54H5	Caenorhabditis elegans	3.62e+02
	547	16	5.4	26581	18	CEF57G8	Caenorhabditis elegans	3.62e+02
c	548	16	5.4	27174	17	AC004089	*** SEQUENCING IN PROG	3.62e+02
c	549	16	5.4	27178	18	CELC09H5	Caenorhabditis elegans	3.62e+02
c	550	16	5.4	27559	23	SC8339	S.cerevisiae chromosom	3.62e+02
	551	16	5.4	27576	18	CEF16B12	Caenorhabditis elegans	3.62e+02
	552	16	5.4	27694	23	HASMT	Hansenula wingei mitoc	3.62e+02

	553	16	5.4	27764	18	CET21E8	Caenorhabditis elegans 3.62e+02
	554	16	5.4	27878	25	HSJ3E10	Human DNA sequence fro 3.62e+02
	555	16	5.4	27935	18	CELC07D10	Caenorhabditis elegans 3.62e+02
	556	16	5.4	28039	18	CET27A8	Caenorhabditis elegans 3.62e+02
c	557	16	5.4	28105	18	CELD2024	Caenorhabditis elegans 3.62e+02
c	558	16	5.4	28122	18	CELT08A9	Caenorhabditis elegans 3.62e+02
c	559	16	5.4	28537	18	CELF19G12	Caenorhabditis elegans 3.62e+02
	560	16	5.4	28684	18	CEC44H9	Caenorhabditis elegans 3.62e+02
c	561	16	5.4	28988	18	CELF59C12	Caenorhabditis elegans 3.62e+02
	562	16	5.4	28999	18	CEF44E5	Caenorhabditis elegans 3.62e+02
c	563	16	5.4	29777	18	CEF46C3	Caenorhabditis elegans 3.62e+02
c	564	16	5.4	29929	18	CET15D6	Caenorhabditis elegans 3.62e+02
	565	16	5.4	30020	18	CELT07H3	Caenorhabditis elegans 3.62e+02
	566	16	5.4	30427	18	CELF44C4	Caenorhabditis elegans 3.62e+02
c	567	16	5.4	30655	18	CET24A11	Caenorhabditis elegans 3.62e+02
c	568	16	5.4	30722	18	CELT19A5	Caenorhabditis elegans 3.62e+02
	569	16	5.4	30919	18	CET08G11	Caenorhabditis elegans 3.62e+02
c	570	16	5.4	31124	18	CELC16B9	Caenorhabditis elegans 3.62e+02
c	571	16	5.4	32664	18	CEF49C12	Caenorhabditis elegans 3.62e+02
	572	16	5.4	32710	23	SPAC23H3	S.pombe chromosome I c 3.62e+02
c	573	16	5.4	32838	18	CELR08F11	Caenorhabditis elegans 3.62e+02
c	574	16	5.4	33318	18	CELF32E10	Caenorhabditis elegans 3.62e+02
	575	16	5.4	33429	18	CEF09F3	Caenorhabditis elegans 3.62e+02
	576	16	5.4	33677	18	CELF40F4	Caenorhabditis elegans 3.62e+02
c	577	16	5.4	33837	18	CEF27C8	Caenorhabditis elegans 3.62e+02
	578	16	5.4	34008	18	CELF42G9	Caenorhabditis elegans 3.62e+02
	579	16	5.4	34224	18	CEC10C5	Caenorhabditis elegans 3.62e+02
	580	16	5.4	34660	18	CELT27F7	Caenorhabditis elegans 3.62e+02
c	581	16	5.4	34841	18	CEC46F11	Caenorhabditis elegans 3.62e+02
	582	16	5.4	35259	24	F3D13	Arabidopsis thaliana B 3.62e+02
	583	16	5.4	35281	25	HSL96A2	Human DNA sequence fro 3.62e+02
	584	16	5.4	35303	18	CELC44B11	Caenorhabditis elegans 3.62e+02
c	585	16	5.4	35303	18	CELC44B11	Caenorhabditis elegans 3.62e+02
c	586	16	5.4	35323	26	U73646	Human Chromosome 11 Co 3.62e+02
c	587	16	5.4	35452	23	SC354CHIV	S.cerevisiae 35.4Kb DN 3.62e+02
	588	16	5.4	35500	18	CEC54C6	Caenorhabditis elegans 3.62e+02
	589	16	5.4	35503	18	CEW09D6	Caenorhabditis elegans 3.62e+02
	590	16	5.4	35564	18	CELT22B11	Caenorhabditis elegans 3.62e+02
c	591	16	5.4	35710	23	SCCHXVCOS	S.cerevisiae DNA from 3.62e+02
	592	16	5.4	35713	18	CELF22D3	Caenorhabditis elegans 3.62e+02
c	593	16	5.4	35798	26	AC004094	Human Cosmid g1248a085 3.62e+02
	594	16	5.4	35880	18	CELK02F3	Caenorhabditis elegans 3.62e+02
	595	16	5.4	36123	17	AC004185	*** SEQUENCING IN PROG 3.62e+02
c	596	16	5.4	36170	18	CER10E4	Caenorhabditis elegans 3.62e+02
c	597	16	5.4	36254	18	CELK12D9	Caenorhabditis elegans 3.62e+02
	598	16	5.4	36431	17	AC005712	*** SEQUENCING IN PROG 3.62e+02
c	599	16	5.4	36510	26	AC004397	Homo sapiens Cosmid g1 3.62e+02
c	600	16	5.4	36798	26	HSE78G1	Human DNA sequence fro 3.62e+02
	601	16	5.4	36811	18	CELT20B12	Caenorhabditis elegans 3.62e+02
	602	16	5.4	36873	18	CELC34E10	Caenorhabditis elegans 3.62e+02
	603	16	5.4	36993	18	CELT10H9	Caenorhabditis elegans 3.62e+02

c	604	16	5.4	37206	18	CELT13H2	Caenorhabditis elegans 3.62e+02
c	605	16	5.4	37215	17	AC005459	*** SEQUENCING IN PROG 3.62e+02
	606	16	5.4	37249	18	CED2045	Caenorhabditis elegans 3.62e+02
	607	16	5.4	37273	23	SPAC31A2	S.pombe chromosome I c 3.62e+02
c	608	16	5.4	37274	26	AC005272	Homo sapiens chromosom 3.62e+02
	609	16	5.4	37353	18	CET12D8	Caenorhabditis elegans 3.62e+02
c	610	16	5.4	37404	23	SC8119	S.cerevisiae chromosom 3.62e+02
	611	16	5.4	37447	18	CEF11A5	Caenorhabditis elegans 3.62e+02
	612	16	5.4	37614	18	CELT10H10	Caenorhabditis elegans 3.62e+02
c	613	16	5.4	37666	18	CEF58G1	Caenorhabditis elegans 3.62e+02
	614	16	5.4	37704	18	CEF36G9	Caenorhabditis elegans 3.62e+02
	615	16	5.4	37773	26	AP027207	Homo sapiens chromosom 3.62e+02
c	616	16	5.4	37895	10	E12050	Genes involved in eico 3.62e+02
c	617	16	5.4	37895	21	I72332	Sequence 1 from patent 3.62e+02
c	618	16	5.4	37895	15	SSU73935	Shewanella sp. SCRC-27 3.62e+02
c	619	16	5.4	37913	21	E06745	DNA encoding eicosapen 3.62e+02
	620	16	5.4	38060	18	CET13F2	Caenorhabditis elegans 3.62e+02
c	621	16	5.4	38394	27	HSN62C4	Human DNA sequence fro 3.62e+02
c	622	16	5.4	38415	18	CELM01G4	Caenorhabditis elegans 3.62e+02
	623	16	5.4	38627	18	CELC01G5	Caenorhabditis elegans 3.62e+02
c	624	16	5.4	38666	23	SPAC24C9	S.pombe chromosome I c 3.62e+02
	625	16	5.4	38790	17	AC005671	*** SEQUENCING IN PROG 3.62e+02
c	626	16	5.4	38906	18	CELF13C5	Caenorhabditis elegans 3.62e+02
	627	16	5.4	39086	18	CEK07A1	Caenorhabditis elegans 3.62e+02
	628	16	5.4	39210	25	HSE118G4	Human DNA sequence fro 3.62e+02
	629	16	5.4	39469	18	CELF19B10	Caenorhabditis elegans 3.62e+02
c	630	16	5.4	39602	26	HSAC000373	Human cosmid g1862d218 3.62e+02
c	631	16	5.4	39784	18	CEC14A4	Caenorhabditis elegans 3.62e+02
c	632	16	5.4	39925	18	CEK06B4	Caenorhabditis elegans 3.62e+02
	633	16	5.4	39931	18	CELC54E4	Caenorhabditis elegans 3.62e+02
c	634	16	5.4	40293	18	CELC45G9	Caenorhabditis elegans 3.62e+02
	635	16	5.4	40344	18	CED2023	Caenorhabditis elegans 3.62e+02
c	636	16	5.4	40460	27	AC005492	Homo sapiens clone UWG 3.62e+02
	637	16	5.4	40572	18	CELC14F11	Caenorhabditis elegans 3.62e+02
c	638	16	5.4	40622	18	CELT22D1	Caenorhabditis elegans 3.62e+02
	639	16	5.4	40901	18	CELF14D12	Caenorhabditis elegans 3.62e+02
	640	16	5.4	40940	18	CED2013	Caenorhabditis elegans 3.62e+02
	641	16	5.4	40977	26	HSAC000362	Human cosmid g1346a094 3.62e+02
	642	16	5.4	41193	18	CELC46H11	Caenorhabditis elegans 3.62e+02
c	643	16	5.4	41275	18	CELF47D12	Caenorhabditis elegans 3.62e+02
c	644	16	5.4	41378	26	AC005381	Homo sapiens chromosom 3.62e+02
	645	16	5.4	41645	18	CELC03H5	Caenorhabditis elegans 3.62e+02
c	646	16	5.4	41645	18	CELC03H5	Caenorhabditis elegans 3.62e+02
	647	16	5.4	41702	26	HSAC000366	Human cosmid g1862x046 3.62e+02
c	648	16	5.4	41702	26	HSAC000366	Human cosmid g1862x046 3.62e+02
	649	16	5.4	41926	18	CELC54D1	Caenorhabditis elegans 3.62e+02
	650	16	5.4	42031	18	CEC35A5	Caenorhabditis elegans 3.62e+02
c	651	16	5.4	42297	18	CEC33A12	Caenorhabditis elegans 3.62e+02
	652	16	5.4	42310	18	CEC18D1	Caenorhabditis elegans 3.62e+02
c	653	16	5.4	42361	18	CEC25D7	Caenorhabditis elegans 3.62e+02
	654	16	5.4	42430	18	CELF59E12	Caenorhabditis elegans 3.62e+02

c	655	16	5.4	43441	18	CED2005	Caenorhabditis elegans 3.62e+02
c	656	16	5.4	43526	18	CELC10H11	Caenorhabditis elegans 3.62e+02
c	657	16	5.4	43676	23	SCCEN12RG	S.cerevisiae DNA from 3.62e+02
	658	16	5.4	43782	18	CEF59B2	Caenorhabditis elegans 3.62e+02
c	659	16	5.4	43788	18	CEC10C6	Caenorhabditis elegans 3.62e+02
	660	16	5.4	43815	18	CEC07E3	Caenorhabditis elegans 3.62e+02
	661	16	5.4	43891	18	CELC24A11	Caenorhabditis elegans 3.62e+02
c	662	16	5.4	43994	18	CEC08E8	Caenorhabditis elegans 3.62e+02
c	663	16	5.4	44371	18	CEC47E8	Caenorhabditis elegans 3.62e+02
c	664	16	5.4	44422	18	L32205	Drosophila melanogaste 3.62e+02
c	665	16	5.4	44733	18	CELC50C3	C. elegans cosmid C50C 3.62e+02
	666	16	5.4	45806	27	HSB33F2	Human DNA sequence fro 3.62e+02
	667	16	5.4	46480	18	CEC53A5	Caenorhabditis elegans 3.62e+02
	668	16	5.4	46717	17	HSM102C10	Human DNA sequence *** 3.62e+02
c	669	16	5.4	52593	18	CEY45F10B	Caenorhabditis elegans 3.62e+02
c	670	16	5.4	53185	25	HS315B17	Human DNA sequence fro 3.62e+02
c	671	16	5.4	54670	28	RATCRYG	Rat gamma-crystallin g 3.62e+02
	672	16	5.4	55795	27	AP000573	Homo sapiens homogenti 3.62e+02
c	673	16	5.4	58736	17	AC005416	*** SEQUENCING IN PROG 3.62e+02
	674	16	5.4	59648	24	AB016870	Arabidopsis thaliana g 3.62e+02
	675	16	5.4	59991	18	AC003120	Drosophila melanogaste 3.62e+02
c	676	16	5.4	62262	26	HSTCRBK17	Homo sapiens T cell re 3.62e+02
	677	16	5.4	63646	25	HSJ256K2A	Human DNA sequence fro 3.62e+02
c	678	16	5.4	63769	17	CER08A2	Caenorhabditis elegans 3.62e+02
c	679	16	5.4	65196	18	AC004274	Drosophila melanogaste 3.62e+02
	680	16	5.4	65325	24	AB013393	Arabidopsis thaliana g 3.62e+02
	681	16	5.4	65958	23	AB009056	Arabidopsis thaliana g 3.62e+02
	682	16	5.4	69013	25	HS138A5	Human DNA sequence fro 3.62e+02
	683	16	5.4	71153	26	HUAE000662	Homo sapiens T-cell re 3.62e+02
	684	16	5.4	72433	17	AC005408	*** SEQUENCING IN PROG 3.62e+02
c	685	16	5.4	73067	26	HS147M19	Homo sapiens DNA seque 3.62e+02
	686	16	5.4	73805	24	AC004135	Genomic sequence for A 3.62e+02
	687	16	5.4	74406	24	F2P3	Arabidopsis thaliana B 3.62e+02
	688	16	5.4	74560	26	HS433B8	Human DNA sequence fro 3.62e+02
	689	16	5.4	75145	26	HS362G17	Human DNA sequence fro 3.62e+02
c	690	16	5.4	75188	23	AB005238	Arabidopsis thaliana g 3.62e+02
c	691	16	5.4	75270	27	AP023268	Homo sapiens clk2 kina 3.62e+02
c	692	16	5.4	76329	23	AB007648	Arabidopsis thaliana g 3.62e+02
c	693	16	5.4	77758	24	AB016885	Arabidopsis thaliana g 3.62e+02
	694	16	5.4	78181	23	AB011477	Arabidopsis thaliana g 3.62e+02
	695	16	5.4	79716	26	AC004502	Homo sapiens chromosom 3.62e+02
c	696	16	5.4	79976	23	AB008264	Arabidopsis thaliana g 3.62e+02
	697	16	5.4	80559	3	AC004570	*** SEQUENCING IN PROG 3.62e+02
	698	16	5.4	80659	27	AC004553	Homo sapiens Xp22 bin 3.62e+02
c	699	16	5.4	80803	3	AC004275	*** SEQUENCING IN PROG 3.62e+02
	700	16	5.4	81004	23	ATAC002521	Arabidopsis thaliana c 3.62e+02
c	701	16	5.4	81053	18	AC005762	Drosophila melanogaste 3.62e+02
c	702	16	5.4	81421	17	CEC53C7	Caenorhabditis elegans 3.62e+02
c	703	16	5.4	82225	25	HS253J14	Human DNA sequence fro 3.62e+02
c	704	16	5.4	82230	26	AC004452	Homo sapiens PAC clone 3.62e+02
c	705	16	5.4	82291	27	AC005376	Homo sapiens PAC clone 3.62e+02

c	706	16	5.4	82484	23	ATAC002341	Arabidopsis thaliana c 3.62e+02
	707	16	5.4	82918	23	AB010071	Arabidopsis thaliana g 3.62e+02
	708	16	5.4	82976	27	AC005031	Homo sapiens P1 clone 3.62e+02
	709	16	5.4	83450	23	AB009054	Arabidopsis thaliana g 3.62e+02
	710	16	5.4	83605	26	AC004220	Homo sapiens chromosom 3.62e+02
c	711	16	5.4	83684	26	AC003959	Homo sapiens chromosom 3.62e+02
	712	16	5.4	84107	26	HS248E1	Human DNA sequence fro 3.62e+02
c	713	16	5.4	85444	26	AP000023	Homo sapiens genomic D 3.62e+02
c	714	16	5.4	85832	26	HS344F17	Human DNA sequence fro 3.62e+02
	715	16	5.4	85938	17	AC004725	*** SEQUENCING IN PROG 3.62e+02
c	716	16	5.4	85953	17	CEH04L24	Caenorhabditis elegans 3.62e+02
c	717	16	5.4	86121	24	AB017068	Arabidopsis thaliana g 3.62e+02
	718	16	5.4	86419	18	AC004345	Drosophila melanogaste 3.62e+02
	719	16	5.4	87077	26	HUAC004097	Human Chromosome 16 BA 3.62e+02
c	720	16	5.4	87124	18	AC004371	Drosophila melanogaste 3.62e+02
c	721	16	5.4	87552	17	HS217016	Human DNA sequence *** 3.62e+02
	722	16	5.4	88057	18	AC004120	Drosophila melanogaste 3.62e+02
c	723	16	5.4	91071	23	ATAC002339	Arabidopsis thaliana c 3.62e+02
	724	16	5.4	91638	17	AC002489	*** SEQUENCING IN PROG 3.62e+02
c	725	16	5.4	93312	26	HS32I10	Human DNA sequence fro 3.62e+02
c	726	16	5.4	94482	24	ATF26P21	Arabidopsis thaliana D 3.62e+02
	727	16	5.4	95129	26	AC003099	Homo sapiens chromosom 3.62e+02
	728	16	5.4	95824	23	ATU90439	Arabidopsis thaliana c 3.62e+02
c	729	16	5.4	95824	23	ATU90439	Arabidopsis thaliana c 3.62e+02
	730	16	5.4	95886	17	CEH27C11	Caenorhabditis elegans 3.62e+02
c	731	16	5.4	95921	26	AC004141	Homo sapiens PAC clone 3.62e+02
	732	16	5.4	96079	26	HS717L17	Human DNA sequence fro 3.62e+02
	733	16	5.4	96574	23	ATT18B16	Arabidopsis thaliana D 3.62e+02
	734	16	5.4	96625	25	HSAC000118	Human BAC clone RG072E 3.62e+02
c	735	16	5.4	96625	25	HSAC000118	Human BAC clone RG072E 3.62e+02
c	736	16	5.4	97494	24	T15F16	Arabidopsis thaliana B 3.62e+02
	737	16	5.4	97789	23	AC004255	Arabidopsis thaliana B 3.62e+02
	738	16	5.4	98645	18	AC005267	Drosophila melanogaste 3.62e+02
	739	16	5.4	98913	24	AF069442	Arabidopsis thaliana B 3.62e+02
	740	16	5.4	99433	17	AC005398	*** SEQUENCING IN PROG 3.62e+02
c	741	16	5.4	99445	17	HS201D17	Human DNA sequence *** 3.62e+02
	742	16	5.4	100000	26	AP000017	Homo sapiens genomic D 3.62e+02
c	743	16	5.4	100000	26	AP000037	Homo sapiens genomic D 3.62e+02
c	744	16	5.4	100000	26	AP000033	Homo sapiens genomic D 3.62e+02
	745	16	5.4	100808	26	HS94G16	Human DNA sequence fro 3.62e+02
c	746	16	5.4	100808	26	HS94G16	Human DNA sequence fro 3.62e+02
	747	16	5.4	100923	26	HSAC002124	Human BAC clone RG1800 3.62e+02
c	748	16	5.4	101196	25	HS463A9	Human DNA sequence fro 3.62e+02
c	749	16	5.4	101370	26	AC003962	Homo sapiens chromosom 3.62e+02
c	750	16	5.4	101933	24	AC004146	Arabidopsis thaliana c 3.62e+02
c	751	16	5.4	102004	24	ATAC005395	Arabidopsis thaliana c 3.62e+02
	752	16	5.4	102653	33	CTSCG	Lymphocystis disease v 3.62e+02
	753	16	5.4	103146	26	HS187N21	Human DNA sequence fro 3.62e+02
	754	16	5.4	103319	26	AC004217	Homo sapiens 12q24.1 P 3.62e+02
c	755	16	5.4	104016	24	ATAC004138	Arabidopsis thaliana c 3.62e+02
c	756	16	5.4	104377	17	CEY62H9	Caenorhabditis elegans 3.62e+02

	757	16	5.4	104817	27	HS681J21	Human DNA sequence fro 3.62e+02
	758	16	5.4	105178	17	HS470K1	Human DNA sequence *** 3.62e+02
c	759	16	5.4	105528	24	ATAC002392	Arabidopsis thaliana c 3.62e+02
	760	16	5.4	106172	26	AC004111	Homo sapiens BAC clone 3.62e+02
c	761	16	5.4	106172	26	AC004111	Homo sapiens BAC clone 3.62e+02
	762	16	5.4	108642	26	AC004451	Homo sapiens PAC clone 3.62e+02
	763	16	5.4	108847	23	ATAF002109	Arabidopsis thaliana c 3.62e+02
c	764	16	5.4	109859	26	HS747E2	Homo sapiens DNA seque 3.62e+02
c	765	16	5.4	110000	17	CEY51A2	Caenorhabditis elegans 3.62e+02
c	766	16	5.4	110000	17	CEY45F10	Caenorhabditis elegans 3.62e+02
c	767	16	5.4	110000	17	CEY105C5	Caenorhabditis elegans 3.62e+02
c	768	16	5.4	110514	23	F7G19	Sequence of BAC F7G19 3.62e+02
c	769	16	5.4	111062	17	HSAC001237	*** SEQUENCING IN PROG 3.62e+02
c	770	16	5.4	111448	18	AC005334	Drosophila melanogaste 3.62e+02
c	771	16	5.4	111650	26	AC005215	Homo sapiens chromosom 3.62e+02
c	772	16	5.4	112253	24	F9D12	Arabidopsis thaliana B 3.62e+02
	773	16	5.4	112726	17	CEY21F11	Caenorhabditis elegans 3.62e+02
	774	16	5.4	113567	23	AF058826	Arabidopsis thaliana B 3.62e+02
c	775	16	5.4	114736	18	PFMAL3P3	Plasmodium falciparum 3.62e+02
c	776	16	5.4	115781	17	AC005407	*** SEQUENCING IN PROG 3.62e+02
c	777	16	5.4	115851	23	ATAC002505	Arabidopsis thaliana c 3.62e+02
c	778	16	5.4	115984	26	HS292P10	Human DNA sequence fro 3.62e+02
c	779	16	5.4	116298	17	AC004942	*** SEQUENCING IN PROG 3.62e+02
c	780	16	5.4	116466	26	HSAC002074	Human BAC clone GS056H 3.62e+02
	781	16	5.4	116961	17	HS633020	Human DNA sequence *** 3.62e+02
	782	16	5.4	117161	17	CEY43F11	Caenorhabditis elegans 3.62e+02
	783	16	5.4	117252	17	AC005091	*** SEQUENCING IN PROG 3.62e+02
	784	16	5.4	117954	26	HSAC002065	Human BAC clone RG021N 3.62e+02
c	785	16	5.4	118067	26	HS163M9	Homo sapiens DNA seque 3.62e+02
	786	16	5.4	118153	26	HS249C1	Human DNA sequence fro 3.62e+02
	787	16	5.4	119298	27	AC005608	Homo sapiens chromosom 3.62e+02
	788	16	5.4	119420	26	HS997K18	Human DNA sequence fro 3.62e+02
	789	16	5.4	119704	23	OSCHLPLXX	O.sinensis complete ch 3.62e+02
c	790	16	5.4	119704	23	OSCHLPLXX	O.sinensis complete ch 3.62e+02
c	791	16	5.4	120067	24	T5C13	Arabidopsis thaliana B 3.62e+02
	792	16	5.4	120832	17	HS836E8	Human DNA sequence *** 3.62e+02
c	793	16	5.4	121019	26	AF042091	Homo sapiens chromosom 3.62e+02
	794	16	5.4	121055	17	AC004823	*** SEQUENCING IN PROG 3.62e+02
	795	16	5.4	121524	24	AF041468	Guillardia theta compl 3.62e+02
c	796	16	5.4	122121	26	AC004543	Homo sapiens BAC clone 3.62e+02
c	797	16	5.4	122256	17	AC005443	*** SEQUENCING IN PROG 3.62e+02
c	798	16	5.4	122322	23	ATF6I18	Arabidopsis thaliana D 3.62e+02
c	799	16	5.4	123331	26	AC005164	Homo sapiens BAC clone 3.62e+02
c	800	16	5.4	123498	17	HS126A5	Human DNA sequence *** 3.62e+02
	801	16	5.4	123891	26	AC004549	Homo sapiens BAC clone 3.62e+02
	802	16	5.4	124010	26	HS340G1	Human DNA sequence fro 3.62e+02
c	803	16	5.4	124801	17	HS292E10	Human DNA sequence *** 3.62e+02
	804	16	5.4	125638	26	AC004706	Homo sapiens chromosom 3.62e+02
c	805	16	5.4	125983	26	HS306D1	Human DNA sequence fro 3.62e+02
	806	16	5.4	127356	17	AC005063	*** SEQUENCING IN PROG 3.62e+02
c	807	16	5.4	127514	26	AC004417	Homo sapiens BAC clone 3.62e+02

c	808	16	5.4	128011	26	HS333E23	Human DNA sequence fro 3.62e+02
	809	16	5.4	128413	33	NPHT3COMP	Bombyx mori nuclear po 3.62e+02
c	810	16	5.4	129515	17	HS886K2	Human DNA sequence *** 3.62e+02
c	811	16	5.4	129969	26	HS204E5	Human DNA sequence fro 3.62e+02
c	812	16	5.4	130223	26	AP004863	Homo sapiens chromosom 3.62e+02
	813	16	5.4	130381	26	AC002402	Human Chromosome 11 pa 3.62e+02
	814	16	5.4	131078	27	HSU80017	Homo sapiens basic tra 3.62e+02
	815	16	5.4	131100	17	AC004666	*** SEQUENCING IN PROG 3.62e+02
c	816	16	5.4	131318	17	CEY42E10	Caenorhabditis elegans 3.62e+02
	817	16	5.4	132012	17	AC005713	*** SEQUENCING IN PROG 3.62e+02
c	818	16	5.4	132012	17	AC005713	*** SEQUENCING IN PROG 3.62e+02
	819	16	5.4	132029	26	HS795G23	Human DNA sequence *** 3.62e+02
c	820	16	5.4	132666	18	AC005270	Drosophila melanogaste 3.62e+02
c	821	16	5.4	133894	21	A48542	Sequence 1 from Patent 3.62e+02
c	822	16	5.4	133894	33	NPHCG	Autographa californica 3.62e+02
	823	16	5.4	133925	26	AC004562	Homo sapiens chromosom 3.62e+02
	824	16	5.4	134149	17	AC004944	*** SEQUENCING IN PROG 3.62e+02
	825	16	5.4	134226	33	IH1CG	Ictalurid herpesvirus 3.62e+02
	826	16	5.4	134525	23	CHOSXX	Rice complete chloropl 3.62e+02
c	827	16	5.4	135007	26	HS423N22	Homo sapiens DNA seque 3.62e+02
	828	16	5.4	135259	17	AC005018	*** SEQUENCING IN PROG 3.62e+02
c	829	16	5.4	135365	17	HS191L6	Human DNA sequence *** 3.62e+02
	830	16	5.4	136037	26	AC004104	Homo sapiens 12q24 PAC 3.62e+02
	831	16	5.4	136142	17	HS37J18	Human DNA sequence *** 3.62e+02
c	832	16	5.4	136335	24	AC005275	Arabidopsis thaliana B 3.62e+02
c	833	16	5.4	136859	17	HS191N21	Human DNA sequence *** 3.62e+02
c	834	16	5.4	137022	26	AC004231	Homo sapiens chromosom 3.62e+02
	835	16	5.4	137094	17	AC005728	*** SEQUENCING IN PROG 3.62e+02
	836	16	5.4	137111	26	HS20J23	Human DNA sequence *** 3.62e+02
c	837	16	5.4	137271	26	AC004552	Homo sapiens Xp22 BAC 3.62e+02
c	838	16	5.4	137277	26	AC002431	Human BAC clone RG180F 3.62e+02
	839	16	5.4	137947	27	HS45I4	Human DNA sequence *** 3.62e+02
c	840	16	5.4	138162	17	AC005100	*** SEQUENCING IN PROG 3.62e+02
	841	16	5.4	141475	26	AC002980	Homo sapiens Xp22 BAC 3.62e+02
	842	16	5.4	141878	26	AC000004	Genomic sequence from 3.62e+02
c	843	16	5.4	142949	17	AC004960	*** SEQUENCING IN PROG 3.62e+02
	844	16	5.4	143092	18	CEY17G7B	Caenorhabditis elegans 3.62e+02
	845	16	5.4	143317	17	AC000029	*** SEQUENCING IN PROG 3.62e+02
c	846	16	5.4	144239	26	AC004015	Human BAC clone RG333F 3.62e+02
	847	16	5.4	145032	17	CEY2H9	Caenorhabditis elegans 3.62e+02
c	848	16	5.4	146360	26	HS1156N12	Human DNA sequence *** 3.62e+02
	849	16	5.4	148280	26	AC004474	Homo sapiens chromosom 3.62e+02
	850	16	5.4	148851	26	HS155D22	Human DNA sequence fro 3.62e+02
c	851	16	5.4	149554	25	HSAC001111	Human BAC clone 068P20 3.62e+02
	852	16	5.4	149752	26	AC004616	Homo sapiens Xp22 BAC 3.62e+02
	853	16	5.4	149764	26	HSAC002076	Human BAC clone GS345D 3.62e+02
	854	16	5.4	150000	17	AC005511	*** SEQUENCING IN PROG 3.62e+02
	855	16	5.4	150357	17	AC004738	*** SEQUENCING IN PROG 3.62e+02
	856	16	5.4	151018	17	PFMAL1P6	Plasmodium falciparum 3.62e+02
	857	16	5.4	151798	26	HS435C23	Human DNA sequence fro 3.62e+02
	858	16	5.4	152184	17	HS593F22	Human DNA sequence *** 3.62e+02

	859	16	5.4	152580	17	HS366L4	Human DNA sequence *** 3.62e+02
	860	16	5.4	153672	17	AC004917	*** SEQUENCING IN PROG 3.62e+02
	861	16	5.4	153692	17	CEY26D4	Caenorhabditis elegans 3.62e+02
	862	16	5.4	153835	17	CEY6D1	Caenorhabditis elegans 3.62e+02
	863	16	5.4	153892	17	HS1121J18	Human DNA sequence *** 3.62e+02
	864	16	5.4	154252	26	AC004062	Homo sapiens chromosom 3.62e+02
	865	16	5.4	154788	27	HS325F22	Human DNA sequence fro 3.62e+02
c	866	16	5.4	154788	27	HS325P22	Human DNA sequence fro 3.62e+02
	867	16	5.4	154985	18	AC004641	Drosophila melanogaste 3.62e+02
	868	16	5.4	155382	27	AC004911	Homo sapiens PAC clone 3.62e+02
	869	16	5.4	155407	26	AF042090	Homo sapiens chromosom 3.62e+02
	870	16	5.4	156214	26	HSAC002064	Human BAC clone RG016J 3.62e+02
	871	16	5.4	156601	26	HS212G6	Homo sapiens DNA seque 3.62e+02
	872	16	5.4	158078	27	AC005094	Homo sapiens BAC clone 3.62e+02
c	873	16	5.4	158078	27	AC005094	Homo sapiens BAC clone 3.62e+02
c	874	16	5.4	158427	26	AC002553	Homo sapiens chromosom 3.62e+02
c	875	16	5.4	158886	26	AC004674	Homo sapiens Xp22 BAC 3.62e+02
	876	16	5.4	159387	26	HS455H14	Homo sapiens DNA seque 3.62e+02
	877	16	5.4	159500	28	AC005742	Mus musculus chromosom 3.62e+02
c	878	16	5.4	159712	17	AC002327	*** SEQUENCING IN PROG 3.62e+02
c	879	16	5.4	160203	26	HS425C14	Human DNA sequence fro 3.62e+02
c	880	16	5.4	160526	17	HS1033E15	Human DNA sequence *** 3.62e+02
c	881	16	5.4	160986	17	AC005505	*** SEQUENCING IN PROG 3.62e+02
	882	16	5.4	161453	27	AC004993	Homo sapiens PAC clone 3.62e+02
c	883	16	5.4	161453	27	AC004993	Homo sapiens PAC clone 3.62e+02
	884	16	5.4	162524	17	HS498I24	Human DNA sequence *** 3.62e+02
	885	16	5.4	162531	26	AC005249	Homo sapiens BAC clone 3.62e+02
	886	16	5.4	162822	27	AC005751	Homo sapiens chromosom 3.62e+02
	887	16	5.4	163889	17	AC004958	*** SEQUENCING IN PROG 3.62e+02
c	888	16	5.4	163974	27	AB011399	Homo sapiens gene for 3.62e+02
c	889	16	5.4	164538	27	AC005774	Homo sapiens chromosom 3.62e+02
	890	16	5.4	165209	17	HS209B9	Human DNA sequence *** 3.62e+02
c	891	16	5.4	165209	17	HS209B9	Human DNA sequence *** 3.62e+02
	892	16	5.4	165228	26	AC005325	Homo sapiens chromosom 3.62e+02
c	893	16	5.4	165608	26	AC004492	Homo sapiens BAC clone 3.62e+02
	894	16	5.4	167343	26	HSU96629	Human chromosome 8 BAC 3.62e+02
c	895	16	5.4	169027	17	AC002370	*** SEQUENCING IN PROG 3.62e+02
c	896	16	5.4	170000	17	AC004066	*** SEQUENCING IN PROG 3.62e+02
c	897	16	5.4	170054	17	HS460J8	Human DNA sequence *** 3.62e+02
c	898	16	5.4	172241	26	AC004014	Human BAC clone RG022K 3.62e+02
	899	16	5.4	173766	26	AC004049	Homo sapiens chromosom 3.62e+02
	900	16	5.4	173767	26	AC002449	Human PAC clone DJ404K 3.62e+02
c	901	16	5.4	174285	17	HS524E15	Human DNA sequence *** 3.62e+02
c	902	16	5.4	175035	17	AC005553	*** SEQUENCING IN PROG 3.62e+02
	903	16	5.4	175162	17	AC005182	*** SEQUENCING IN PROG 3.62e+02
	904	16	5.4	175488	17	AC004887	*** SEQUENCING IN PROG 3.62e+02
	905	16	5.4	176781	17	AC002989	*** SEQUENCING IN PROG 3.62e+02
c	906	16	5.4	177137	17	AC005089	*** SEQUENCING IN PROG 3.62e+02
c	907	16	5.4	177163	26	HS134E15	Human DNA sequence *** 3.62e+02
c	908	16	5.4	177189	17	HS428A13	*** SEQUENCING IN PROG 3.62e+02
c	909	16	5.4	178440	17	CEY38F1	Caenorhabditis elegans 3.62e+02

c	910	16	5.4	179245	26	HUAC002492	Human Chromosome 16 BA 3.62e+02
c	911	16	5.4	179319	17	AC005382	*** SEQUENCING IN PROG 3.62e+02
c	912	16	5.4	179544	17	AC005720	Homo sapiens chromosom 3.62e+02
	913	16	5.4	180388	25	HUMRETBLAS	Human retinoblastoma s 3.62e+02
c	914	16	5.4	180388	25	HUMRETBLAS	Human retinoblastoma s 3.62e+02
	915	16	5.4	180551	26	HUAC004158	Homo sapiens Chromosom 3.62e+02
	916	16	5.4	182329	17	AC005022	*** SEQUENCING IN PROG 3.62e+02
c	917	16	5.4	184919	27	AF069291	Homo sapiens BAC clone 3.62e+02
	918	16	5.4	185364	17	AC005477	*** SEQUENCING IN PROG 3.62e+02
c	919	16	5.4	186306	17	CEY6E2	Caenorhabditis elegans 3.62e+02
	920	16	5.4	186510	26	HS451B15	Human DNA sequence fro 3.62e+02
c	921	16	5.4	186550	17	CEY44F5	Caenorhabditis elegans 3.62e+02
	922	16	5.4	187816	17	CEY32F6	Caenorhabditis elegans 3.62e+02
	923	16	5.4	188205	27	AC003986	Homo sapiens BAC clone 3.62e+02
	924	16	5.4	188477	17	AC004971	*** SEQUENCING IN PROG 3.62e+02
c	925	16	5.4	188741	26	AC002542	Human BAC clone RG114A 3.62e+02
c	926	16	5.4	189215	17	AC005675	*** SEQUENCING IN PROG 3.62e+02
c	927	16	5.4	189357	17	HS596H12	Human DNA sequence *** 3.62e+02
	928	16	5.4	190000	17	AC004168	*** SEQUENCING IN PROG 3.62e+02
c	929	16	5.4	190248	17	AC004988	*** SEQUENCING IN PROG 3.62e+02
c	930	16	5.4	190907	26	AC005261	Homo sapiens chromosom 3.62e+02
c	931	16	5.4	191233	17	AC005409	*** SEQUENCING IN PROG 3.62e+02
	932	16	5.4	191908	17	CEY51B9	Caenorhabditis elegans 3.62e+02
c	933	16	5.4	193374	17	AC005139	*** SEQUENCING IN PROG 3.62e+02
c	934	16	5.4	195033	17	HS608E8	Human DNA sequence *** 3.62e+02
c	935	16	5.4	196475	17	AC005386	*** SEQUENCING IN PROG 3.62e+02
c	936	16	5.4	196542	27	AC002365	Homo sapiens chromosom 3.62e+02
c	937	16	5.4	197273	26	AC002461	Human BAC clone RG204I 3.62e+02
c	938	16	5.4	197391	17	AC005140	*** SEQUENCING IN PROG 3.62e+02
c	939	16	5.4	197620	26	AF001905	Homo sapiens cosmids E 3.62e+02
	940	16	5.4	198499	17	AC004870	*** SEQUENCING IN PROG 3.62e+02
	941	16	5.4	199496	17	HS357K22	Human DNA sequence *** 3.62e+02
c	942	16	5.4	200000	17	AC005510	*** SEQUENCING IN PROG 3.62e+02
	943	16	5.4	200811	23	ATFCA0	Arabidopsis thaliana D 3.62e+02
	944	16	5.4	204082	27	AC005323	Homo sapiens chromosom 3.62e+02
	945	16	5.4	207139	17	CEY17G7	Caenorhabditis elegans 3.62e+02
	946	16	5.4	208438	17	AC005034	*** SEQUENCING IN PROG 3.62e+02
c	947	16	5.4	209165	23	ATFCA5	Arabidopsis thaliana D 3.62e+02
	948	16	5.4	210964	17	AC005497	*** SEQUENCING IN PROG 3.62e+02
c	949	16	5.4	212610	15	BSUB0019	Bacillus subtilis comp 3.62e+02
	950	16	5.4	213541	27	HUAC004381	Homo sapiens Chromosom 3.62e+02
	951	16	5.4	220731	17	AC004671	*** SEQUENCING IN PROG 3.62e+02
c	952	16	5.4	220731	17	AC004671	*** SEQUENCING IN PROG 3.62e+02
	953	16	5.4	221273	17	CEY40H4	Caenorhabditis elegans 3.62e+02
c	954	16	5.4	221273	17	CEY40H4	Caenorhabditis elegans 3.62e+02
c	955	16	5.4	222922	17	AC005015	*** SEQUENCING IN PROG 3.62e+02
	956	16	5.4	222930	26	HSU47924	Human chromosome 12p13 3.62e+02
	957	16	5.4	225432	26	AF027390	Homo sapiens 7q telome 3.62e+02
	958	16	5.4	227538	28	AC002397	Mouse chromosome 6 BAC 3.62e+02
c	959	16	5.4	228500	17	HS212A2	Human DNA sequence *** 3.62e+02
	960	16	5.4	228516	26	AC002523	Homo sapiens Xq28 BACs 3.62e+02

c	961	16	5.4	236157	17	AC005703	*** SEQUENCING IN PROG 3.62e+02
c	962	16	5.4	236822	26	D84394	Homo sapiens genomic D 3.62e+02
	963	16	5.4	242825	26	HSPEX	H.sapiens PEX gene. 3.62e+02
c	964	16	5.4	244362	17	HS618F1	Human DNA sequence *** 3.62e+02
c	965	16	5.4	244455	17	AC002470	*** SEQUENCING IN PROG 3.62e+02
	966	16	5.4	246282	26	HSU91328	Human hereditary haemo 3.62e+02
c	967	16	5.4	250448	17	CEY36E3	Caenorhabditis elegans 3.62e+02
	968	16	5.4	250685	17	CEY62F10	Caenorhabditis elegans 3.62e+02
	969	16	5.4	257000	16	AP000002	Pyrococcus horikoshii 3.62e+02
	970	16	5.4	259202	26	AC002366	Human Xp22 BAC CT-2851 3.62e+02
c	971	16	5.4	267156	26	U66059	Human germline T-cell 3.62e+02
	972	16	5.4	269078	17	CEY76A2	Caenorhabditis elegans 3.62e+02
c	973	16	5.4	287986	17	CEY51H4	Caenorhabditis elegans 3.62e+02
	974	16	5.4	296699	17	CEY48A6	Caenorhabditis elegans 3.62e+02
c	975	16	5.4	302988	17	CEY46G5	Caenorhabditis elegans 3.62e+02
c	976	16	5.4	323000	26	HS229041	Homo sapiens 959 kb co 3.62e+02
c	977	16	5.4	338234	18	DMU31961	Drosophila melanogaste 3.62e+02
c	978	16	5.4	338534	15	ECOUW93	Escherichia coli K-12 3.62e+02
	979	16	5.4	348050	26	HS229042	Homo sapiens 959 kb co 3.62e+02
c	980	15	5.1	454	33	XCU70889	Xestia c-nigrum granul 1.29e+03
c	981	15	5.1	967	33	VACITR	Vaccinia virus inverte 1.29e+03
c	982	15	5.1	982	33	VVU23801	Variola virus strain S 1.29e+03
c	983	15	5.1	1173	15	BSU42331	Bacillus sphaericus 23 1.29e+03
c	984	15	5.1	2870	25	HSU19603	Human gonadotropin-rel 1.29e+03
c	985	15	5.1	3468	18	DMU47849	Drosophila melanogaste 1.29e+03
c	986	15	5.1	6335	26	AB011087	Homo sapiens mRNA for 1.29e+03
	987	15	5.1	11807	26	D86990	Human (lambda) DNA fo 1.29e+03
	988	15	5.1	19712	15	D78509	Bacillus subtilis YfjG 1.29e+03
c	989	15	5.1	29539	33	VVU18340	Variola virus Somalia- 1.29e+03
c	990	15	5.1	40685	26	AC002052	Homo sapiens Chromosom 1.29e+03
c	991	15	5.1	101167	26	HS32F7	Homo sapiens DNA seque 1.29e+03
c	992	15	5.1	128915	26	AC002418	Human Chromosome X, co 1.29e+03
c	993	15	5.1	149140	26	HS256G22	Homo sapiens DNA seque 1.29e+03
c	994	15	5.1	153356	17	AC004995	*** SEQUENCING IN PROG 1.29e+03
	995	15	5.1	172108	17	HS3100I3	Human DNA sequence *** 1.29e+03
c	996	15	5.1	173140	17	AC004867	*** SEQUENCING IN PROG 1.29e+03
c	997	15	5.1	177311	26	AC004385	Human Chromosome X clo 1.29e+03
c	998	15	5.1	177923	33	U94848	Vaccinia virus strain 1.29e+03
c	999	15	5.1	191737	33	VACCG	Vaccinia virus, comple 1.29e+03
	1000	15	5.1	328808	17	CEY44A6	Caenorhabditis elegans 1.29e+03

ALIGNMENTS

```
RESULT    1
LOCUS       HS11703     153617 bp     DNA              HTG       23-SEP-1998
DEFINITION  Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone 11703;
            HTGS phase 1.
ACCESSION   AL020995
NID         g3421021
KEYWORDS    HTG; HTGS_PHASE1.
```



```
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 153617)
  AUTHORS   McDougall,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-SEP-1998) Wellcome Trust Genome Campus, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquires:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 18, 1998 this sequence version replaced gi:2689208.
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and the release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc. Order of segments is not known; 800 n's separate
            segments. Unfinished: dJ11703  Contig_ID: 02265  acc=AL020995
            Length: 62095 bp Unfinished: dJ11703  Contig_ID: 02899
            acc=AL020995  Length: 1255 bp Unfinished: dJ11703  Contig_ID: 02974
             acc=AL020995  Length: 88667 bp.
            ***                                                          ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                          ***
            * This sequence is unfinished. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
FEATURES              Location/Qualifiers
     source           1. .153617
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="1"
                      /clone="11703"
BASE COUNT    40391 a  35449 c  35412 g  40765 t   1600 others
ORIGIN

  Query Match           7.4%;  Score 22;  DB 17;  Length 153617;
  Best Local Similarity 100.0%;  Pred. No. 5.37e-02;
  Matches    22;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  41936 AAATAAAGTTATTTAAATAAAA 41957
          ||||||||||||||||||||||
Cp    133 AAATAAAGTTATTTAAATAAAA 112


RESULT    2
LOCUS       AC005116    121434 bp     DNA              HTG       26-SEP-1998
DEFINITION  *** SEQUENCING IN PROGRESS *** DS01831; D311; Drosophila
            melanogaster; Chromosome 2L; Region 32D1-32D3; HTGS phase 1, 42
            unordered pieces.
ACCESSION   AC005116
```



```
NID         g3659472
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea;
            Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 121434)
  AUTHORS   Celniker,S.E., George,R.A., Galle,R., Svirskas,R.R., Hoskins,R.A.,
            Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Chavez,C.,
            Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A.,
            Hummasti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B.,
            Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S., Moshrefi,A.R.,
            Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,E.,
            Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R.,
            Zieran,L.L. and Kimmel,B.E.
  TITLE     Sequencing of Drosophila melanogaster
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 121434)
  AUTHORS   Celniker,S.E., George,R.A., Galle,R., Svirskas,R.R., Hoskins,R.A.,
            Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Chavez,C.,
            Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A.,
            Hummasti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B.,
            Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S., Moshrefi,A.R.,
            Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,E.,
            Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R.,
            Zieran,L.L. and Kimmel,B.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUN-1998) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Sep 26, 1998 this sequence version replaced gi:3419773.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://fruitfly.berkeley.edu/sequence/) or send
            email to drosophila@mhgc.lbl.gov. All contigs in this submission
            meet the following cutoffs: length >= 400 bases, phrap computed
            error rate <= 1/10.
            ***                                                          ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                          ***
            * This sequence is unfinished. It consists of 42 contigs for
            * which the order is not known; their order in this record is
            * arbitrary. In some cases, the exact lengths of the gaps
            * between the contigs are also unknown; these gaps are  presented
            * as runs of N as a convenience only. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
            *       1      998: contig of 998 bp in length
            *     999     1078: gap of unknown length
            *    1079     2061: contig of 983 bp in length
            *    2062     2141: gap of unknown length
```



```
            *    2142     2983: contig of 842 bp in length
            *    2984     3063: gap of unknown length
            *    3064     3792: contig of 729 bp in length
            *    3793     3872: gap of unknown length
            *    3873     4753: contig of 881 bp in length
            *    4754     4833: gap of unknown length
            *    4834     5477: contig of 644 bp in length
            *    5478     5557: gap of unknown length
            *    5558     6433: contig of 876 bp in length
            *    6434     6513: gap of unknown length
            *    6514     7509: contig of 996 bp in length
            *    7510     7589: gap of unknown length
            *    7590     8557: contig of 968 bp in length
            *    8558     8637: gap of unknown length
            *    8638     9312: contig of 675 bp in length
            *    9313     9392: gap of unknown length
            *    9393    10345: contig of 953 bp in length
            *   10346    10425: gap of unknown length
            *   10426    11407: contig of 982 bp in length
            *   11408    11487: gap of unknown length
            *   11488    12375: contig of 888 bp in length
            *   12376    12455: gap of unknown length
            *   12456    13308: contig of 853 bp in length
            *   13309    13388: gap of unknown length
            *   13389    14438: contig of 1050 bp in length
            *   14439    14518: gap of unknown length
            *   14519    15630: contig of 1112 bp in length
            *   15631    15710: gap of unknown length
            *   15711    17271: contig of 1561 bp in length
            *   17272    17351: gap of unknown length
            *   17352    18686: contig of 1335 bp in length
            *   18687    18766: gap of unknown length
            *   18767    20084: contig of 1318 bp in length
            *   20085    20164: gap of unknown length
            *   20165    21518: contig of 1354 bp in length
            *   21519    21598: gap of unknown length
            *   21599    22766: contig of 1168 bp in length
            *   22767    22846: gap of unknown length
            *   22847    23959: contig of 1113 bp in length
            *   23960    24039: gap of unknown length
            *   24040    25376: contig of 1337 bp in length
            *   25377    25456: gap of unknown length
            *   25457    26957: contig of 1501 bp in length
            *   26958    27037: gap of unknown length
            *   27038    27822: contig of 785 bp in length
            *   27823    27902: gap of unknown length
            *   27903    29569: contig of 1667 bp in length
            *   29570    29649: gap of unknown length
            *   29650    31507: contig of 1858 bp in length
            *   31508    31587: gap of unknown length
            *   31588    33451: contig of 1864 bp in length
```

```
             *    33452   33531: gap of unknown length
             *    33532   35347: contig of 1816 bp in length
             *    35348   35427: gap of unknown length
             *    35428   36949: contig of 1522 bp in length
             *    36950   37029: gap of unknown length
             *    37030   39239: contig of 2210 bp in length
             *    39240   39319: gap of unknown length
             *    39320   41015: contig of 1696 bp in length
             *    41016   41095: gap of unknown length
             *    41096   43237: contig of 2142 bp in length
             *    43238   43317: gap of unknown length
             *    43318   45455: contig of 2138 bp in length
             *    45456   45535: gap of unknown length
             *    45536   48197: contig of 2662 bp in length
             *    48198   48277: gap of unknown length
             *    48278   51254: contig of 2977 bp in length
             *    51255   51334: gap of unknown length
             *    51335   56024: contig of 4690 bp in length
             *    56025   56104: gap of unknown length
             *    56105   63712: contig of 7608 bp in length
             *    63713   63792: gap of unknown length
             *    63793   71752: contig of 7960 bp in length
             *    71753   71832: gap of unknown length
             *    71833   86505: contig of 14673 bp in length
             *    86506   86585: gap of unknown length
             *    86586   98210: contig of 11625 bp in length
             *    98211   98290: gap of unknown length
             *    98291  121434: contig of 23144 bp in length.
FEATURES             Location/Qualifiers
     source          1. .121434
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="DS01831; DJ11"
                     /chromosome="2"
                     /map="32D1-32D3"
BASE COUNT    35466 a  23163 c  23109 g  36409 t   3287 others
ORIGIN

  Query Match           7.1%;  Score 21;  DB 17;  Length 121434;
  Best Local Similarity 100.0%;  Pred. No. 2.62e-01;
  Matches   21;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  77374 AAATAACTTTATTTATAACTT 77394
          |||||||||||||||||||||
Qy    120 AAATAACTTTATTTATAACTT 140


RESULT    3
LOCUS       SPAC22E12   37172 bp    DNA             PLN       09-MAR-1996
DEFINITION  S.pombe chromosome I cosmid c22E12.
ACCESSION   Z70043
```



```
NID         g1220275
KEYWORDS    krp1; LTR; membrane protein; protease; protein kinase; ribosomal
            protein; RNA 1; RNA-binding protein; tf2.
SOURCE      fission yeast.
  ORGANISM  Schizosaccharomyces pombe
            Eukaryotae; mitochondrial eukaryotes; Fungi; Ascomycota;
            Archaeascomycetes; Schizosaccharomycetales;
            Schizosaccharomycetaceae; Schizosaccharomyces.
REFERENCE   1  (bases 1 to 37172)
  AUTHORS   Devlin,K. and Churcher,C.M.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 37172)
  AUTHORS   Barrell,B.G., Rajandream,M.A. and Walsh,S.V.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-1996) Schizosaccharomyces pombe chromosome I
            sequencing project, Sanger Centre, Hinxton Hall, Hinxton, Cambridge
            CB10 1RQ E-mail: barrell@sanger.ac.uk
COMMENT     Notes:
            Details of yeast sequencing at the Sanger Centre are available on
            the World Wide Web.
            (URL, http://www.sanger.ac.uk/yeast/home.html)
            Protein coding regions (CDS) have been predicted with the help of
            computer analysis using the Genefinder program in PomBase (an ACEDB
            database) with additional predictions for the branch-acceptor sites
            supplied by the program Sp3splice. CAUTION: It is possible that for
            any individual CDS we may have underestimated or overestimated the
            number of introns/exons or we may not have chosen the correct
            splice donor/acceptor sites.
            CDS are numbered using the following system eg SPAC5H10.01c. SP (S.
            pombe), A (chromosome 1), c5H10 (cosmid name), .01 (first CDS), c
            (complementary strand).
            The more significant matches with motifs in the PROSITE database
            are also included but some of these may be fortuitous.
            The length in codons is given for each CDS.
            IMPORTANT: This sequence MAY NOT be the entire insert of the
            sequenced clone.  It may be shorter because we only sequence
            overlapping sections once, or longer, because we arrange for a
            small overlap between neighbouring submissions.
            Cosmid c22E12 is overlapped at the 3' end by cosmid c22E12.
FEATURES             Location/Qualifiers
     source          1. .37172
                     /organism="Schizosaccharomyces pombe"
                     /strain="972h-"
                     /db_xref="taxon:4896"
                     /chromosome="I"
                     /clone="cosmid c22E12"
                     /map="IR"
     gene            1. .441
                     /gene="SPAC22E12.01"
     CDS             <1. .441
                     /gene="SPAC22E12.01"
```



```
                     /note="SPAC22E12.01, unknown, partial orf, len: > 146, low
                     similarity with PIR:S49741 hypothetical protein (29.1%
                     identity in 79 aa overlap)"
                     /codon_start=1
                     /product="unknown"
                     /db_xref="PID:e227661"
                     /db_xref="PID:g1220276"
                     /translation="DHPWTSNPPTSLFALTPLMFLFLLVAGLIFEGPVRFIESPAWKE
                     PGPPMSVVILVPGTLAPFMVASEPGLIQKTSIVTLSVCGILKEIITIIASTLFYHDIL
                     LPINIVGLVITLCGIGVYNYYRITKGNKKEAEKEVEYIVLWENA"
     LTR             1334. .1682
                     /gene="tf2"
                     /note="solo tf2 ltr element"
     gene            1334. .1682
                     /gene="tf2"
     gene            1958. .2851
                     /gene="SPAC22E12.02"
     CDS             join(1958. .2265,2500. .2851)
                     /gene="SPAC22E12.02"
                     /note="SPAC22E12.02, unknown, len: : 219, similarity with
                     RNA-binding proteins,: 219, contains PS00030 Eukaryotic
                     putative RNA-binding region RNP-1 signature"
                     /codon_start=1
                     /product="unknown"
                     /db_xref="PID:e227662"
                     /db_xref="PID:g1220277"
                     /translation="MSRVKTTVVRKAGGQVWEDPTLLEWDPNHFRLFVGNLGNDVNDE
                     SLYQAFSEYPSLVKTKVVRDREGKTRGFGFVSFKDSDQFLKAWREKNGKYIGSRPVKL
                     SRATSDVKPNEVNEKTIDSKIRNSLYSRTIHHGNRVQKKIKNKHGKNSSKSSRAAQSA
                     AAELISSSSITGARPANSTSVPNAVNTEISAARATESEASRNQTKASRDYSRASSFRR
                     V"
     misc_feature    2165. .2188
                     /gene="SPAC22E12.02"
                     /note="PS00030 Eukaryotic putative RNA-binding region
                     RNP-1 signature"
     misc_feature    2266. .2271
                     /gene="SPAC22E12.02"
                     /note="splice donor sequence, gtatgt"
     misc_feature    2480. .2499
                     /gene="SPAC22E12.02"
                     /note="splice branch and acceptor sequence,
                     ctaacatgtagttttctag"
     gene            complement(2853. .3526)
                     /gene="SPAC22E12.03c"
     CDS             complement(join(2853. .2949,3048. .3526))
                     /gene="SPAC22E12.03c"
                     /note="SPAC22E12.03c, unknown, len: 191"
                     /codon_start=1
                     /product="unknown"
                     /db_xref="PID:e228197"
                     /db_xref="PID:g1220278"
```



```
                     /translation="MVKVCLFVADGTDEIEFSAPWGIFKRAEIPIDSVYVGENKDRLV
                     KMSRDVEMYANRSYKEIPSADDFAKQYDIAIIPGGGLGAKTLSTTPFVQQVVKEFYKK
                     PNKWIGMICAGTLTAKTSGLPNKQITGHPSVRGQLEEGGYKYLDQPVVLEENLITSQG
                     PGTAMLFGLKLLEQVASKDKTNAVYKSLSKP"
     misc_feature    complement(2950. .2969)
                     /gene="SPAC22E12.03c"
                     /note="splice branch and acceptor sequence,
                     ctaacatgttcttttcattag"
     misc_feature    complement(3042. .3047)
                     /gene="SPAC22E12.03c"
                     /note="splice donor sequence, gtaagt"
     gene            4710. .5810
                     /gene="SPAC22E12.04"
     CDS             join(4710. .4718,4771. .4897,5053. .5810)
                     /gene="SPAC22E12.04"
                     /note="SPAC22E12.04, unknown, len: 297"
                     /codon_start=1
                     /product="unknown"
                     /db_xref="PID:e227663"
                     /db_xref="PID:g1220279"
                     /translation="MFEVEYLIKDCDDVNKNTLEQEFQDLNIEDWKWDAATGQLIVKG
                     SVSPSKVLRRLENATSKPILIRGASNKESGVSILYEANEDITQIPKVYGLCRFIPTEE
                     KIFLDLIATQLLPNRETYGLVTISGDISRGLKSAGDSLVTLFNANSNEQGKIVLDKEV
                     SGSLPNWIGHCFVLKCVDDSDSATMGIISRSAGLGQNTKQICACTGKSLWTEHAELKS
                     VNEGSSCCSKKDSSPSEKPSCCSQEKKSCCSSKKPSCCSQEKKGCCSTEKTSCCSQEK
                     KSCCTSEKPSCCSNGKSTVCA"
     misc_feature    4719. .4724
                     /gene="SPAC22E12.04"
                     /note="splice donor sequence, gtatgg"
     misc_feature    4755. .4770
                     /gene="SPAC22E12.04"
                     /note="splice branch and acceptor sequence,
                     ctaataaaatttttag"
     misc_feature    4898. .4903
                     /gene="SPAC22E12.04"
                     /note="splice donor sequence, gtaagt"
     misc_feature    5036. .5052
                     /gene="SPAC22E12.04"
                     /note="splice branch and acceptor sequence,
                     ctaacgtagatttttcag"
     gene            complement(6574. .7181)
                     /gene="SPAC22E12.05c"
     CDS             complement(join(6574. .7068,7122. .7181))
                     /gene="SPAC22E12.05c"
                     /note="SPAC22E12.05c, possible transmembrane protein, len:
                     184, similar to SW:RER1_YEAST P25560 RER1 protein
                     PIR:S19346 (56.8% identity in 162 aa overlap)"
                     /codon_start=1
                     /product="unknown"
                     /db_xref="PID:e227664"
                     /db_xref="PID:g1220280"
```

```
                     /translation="MEFIQRHIENVKEKKNFAVRLYRHWVDRTIPYTTYRWLTVSGLI
                     ALPFIRILLVRGWYIVCYTLAIYLLNLFLAFLTPKFDPSVEQAMKDEEIEEGVLPTSK
                     DDEFRPPIRRLPEPKFWYSSMRATLFALVASFFRIFDVPVFWPILVVYYLVLSFFCFR
                     RQIQHMLKYRYVPFDIGKKKFGSH"
     misc_feature    complement(7069. .7083)
                     /gene="SPAC22E12.05c"
                     /note="splice branch and acceptor sequence,
                     ctaacgtgttttag"
     misc_feature    complement(7116. .7121)
                     /gene="SPAC22E12.05c"
                     /note="splice donor sequence, gtaaga"
     gene            complement(8774. .9772)
                     /gene="SPAC22E12.06c"
     CDS             complement(8774. .9772)
                     /gene="SPAC22E12.06c"
                     /note="SPAC22E12.06c, unknown, len: 332, similar to
                     SW:YA0B_SCHPO Q09679 hypothetical 37.6 KD protein
                     SPAC5H10.11 (47.9% identity in 317 aa overlap) and
                     SW:YA0D_SCHPO Q09681 hypothetical 39.1 KD protein
                     SPAC5H10.13C (43.8% identity in 338 aa overlap)"
                     /codon_start=1
                     /product="unknown"
                     /db_xref="PID:e228198"
                     /db_xref="PID:g1220281"
                     /translation="MAVLQWKLSRVIFVLLIIGISSFLIYFQPGNSFSASPALERLQD
                     VFSKPSVQEYDYLVPLDEDMQGLSTTYHQATINEDDPQEPRSHEIVILLASDGRSSGN
                     MAPETFNQCIENRINYAKHHNYGFEYVNVSQMNIPPVWAKMPAIIQTMNKEPHAKWIW
                     WLDQDALILNTELSIQEHILSPDVLVEKLWKENEPMISPFSADLERLTPSSYTVDSARS
                     LGLLISQDLNGLNAGSFFVRRSPMMALFLDLWGDKSFRENKVADHEQALLGTFVRYHP
                     EIAAIIGILPQTLINSYPVGQPEMGWKEGHLVIHFAGCWVENRCQVLWDEYRERVH"
     gene            11558. .12718
                     /gene="rna1"
     CDS             11558. .12718
                     /gene="rna1"
                     /note="SPAC22E12.07, rna1 gene, len: 386, SW:RNA1_SCHPO
                     P41391 rna11 protein, PIR:S37691"
                     /codon_start=1
                     /product="Rna1"
                     /db_xref="PID:e227665"
                     /db_xref="PID:g1220282"
                     /translation="MSRFSIEGKSLKLDAITTEDEKSVFAVLLEDDSVKEIVLSGNTI
                     GTEAARWLSENIASKKDLEIAEFSDIPTGRVKDEIPEALRLLLQALLKCPKLHTVRLS
                     DNAPGPTAQEPLIDFLSKHTPLEHLYLHNNGLGPQAGAKIARALQELAVNKKAKNAPP
                     LRSIICGRNRLENGSMKEWAKTPQSHRLLHTVKMVQNGIRPEGIEHLLLEGLAYCQEL
...
Note: remainder of annotations omitted.

  Query Match           6.8%;  Score 20;  DB 23;  Length 37172;
  Best Local Similarity 100.0%;  Pred. No. 1.23e+00;
  Matches   20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```



```
Db  28992 AAAATATTTTTAATGCATTA 29011
          ||||||||||||||||||||
Cp    114 AAAATATTTTTAATGCATTA 95


RESULT    4
LOCUS       L49405      81729 bp    DNA                 INV       17-JUL-1998
DEFINITION  Drosophila melanogaster DNA sequence (P1 DS08249 (D25)), complete
            sequence.
ACCESSION   L49405 L36274 L36273 L39707 L39703 L36276 L36269 L39706 L36282
            L36275 L36268 L36283 L39705 L36271 L36278 L36279 L36280 L36270
            L39702 L39704 L39709 L39708 L39701 L36265 L36266 L36281 L36267
            L36277 L36272
NID         g2822113
KEYWORDS    HTG.
SOURCE      Drosophila melanogaster (Subclones in sac from P1 clone DS08249
            (D25)) DNA.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea;
            Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 81729)
  AUTHORS   Celniker,S.E., George,R.A., Galle,R.F., Hoskins,R.A.,
            Svirskas,R.R., Harris,N.L., Agbayani,A., Arcaina,T.T., Baxter,E.,
            Blazej,R.G., Chavez,C., Chew,M., Doyle,C.M., Farfan,D.E.,
            Flanagan,J., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,S.H., Lee,B., Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Punch,E., Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R.,
            Yee,A., Zhang,R., Zieran,L.L. and Kimmel,B.
  TITLE     Sequencing of Alcohol dehydrogenase region
  JOURNAL   Unpublished (1997)
REFERENCE   2  (bases 1 to 81729)
  AUTHORS   Martin,C.H., Alcivare,D.A., Arcaina,T.T., Bondoc,M.M., Chiang,A.,
            Critz,P.A., Davis,C.A., Doyle,C.M., Ericsson,C.L., Farfan,D.E.,
            Gunning,K.M., Houston,K.A., Hummasti,S.R., Jaklevic,M.A.,
            Kadner,K.E., Karra,K., Kim,K., Kim,S.F., Ko,C.L., Lewis,K.D.,
            Li,M., Lindquist,K.J., Lomotan,M.A., Lustre,V.M., Machrus,M.U.,
            Mayeda,C.A., Mazda,P., Miguel,T.M., Miller,C.A., Mok,M.S.,
            Pacleb,J.M., Patel,S.G., Salva,A., Santos,R.F., Snir,E.,
            Subramanian,S., Wan,K.H., Whitelaw,K.R., Yee,A., Yeh,R.T., Yu,C.,
            Zieran,L.L. and Palazzolo,M.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-AUG-1997) Berkeley Drosophila Genome Project, MS
            74-157, Lawrence Berkeley National Laboratory, One Cyclotron Road,
            Berkeley, CA 94720, US
COMMENT     Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            For further information about this sequence, including its location
```



```
            and relationship to other sequences, please visit our sequence
            archive Web site (http://fruitfly.berkeley.edu/sequence/) or send
            email to drosophila@mhgc.lbl.gov.
            Library location: 90-86.
FEATURES                Location/Qualifiers
     source             1. .81729
                        /organism="Drosophila melanogaster"
                        /db_xref="taxon:7227"
                        /chromosome="2L"
                        /map="34D1-34D2"
                        /clone="P1 DS08249 (D25)"
                        /note="This sequence has not changed since its original
                        submission.  It was resubmitted in order to include all
                        secondary accession numbers for the subclones belonging to
                        this clone."
BASE COUNT    24230 a  16948 c  16975 g  23576 t
ORIGIN

  Query Match           6.8%;  Score 20;  DB 18;  Length 81729;
  Best Local Similarity 100.0%;  Pred. No. 1.23e+00;
  Matches   20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  48840 ATTAAAAATATTTTTATTTA 48859
          ||||||||||||||||||||
Qy    101 ATTAAAAATATTTTTATTTA 120


RESULT    5
LOCUS       AC004744    85579 bp    DNA             PRI       23-MAY-1998
DEFINITION  Homo sapiens BAC clone GS465N13 from 7p15-p21, complete sequence.
ACCESSION   AC004744
NID         g3152628
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 85579)
  AUTHORS   Stoneking,T., Rohlfing,T., Langston,Y. and Ozersky,P.
  TITLE     The sequence of Homo sapiens BAC clone GS465N13
  JOURNAL   Unpublished (1998)
REFERENCE   2  (bases 1 to 85579)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-MAY-1998) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     SUBMITTED BY:  WUGSC
            Genome Sequencing Center
            Department of Genetics
            Washington University
            St. Louis MO 63108, USA
```



```
            http://genome.wustl.edu/gsc
            mailto:sapiens@watson.wustl.edu

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded or sequenced with an alternate
            chemistry; an attempt was made to resolve all sequencing problems,
            such as compressions and repeats; all regions were covered by
            sequence from more than one subclone; and the assembly was
            confirmed by restriction digest.

            MAPPING INFORMATION:
            The sequence of this clone was established as part of a mapping and
            sequencing collaboration between the NHGRI Chromosome 7 Mapping
            Project (Eric D. Green, Director), John D. McPherson in the
            Department of Genetics (Washington University), and the Washington
            University Genome Sequencing Center.  For additional information
            about the map position of this sequence, see
            http://www.nhgri.nih.gov/DIR/GTB/CHR7 , send
            mailto:egreen@nhgri.nih.gov , or see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            This clone is from the first BAC library from Genome Systems, Inc.
            (http://www.genomesystems.com).
            Cell line:  lymphoblastoid
            Haplotypes:  two
            VECTOR:  pBeloBAC
            Selection:  chloramphenicol
            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is DJ1194E15; the clone sequenced
            to the right is RG370M10, 200 bp overlap.  Actual start of this
            clone is at base position 1 of GS465N13; actual end is at 85579 of
            GS465N13.
FEATURES                Location/Qualifiers
     source             1. .85579
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="7"
                        /map="7p15-p21"
                        /clone_lib="GSBAC1"
                        /clone="GS465N13"
     repeat_region      1. .5376
                        /rpt_family="L1"
     repeat_region      5392. .5669
                        /rpt_family="L2"
     repeat_region      5748. .5814
                        /rpt_family="MER1_type?"
```

```
     repeat_region   5880. .6080
                     /rpt_family="L2"
     repeat_region   6106. .6966
                     /rpt_family="L1"
     repeat_region   7771. .8073
                     /rpt_family="Alu"
     repeat_region   10130. .11291
                     /rpt_family="L1"
     repeat_region   11491. .11701
                     /rpt_family="MIR"
     repeat_region   14381. .14676
                     /rpt_family="Alu"
     repeat_region   15967. .19011
                     /rpt_family="Retroviral"
     repeat_region   19016. .19080
                     /rpt_family="(CAGA)n"
     repeat_region   19165. .20032
                     /rpt_family="MER2_type"
     repeat_region   20240. .20718
                     /rpt_family="Retroviral"
     repeat_region   20719. .23148
                     /rpt_family="L1"
     repeat_region   23143. .23594
                     /rpt_family="L1"
     repeat_region   23586. .24864
                     /rpt_family="L1"
     repeat_region   24793. .24885
                     /rpt_family="L1"
     repeat_region   24935. .26701
                     /rpt_family="L1"
     repeat_region   26702. .27001
                     /rpt_family="Alu"
     repeat_region   27002. .27245
                     /rpt_family="L1"
     repeat_region   27611. .27854
                     /rpt_family="L1"
     repeat_region   28328. .28700
                     /rpt_family="MaLR"
     repeat_region   29062. .29358
                     /rpt_family="Alu"
     repeat_region   30617. .30733
                     /rpt_family="MER86"
     repeat_region   31740. .32099
                     /rpt_family="L1"
     repeat_region   32745. .33839
                     /rpt_family="L1"
     repeat_region   33840. .34125
                     /rpt_family="Alu"
     repeat_region   35333. .35379
                     /rpt_family="MIR"
     repeat_region   36805. .36956
```



```
                     /rpt_family="MIR"
     repeat_region   39884. .40160
                     /rpt_family="Alu"
     misc_feature    complement(43425. .43602)
                     /note="match to EST T06037 (NID:g317186)"
     repeat_region   45803. .46102
                     /rpt_family="Alu"
     repeat_region   46186. .46340
                     /rpt_family="MIR"
     repeat_region   47234. .47605
                     /rpt_family="L1"
     repeat_region   47638. .47752
                     /rpt_family="L2"
     repeat_region   47967. .48263
                     /rpt_family="Alu"
     repeat_region   48553. .49328
                     /rpt_family="L1"
     repeat_region   49324. .54540
                     /rpt_family="L1"
     repeat_region   55709. .56097
                     /rpt_family="L1"
     repeat_region   56097. .56676
                     /rpt_family="L1"
     repeat_region   56686. .57131
                     /rpt_family="L1"
     repeat_region   57171. .57464
                     /rpt_family="L2"
     repeat_region   57748. .58223
                     /rpt_family="MaLR"
     repeat_region   58499. .58799
                     /rpt_family="Alu"
     repeat_region   58830. .58954
                     /rpt_family="(GAA)n"
     repeat_region   59033. .59197
                     /rpt_family="MIR"
     repeat_region   59645. .60077
                     /rpt_family="MaLR"
     repeat_region   63705. .63785
                     /rpt_family="L2"
     repeat_region   65588. .65820
                     /rpt_family="L2"
     repeat_region   66392. .66483
                     /rpt_family="MIR"
     repeat_region   66942. .67243
                     /rpt_family="Alu"
     repeat_region   69791. .70616
                     /rpt_family="L1"
     repeat_region   70644. .70942
                     /rpt_family="Alu"
     repeat_region   70976. .71055
                     /rpt_family="L1"
```



```
     repeat_region   71440. .71741
                     /rpt_family="Alu"
     repeat_region   72898. .73202
                     /rpt_family="Alu"
     repeat_region   73840. .74167
                     /rpt_family="Retroviral"
     repeat_region   74705. .75078
                     /rpt_family="MaLR"
     repeat_region   76688. .76847
                     /rpt_family="MIR"
     repeat_region   77734. .77976
                     /rpt_family="MIR"
     repeat_region   80931. .81042
                     /rpt_family="L2"
     repeat_region   81057. .81336
                     /rpt_family="L1"
     repeat_region   81896. .82072
                     /rpt_family="MER1_type"
     repeat_region   82618. .82793
                     /rpt_family="MER1_type"
     misc_feature    complement(82885. .83292)
                     /note="match to EST T94583 (NID:g728071) ye31h05.s1"
     repeat_region   83910. .84067
                     /rpt_family="L2"
...
Note: remainder of annotations omitted.

  Query Match             6.8%;  Score 20;  DB 26;  Length 85579;
  Best Local Similarity 100.0%;  Pred. No. 1.23e+00;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  73582 AGATAATAATACTTCATTCA 73601
          ||||||||||||||||||||
Cp    234 AGATAATAATACTTCATTCA 215


RESULT    6
LOCUS       AC004130   116215 bp    DNA             PRI       06-FEB-1998
DEFINITION  Homo sapiens BAC clone RG293F17 from 7p15-p21, complete sequence.
ACCESSION   AC004130
NID         g2842785
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primatea; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 116215)
  AUTHORS   Tin-Wollam,A., Sutterer,C. and Fronick,B.
  TITLE     The sequence of Homo sapiens BAC clone RG293F17
  JOURNAL   Unpublished (1998)
REFERENCE   2  (bases 1 to 116215)
```



```
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-FEB-1998) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     SUBMITTED BY:  WUGSC
            Genome Sequencing Center
            Department of Genetics
            Washington University
            St. Louis MO 63108, USA
            http://genome.wustl.edu/gsc
            mailto:sapiens@watson.wustl.edu

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded or sequenced with an alternate
            chemistry; an attempt was made to resolve all sequencing problems,
            such as compressions and repeats; all regions were covered by
            sequence from more than one subclone; and the assembly was
            confirmed by restriction digest.

            MAPPING INFORMATION:
            The sequence of this clone was established as part of a mapping and
            sequencing collaboration between the NHGRI Chromosome 7 Mapping
            Project (Eric D. Green, Director), John D. McPherson in the
            Department of Genetics (Washington University), and the Washington
            University Genome Sequencing Center.  For additional information
            about the map position of this sequence, see
            http://www.nhgri.nih.gov/DIR/GTB/CHR7 or send
            mailto:egreen@nhgri.nih.gov

            SOURCE INFORMATION:
            Clone RG293F17 is from a release of the human BAC library
            CITB-HS-A.  The library contains cloned DNA from human sperm.  See:
             Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J.
            Kim et al., Genomics 34:213-8 (1996).  The clone is available from
            Research Genetics, Inc. (http://www.resgen.com).
            VECTOR:  pBeloBAC11
            Selection:  chloramphenicol

            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the right is RG326G04, 200 bp overlap.
            Actual start of this clone is at base position 1 of RG293F17;
            actual end is at 116215 of RG293F17.

            This clone contains STS sWSS2601 (NID:g1222836).

            This clone RG293F17 contains a bacterial transposon from base
```

```
          position 4508-5845, which has been cut from the submitted sequence.
  FEATURES             Location/Qualifiers
       source          1. .116215
                       /organism="Homo sapiens"
                       /db_xref="taxon:9606"
                       /chromosome="7"
                       /clone="RG293F17"
                       /clone_lib="CITB-HS-A"
                       /map="7p15-p21"
       repeat_region   1078. .1302
                       /rpt_family="MIR"
       repeat_region   1939. .2240
                       /rpt_family="ALU"
       repeat_region   4266. .4475
                       /rpt_family="L1"
       repeat_region   6166. .6460
                       /rpt_family="ALU"
       repeat_region   8190. .9034
                       /rpt_family="L1"
       repeat_region   9046. .9128
                       /rpt_family="MIR"
       repeat_region   9629. .9780
                       /rpt_family="RETROVIRAL"
       repeat_region   10256. .10552
                       /rpt_family="ALU"
       repeat_region   10693. .10742
                       /rpt_family="MIR"
       repeat_region   11122. .11270
                       /rpt_family="MER1_TYPE"
       repeat_region   11568. .11677
                       /rpt_family="ALU"
       misc_feature    14096. .15596
                       /note="CpG_island (%GC=67.1, o/e=0.70, #CpGs=138)"
       gene            15522. .93712
                       /gene="ITB8"
       CDS             join(15522. .15648,47352. .47437,50727. .50901,62766. .63012,
                       64381. .64546,65442. .65600,75118. .75213,78611. .78700,
                       82575. .82709,85436. .85841,88343. .88568,89777. .89886,
                       93330. .93493,93590. .93712)
                       /gene="ITB8"
                       /note="match to P26012 (PID:g124975); H_RG293F17.1"
                       /codon_start=1
                       /product="INTEGRIN BETA-8 SUBUNIT PRECURSOR"
                       /db_xref="PID:g2842786"
                       /translation="MCGSALAFTTAAFVCLQNDRRGPASFLWAAWVFSLVLGLQQED
                       NRCASSNAASCARCLALGPECGWCVQEDFISGGSRSERCDIVSNLISKGCSVDSIEYP
                       SVHVIIPTENEINTQVTPGEVSIQLRPGAEANFMLKVHPLKKYPVDLYYLVDVSASMH
                       NNIEKLNSVGNDLSRKMAFFSRDFRLGFGSIVDKTVSPYISIHPERIHNQCSDYNLDC
                       MPPHGYIHVLSLTENITEFEKAVHRQKISGNIDTPEGGFDAMLQAAVCESHIGWRKEA
                       KRLLLVMTDQTSHLALDSKLAGIVVPNDGNCHLKNNVYVKSTTMEHPSLGQLSEKLID
                       NNINVIFAVQGKQFHWYKDLLPLLPGTIAGEIESKAANLNNLVVEAYQKLISEVKVQV
```



```
                       ENQVQGIYFNITAICPDGSRKPGMEGCRNVTSNDEVLFNVTVTMKKCDVTGGKNYAII
                       KPIGFNETAKIHIHRNCSCQCEDNRGPKGKCVDETFLDSKCFQCDENKCHFDEDQFSS
                       ESCKSHKDQPVCSGRGVCVCGKCSCHKIKLGKVYGKYCEKDDFSCPYHHGNLCAGHGE
                       CEAGRCQCFSGWEGDRCQCPSAAAQHCVNSKGQVCSGRGTCVCGRCECTDPRSIGRFC
                       EHCPTCYTACKENWNCMQCLHPHNLSQAILDQCKTSCALMEQQHYVDQTSECPSSPSY
                       LRIFFIIFIVTFLIGLLKVLIIRQVILQWNSNKIKSSSDYRVSASKKDKLILQSVCTR
                       AVTYRREKPEEIKMDISKLNAHETFRCNF"
       repeat_region   16367. .16411
                       /rpt_family="MIR"
       repeat_region   19318. .19595
                       /rpt_family="L2"
       repeat_region   19719. .20019
                       /rpt_family="ALU"
       repeat_region   20171. .20303
                       /rpt_family="MIR"
       repeat_region   21192. .21371
                       /rpt_family="MIR"
       repeat_region   22060. .22362
                       /rpt_family="ALU"
       repeat_region   22627. .22923
                       /rpt_family="ALU"
       repeat_region   23461. .23764
                       /rpt_family="ALU"
       repeat_region   25495. .25561
                       /rpt_family="MARINER"
       repeat_region   28545. .28641
                       /rpt_family="L2"
       repeat_region   28642. .28695
                       /rpt_family="RETROVIRAL"
       repeat_region   28981. .29310
                       /rpt_family="L2"
       repeat_region   29362. .29706
                       /rpt_family="L2"
       repeat_region   30602. .30960
                       /rpt_family="MALR"
       repeat_region   33045. .33124
                       /rpt_family="MARINER"
       repeat_region   33574. .33727
                       /rpt_family="L1"
       repeat_region   33861. .34162
                       /rpt_family="ALU"
       repeat_region   35013. .35240
                       /rpt_family="MIR"
       repeat_region   37367. .37666
                       /rpt_family="ALU"
       repeat_region   38876. .39009
                       /rpt_family="RETROVIRAL"
       repeat_region   41291. .41433
                       /rpt_family="MIR"
       repeat_region   41679. .41977
                       /rpt_family="ALU"
```



```
       repeat_region   42244. .42417
                       /rpt_family="ALU"
       repeat_region   42430. .42876
                       /rpt_family="L1"
       repeat_region   44091. .44913
                       /rpt_family="L1"
       repeat_region   45045. .45375
                       /rpt_family="MALR"
       repeat_region   49090. .49284
                       /rpt_family="MIR"
       repeat_region   50198. .50495
                       /rpt_family="ALU"
       repeat_region   51157. .51274
                       /rpt_family="L2"
       repeat_region   52969. .53361
                       /rpt_family="MALR"
       repeat_region   53486. .54129
                       /rpt_family="L1"
       repeat_region   54325. .54628
                       /rpt_family="ALU"
       repeat_region   54702. .54921
                       /rpt_family="L1"
       repeat_region   56264. .56393
                       /rpt_family="L2"
       misc_feature    complement(62953. .63034)
                       /gene="ITB8"
                       /note="match to EST AA644567 (NID:g2569785) ab63h07.s1"
       repeat_region   64883. .65176
                       /rpt_family="ALU"
       misc_feature    complement(65432. .65781)
                       /gene="ITB8"
                       /note="match to EST AA644567 (NID:g2569785) ab63h07.s1"
       repeat_region   68145. .68326
...
Note: remainder of annotations omitted.

  Query Match            6.8%;  Score 20;  DB 26;  Length 116215;
  Best Local Similarity 100.0%;  Pred. No. 1.23e+00;
  Matches    20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  19060 CATTAAAAATATTTTTATTT 19079
          ||||||||||||||||||||
Qy    100 CATTAAAAATATTTTTATTT 119


RESULT    7
LOCUS       AC002382   125590 bp    DNA             PRI       23-JUL-1997
DEFINITION  Human BAC clone RG022J17 from 7q21, complete sequence.
ACCESSION   AC002382
NID         g2275185
KEYWORDS    HTG.
```



```
  SOURCE      human.
    ORGANISM  Homo sapiens
              Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
              Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
              Homo.
  REFERENCE   1  (bases 1 to 125590)
    AUTHORS   Graves,T, Sutterer,C and Hawkins,M.
    TITLE     The sequence of H. sapiens BAC clone RG022J17
    JOURNAL   Unpublished (1997)
  REFERENCE   2  (bases 1 to 125590)
    AUTHORS   Waterston,R.
    TITLE     Direct Submission
    JOURNAL   Submitted (23-JUL-1997)
  COMMENT     SUBMITTED BY:
              Genome Sequencing Center
              Department of Genetics
              Washington University
              St. Louis MO 63108, USA
              http://genome.wustl.edu/gsc
              mailto:sapiens@watson.wustl.edu

              NOTICE:  This sequence may not represent the entire insert of this
              clone.  It may be shorter because we only sequence overlapping
              clone sections once, or longer because we provide a small overlap
              between neighboring data submissions.

              This sequence was finished as follows unless otherwise noted:
              all regions were double stranded or sequenced with an alternate
              chemistry; an attempt was made to resolve all sequencing problems,
              such as compreasions and repeats; all regions were covered by
              sequence from more than one subclone; and the assembly was
              confirmed by restriction digest.

              MAPPING INFORMATION:
              The sequence of this clone was established as part of a mapping and
              sequencing collaboration between the NHGRI Chromosome 7 Mapping
              Project and the Washington University Genome Sequencing Center.
              For additional information about the map position of this sequence,
              see http://www.nhgri.nih.gov/DIR/GTB/CHR7 or send an E-mail to
              egreen@nhgri.nih.gov

              SOURCE INFORMATION:
              This clone is from the first release of the human BAC library.  The
              library contains cloned DNA from a human male fibroblast cell line
              978SK.  See:  Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7
              (1992); Kim et al., Genomics 34:213-8 (1996).  The clone is
              available from Research Genetics, Inc. (http://www.resgen.com).
              VECTOR:  pBeloBAC11
              Selection:  chloramphenicol

              NEIGHBORING SEQUENCE INFORMATION:
```

The clone sequenced to the right is RG067E13, 200 bp overlap.
Actual start of this clone is at base position 1 of RG022J17;
actual end is at 64313 of RG067E13.

This clone contains STS sWSS1762 (NID:g1254729).

```
FEATURES             Location/Qualifiers
     source          1. .125590
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /clone="RG022J17"
                     /clone_lib="CITB-978SK-B"
                     /map="7q21"
     repeat_region   complement(3732. .3768)
                     /rpt_family="L1"
     repeat_region   complement(8187. .8567)
                     /rpt_family="L1"
     repeat_region   8697. .8986
                     /rpt_family="ALU"
     repeat_region   11929. .11996
                     /rpt_family="ALU"
     repeat_region   12796. .12824
                     /rpt_family="L1"
     repeat_region   12885. .12911
                     /rpt_family="L1"
     repeat_region   complement(14261. .14349)
                     /rpt_family="THE"
     repeat_region   complement(14449. .14579)
                     /rpt_family="THE"
     repeat_region   complement(15220. .15239)
                     /rpt_family="L1"
     repeat_region   17699. .17975
                     /rpt_family="ALU"
     repeat_region   17976. .18466
                     /rpt_family="MER"
     repeat_region   19265. .19430
                     /rpt_family="L1"
     repeat_region   complement(20334. .20363)
                     /rpt_family="L1"
     repeat_region   20416. .20705
                     /rpt_family="ALU"
     repeat_region   20832. .20883
                     /rpt_family="L1"
     repeat_region   20954. .21232
                     /rpt_family="L1"
     repeat_region   21823. .22011
                     /rpt_family="L1"
     repeat_region   22102. .22161
                     /rpt_family="L1"
     repeat_region   22529. .22721
                     /rpt_family="L1"
```



```
     repeat_region   complement(22778. .22812)
                     /rpt_family="L1"
     repeat_region   complement(23818. .23853)
                     /rpt_family="L1"
     repeat_region   complement(24405. .24432)
                     /rpt_family="L1"
     repeat_region   complement(24918. .24937)
                     /rpt_family="L1"
     repeat_region   30642. .30679
                     /rpt_family="L1"
     repeat_region   30938. .30988
                     /rpt_family="L1"
     repeat_region   complement(31198. .31300)
                     /rpt_family="L1"
     repeat_region   33598. .33626
                     /rpt_family="L1"
     repeat_region   35327. .35359
                     /rpt_family="L1"
     repeat_region   complement(42189. .42467)
                     /rpt_family="ALU"
     repeat_region   complement(42654. .42947)
                     /rpt_family="ALU"
     repeat_region   complement(45536. .45626)
                     /rpt_family="L1"
     repeat_region   complement(45744. .45794)
                     /rpt_family="L1"
     repeat_region   47538. .47563
                     /rpt_family="L1"
     repeat_region   complement(53577. .53615)
                     /rpt_family="L1"
     repeat_region   53622. .53659
                     /rpt_family="L1"
     repeat_region   53910. .54195
                     /rpt_family="ALU"
     repeat_region   complement(55628. .55659)
                     /rpt_family="L1"
     repeat_region   complement(58950. .60613)
                     /rpt_family="L1"
     repeat_region   59953. .60371
                     /rpt_family="L1"
     repeat_region   60530. .60555
                     /rpt_family="L1"
     repeat_region   60571. .61088
                     /rpt_family="L1"
     repeat_region   complement(61438. .61521)
                     /rpt_family="L1"
     repeat_region   61444. .61499
                     /rpt_family="L1"
     repeat_region   63094. .63113
                     /rpt_family="L1"
     repeat_region   complement(63609. .63632)
```



```
                     /rpt_family="L1"
     repeat_region   complement(63637. .63925)
                     /rpt_family="ALU"
     repeat_region   65330. .65497
                     /rpt_family="L1"
     repeat_region   65770. .66436
                     /rpt_family="L1"
     repeat_region   66837. .67169
                     /rpt_family="L1"
     repeat_region   complement(67187. .67482)
                     /rpt_family="ALU"
     repeat_region   67484. .67600
                     /rpt_family="L1"
     repeat_region   69644. .69927
                     /rpt_family="ALU"
     repeat_region   71037. .71333
                     /rpt_family="ALU"
     repeat_region   complement(71761. .71795)
                     /rpt_family="L1"
     repeat_region   complement(72361. .72420)
                     /rpt_family="L1"
     repeat_region   complement(73908. .74195)
                     /rpt_family="ALU"
     repeat_region   complement(74575. .74959)
                     /rpt_family="L1"
     repeat_region   complement(75038. .76147)
                     /rpt_family="L1"
     repeat_region   75103. .75125
                     /rpt_family="L1"
     repeat_region   76145. .76992
                     /rpt_family="L1"
     repeat_region   complement(76286. .76317)
                     /rpt_family="L1"
     repeat_region   complement(76998. .77207)
                     /rpt_family="L1"
     repeat_region   80223. .80263
                     /rpt_family="L1"
     repeat_region   complement(80701. .80733)
                     /rpt_family="L1"
     repeat_region   81143. .81184
                     /rpt_family="L1"
     repeat_region   81952. .82065
...
Note: remainder of annotations omitted.
```

```
  Query Match           6.8%;  Score 20;  DB 26;  Length 125590;
  Best Local Similarity 100.0%;  Pred. No. 1.23e+00;
  Matches   20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  27568 TTTATTTATAACTTTTAGAA 27587
          ||||||||||||||||||||
```



```
Qy    127 TTTATTTATAACTTTTAGAA 146


RESULT    8
LOCUS     HS501N12   170952 bp    DNA             PRI       24-JUL-1998
DEFINITION  Homo sapiens DNA sequence from clone 501N12 on chromosome
            6p22.1-22.3 Contains a gene almost identical to four genes of
            unknown function, a pseudogene, three (pseudo?) genes similar to
            genes of unknown function, an unknown gene similar to a rat EST, a
            PX19 LIKE pseudogene and another unknown gene. Contains ESTs, STSs
            and GSSs, complete sequence.
ACCESSION   AL022170
NID         g3281976
KEYWORDS    HTG; PX19.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 170952)
  AUTHORS   Tubby,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUL-1998) E-mail enquires: humquery@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 1, 1998 this sequence version replaced gi:2980811.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above. This sequence is the
            entire insert of clone 501N12. This sequence has been finished
            according to sequence map criteria as follows. An attempt is made
            to resolve all sequencing problems, such as compressions and
            repeats, but not necessarily within known annotated human repeat
            sequence elements (e.g. Alu). Where the sequence is ambiguous,
            there is an annotation using the 'unsure' feature key.
            This sequence was generated from part of bacterial clone contigs of
            human chromosome 6, constructed in collaboration by the Sanger
            Centre chromosome 6 mapping group and Armin Volz & Andreas Ziegler.
            Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6/ 501N12 is from the library RPCI3
            constructed at the Roswell Park Cancer Institute by the group of
            Pieter de Jong. For further details see
            http://bacpac.med.buffalo.edu/ VECTOR: pCYPAC2.
FEATURES              Location/Qualifiers
     source           1. .170952
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="6"
                      /clone="501N12"
                      /map="p22.1-22.3"
```

```
                /clone_lib="RPCI3"
repeat_region   444. .516
                /note="MER5A repeat: matches 101. .188 of consensus"
repeat_region   complement(832. .1100)
                /note="AluJb repeat: matches 280. .1 of consensus"
repeat_region   complement(1341. .1477)
                /note="MIR repeat: matches 237. .85 of consensus"
repeat_region   2942. .3479
                /note="L1 repeat: matches 1432. .1968 of consensus"
repeat_region   4030. .4359
                /note="AluY repeat: matches 1. .296 of consensus"
repeat_region   complement(4554. .4853)
                /note="AluJo repeat: matches 301. .1 of consensus"
gene            6829. .34852
                /gene="dJ501N12.1"
CDS             join(<6829. .6895,15857. .15985,17949. .18080,28912. .29087,
                32011. .32098,34703. .>34852)
                /gene="dJ501N12.1"
                /note="match: cDNAs Y10518 Y10519 Z83950 Y10508"
                /codon_start=2
                /evidence=not_experimental
                /product="dJ501N12.1"
                /db_xref="PID:e1312796"
                /db_xref="PID:g3355565"
                /translation="RKYLQEQAYRLQQGIVTSTTQQMIDRICVKVQDHLNSLRNCGGD
                AIQEDLKSAEELMRDAKNSKTLLPNLYHVGGASWAGASGLLSSPIQETLESMAGEVTR
                VVDEQLKALLESKVDAAEMLCPNVMKKAHIRQDLIHASTEKISIPRTFVKNVLLEQSG
                IDILNKISEVKLTVASFLSDRIVDEILDALSHCHHKLADHFSRRGKTLPQQESLEIEL
                AEEKPVKRSIITVEELTEIERLEDLDTCM"
prim_transcript 7271. .7658
                /gene="dJ501N12.1"
                /note="match: EST AA526622"
repeat_region   7659. .7776
                /note="AluYb8 repeat: matches 191. .308 of consensus"
repeat_region   8239. .8475
                /note="AluSg repeat: matches 1. .239 of consensus"
misc_feature    complement(8536. .8952)
                /gene="dJ501N12.1"
                /note="match: GSS B37778"
repeat_region   complement(9800. .10072)
                /note="AluSx repeat: matches 273. .1 of consensus"
repeat_region   complement(10099. .10222)
                /note="AluSp repeat: matches 303. .179 of consensus"
repeat_region   11294. .11503
                /note="MIR repeat: matches 4. .214 of consensus"
prim_transcript complement(12237. .12525)
                /gene="dJ501N12.1"
                /note="match: EST AA639318"
repeat_region   complement(12599. .12717)
                /note="MIR2 repeat: matches 146. .28 of consensus"
repeat_region   15448. .15519
```



```
                /note="MIR2 repeat: matches 74. .146 of consensus"
misc_feature    complement(15866. .16333)
                /gene="dJ501N12.1"
                /note="match: GSS AQ016761"
repeat_region   complement(16512. .16606)
                /note="MER21B repeat: matches 790. .703 of consensus"
repeat_region   complement(16545. .16637)
                /note="MER39 repeat: matches 676. .582 of consensus"
repeat_region   complement(16603. .17427)
                /note="MER21B repeat: matches 792. .4 of consensus"
repeat_region   17480. .17647
                /note="AluSp repeat: matches 136. .303 of consensus"
repeat_region   complement(19382. .19433)
                /note="L1HS repeat: matches 578. .526 of consensus"
repeat_region   complement(19656. .19960)
                /note="AluSx repeat: matches 300. .1 of consensus"
repeat_region   22609. .22648
                /note="20 copies 2 mer ac 100% conserved"
repeat_region   complement(22914. .23103)
                /note="MER3 repeat: matches 209. .1 of consensus"
repeat_region   24161. .24346
                /note="L1 repeat: matches 4042. .4231 of consensus"
repeat_region   24420. .24689
                /note="AluJo repeat: matches 1. .302 of consensus"
repeat_region   24771. .25083
                /note="AluSq repeat: matches 2. .303 of consensus"
repeat_region   26805. .26904
                /note="MIR repeat: matches 35. .134 of consensus"
repeat_region   complement(27877. .28054)
                /note="MIR repeat: matches 260. .77 of consensus"
repeat_region   28097. .28238
                /note="MIR repeat: matches 71. .201 of consensus"
misc_feature    complement(28312. .28688)
                /gene="dJ501N12.1"
                /note="match: GSS AQ020591"
repeat_region   complement(30856. .30913)
                /note="MER5A repeat: matches 111. .57 of consensus"
repeat_region   complement(33402. .33698)
                /note="AluSx repeat: matches 301. .1 of consensus"
repeat_region   complement(38477. .38699)
                /note="MIR repeat: matches 253. .2 of consensus"
repeat_region   complement(40249. .40548)
                /note="AluSg repeat: matches 299. .1 of consensus"
repeat_region   complement(41636. .41692)
                /note="MIR2 repeat: matches 146. .87 of consensus"
repeat_region   42990. .43016
                /note="MIR2 repeat: matches 98. .124 of consensus"
repeat_region   43567. .43860
                /note="AluSc repeat: matches 1. .298 of consensus"
repeat_region   44628. .44723
                /note="MIR2 repeat: matches 30. .134 of consensus"
```



```
repeat_region   complement(45009. .45166)
                /note="MIR repeat: matches 180. .21 of consensus"
repeat_region   complement(47974. .48264)
                /note="AluY repeat: matches 301. .2 of consensus"
mRNA            complement(48410. .49073)
                /gene="dJ501N12.2"
                /note="match: cDNA Y10205"
                /pseudo
                /evidence=not_experimental
gene            complement(48410. .49073)
                /gene="dJ501N12.2"
repeat_region   49166. .50348
                /note="MER42c repeat: matches 348. .1536 of consensus"
repeat_region   50415. .50724
                /note="AluJo repeat: matches 2. .302 of consensus"
repeat_region   complement(50998. .51085)
                /note="MIR2 repeat: matches 145. .58 of consensus"
repeat_region   complement(52655. .52797)
                /note="L1ME2 repeat: matches 699. .554 of consensus"
repeat_region   complement(52991. .53050)
                /note="MIR repeat: matches 234. .175 of consensus"
repeat_region   complement(60954. .60997)
                /note="MER34 repeat: matches 524. .480 of consensus"
repeat_region   complement(61036. .61200)
                /note="MER39 repeat: matches 409. .251 of consensus"
repeat_region   61430. .61537
                /note="MIR2 repeat: matches 32. .142 of consensus"
repeat_region   complement(62029. .62337)
                /note="AluJo repeat: matches 301. .6 of consensus"
repeat_region   62452. .62815
                /note="MER39 repeat: matches 48. .409 of consensus"
repeat_region   62854. .62898
                /note="MER34 repeat: matches 480. .525 of consensus"
repeat_region   63567. .65020
                /note="SVA repeat: matches 4. .1372 of consensus"
misc_feature    complement(65545. .65881)
                /note="match: STS G05436"
repeat_region   complement(65917. .66232)
                /note="MER2 repeat: matches 345. .9 of consensus"
repeat_region   complement(67491. .67789)
                /note="AluJb repeat: matches 300. .1 of consensus"
repeat_region   complement(68357. .68545)
                /note="L1 repeat: matches 3768. .3579 of consensus"
prim_transcript complement(70789. .71348)
                /gene="dJ501N12.3"
                /note="possibly a pseudogene; match: cDNAs Y10202 Z83943"
                /evidence=not_experimental
gene            complement(70789. .71348)
                /gene="dJ501N12.3"
misc_feature    complement(71819. .72143)
                /note="match: GSS B53022"
```



```
...
Note: remainder of annotations omitted.

  Query Match           6.8%;  Score 20;  DB 26;  Length 170952;
  Best Local Similarity 100.0%;  Pred. No. 1.23e+00;
  Matches   20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  83843 AACAACCAAAAAGTTTATTG 83862
          ||||||||||||||||||||
Cp    287 AACAACCAAAAAGTTTATTG 268


RESULT    9
LOCUS     AC003664    202233 bp    DNA             PRI       03-FEB-1998
DEFINITION  Homo sapiens chromosome 17, clone hCIT54K19, complete sequence.
ACCESSION   AC003664
NID         g2828774
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 202233)
  AUTHORS   Birren,B., Fasman,K., McKernan,K., Nusbaum,C., Richardson,P. and
            Lander,E.
  TITLE     Homo sapiens chromosome 17, clone hCIT54K19
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 202233)
  AUTHORS   Birren,B., Fasman,K., McKernan,K., Munro,C., Nusbaum,C.,
            Richardson,P., Lander,E., Baldwin,J., Barna,N., Cantu,C., Chang,A.,
            Cooke,P., Daly,M.J., Depayre,E., Devon,K., Dewar,K., DuRette,B.,
            Forrest,C., Gage,D., Gensheimer,S., Geraigery,K., Gilmartin,T.,
            Hagos,B., Halphen,I., Harris,K., Howland,J.C., Huang,J., Hui,L.,
            Jacotot,L., Linton,L., MacKenzie,J., Marquis,N., McDermott,J.,
            McGurk,A., Molla,M., Morris,W., Morrow,J., Nachman,A., Naylor,J.,
            O'Connor,T., Olotu,A., Peterson,K., Roberts,D., Rollins,G.,
            Sarnaik,A., Shiu,P., Shyam,R., Stilwell,J., Stone,C.,
            Strickland,C., Sydney,K., Tang,L., Zemtseva,I., Zhao,J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-DEC-1997) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 202233)
  AUTHORS   Birren,B., Fasman,K., McKernan,K., Nusbaum,C., Richardson,P.,
            Lander,E., Allen,N., Baker,J., Baldwin,J., Barna,N., Beckerly,R.,
            Boutwell,C., Byrne,S., Cantu,C., Castle,A., Cooke,P., Daly,M.J.,
            Depayre,E., Devon,K., Dewar,K., DuRette,B., Etemadi,S.,
            Ferreira,P., Forrest,C., Gage,D., Gardyna,S., Gensheimer,S.,
            Geraigery,K., Gilmartin,T., Gray,D., Hagos,B., Harris,K.,
            Horton,L., Howland,J.C., Hui,L., Jacotot,L., Linton,L.,
            MacKenzie,J., Marquis,N., McEwan,P., McGurk,A., Meldrim,J.,
            Molla,M., Morris,W., Morrow,J., Nachman,A., Naylor,J., O'Connor,T.,
```

```
            Pavlin,B., Peterson,K., Ranganath,S., Riley,R., Roberts,D.,
            Rollins,G., Rossello,R., Roy,A., Shyam,R., Soohoo,S., Stilwell,J.,
            Stone,C., Strickland,C., Sydney,K., Tang,L., Vassiliev,H., Vo,A.,
            Wagner,A., Wheeler,J., Wu,Y., Ye,W.J., Zemtseva,I., Zhao,J. and
            Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-FEB-1998) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 3, 1998 this sequence version replaced gi:2815559.
            The Staden databases, finishing information, and all
            chromatographic files used in the assembly of this clone are
            available from our anonymous ftp site.

            All repeats were identified using RepeatMasker: Smit, A.F.A. &
            Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html.
FEATURES             Location/Qualifiers
     source          1. .202233
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="hCIT54K19"
                     /clone_lib="Research Genetics/Cal Tech CITB978SK-B (plates
                     1-194)"
                     /map="17"
                     /chromosome="17"
     repeat_region   complement(1447. .1509)
                     /rpt_family="AT_rich"
     repeat_region   complement(1977. .2413)
                     /rpt_family="MER28"
     repeat_region   2794. .2999
                     /rpt_family="L1"
     repeat_region   3093. .3158
                     /rpt_family="AT_rich"
     repeat_region   3363. .3661
                     /rpt_family="AluSg"
     repeat_region   complement(3748. .3773)
                     /rpt_family="AT_rich"
     repeat_region   3828. .3899
                     /rpt_family="AT_rich"
     repeat_region   4086. .4148
                     /rpt_family="AT_rich"
     repeat_region   complement(4318. .4397)
                     /rpt_family="LINE2"
     repeat_region   complement(4401. .4990)
                     /rpt_family="LINE2"
     repeat_region   complement(5048. .5112)
                     /rpt_family="AT_rich"
     repeat_region   complement(5173. .5213)
                     /rpt_family="AT_rich"
     repeat_region   6572. .6874
                     /rpt_family="AluSq"
```



```
     repeat_region   complement(7529. .8617)
                     /rpt_family="LINE2"
     repeat_region   complement(8992. .9065)
                     /rpt_family="MER77"
     repeat_region   complement(9109. .9560)
                     /rpt_family="MLT2A"
     repeat_region   complement(9561. .9890)
                     /rpt_family="MER77"
     repeat_region   9895. .10244
                     /rpt_family="THE1A"
     repeat_region   10916. .11261
                     /rpt_family="MER33"
     repeat_region   complement(13745. .14445)
                     /rpt_family="L1PA16"
     repeat_region   14444. .14717
                     /rpt_family="L1"
     repeat_region   14705. .15067
                     /rpt_family="L1"
     repeat_region   15761. .15824
                     /rpt_family="MIR"
     repeat_region   16004. .16827
                     /rpt_family="MER21A"
     repeat_region   complement(17013. .17282)
                     /rpt_family="(GAAA)n"
     repeat_region   complement(17284. .17422)
                     /rpt_family="(GA)n"
     repeat_region   complement(17646. .17692)
                     /rpt_family="(CA)n"
     repeat_region   complement(17800. .17854)
                     /rpt_family="MLT1F"
     repeat_region   complement(17872. .17953)
                     /rpt_family="MLT1F"
     repeat_region   complement(17984. .18047)
                     /rpt_family="MST-INTERNAL"
     repeat_region   complement(18134. .18299)
                     /rpt_family="MLT1-INTERNAL"
     repeat_region   complement(18589. .18857)
                     /rpt_family="MLT1-INTERNAL"
     repeat_region   complement(18872. .19088)
                     /rpt_family="AluSg/x"
     repeat_region   complement(20173. .20187)
                     /rpt_family="LINE2"
     repeat_region   complement(20188. .20220)
                     /rpt_family="MIR"
     repeat_region   complement(20221. .20423)
                     /rpt_family="LINE2"
     repeat_region   complement(20704. .20943)
                     /rpt_family="MER69B"
     repeat_region   20946. .21236
                     /rpt_family="AluSq"
     repeat_region   complement(21237. .21389)
```



```
                     /rpt_family="MER69B"
     repeat_region   complement(21390. .21679)
                     /rpt_family="AluJo"
     repeat_region   complement(21688. .21842)
                     /rpt_family="MER69B"
     repeat_region   complement(21832. .21975)
                     /rpt_family="MER69B"
     repeat_region   complement(23805. .23903)
                     /rpt_family="L1PA16"
     repeat_region   complement(23906. .24336)
                     /rpt_family="MER57A"
     repeat_region   complement(24337. .25040)
                     /rpt_family="L1PA15"
     repeat_region   25283. .26868
                     /rpt_family="L1M4"
     repeat_region   26908. .27002
                     /rpt_family="L1M4"
     repeat_region   complement(27028. .28892)
                     /rpt_family="L1PA16"
     repeat_region   28891. .29505
                     /rpt_family="L1PA16"
     repeat_region   29507. .30813
                     /rpt_family="L1MC3"
     repeat_region   complement(30814. .30986)
                     /rpt_family="MSTD"
     repeat_region   30987. .31633
                     /rpt_family="L1MC3"
     repeat_region   31671. .32128
                     /rpt_family="MLT1D"
     repeat_region   32129. .35954
                     /rpt_family="L1MC3"
     repeat_region   complement(36209. .36313)
                     /rpt_family="(GAAAA)n"
     repeat_region   complement(36314. .36451)
                     /rpt_family="(GAAA)n"
     repeat_region   complement(36501. .36787)
                     /rpt_family="AluSx"
     repeat_region   complement(36805. .36850)
                     /rpt_family="(CA)n"
     repeat_region   complement(36865. .37044)
                     /rpt_family="L1PA5"
     repeat_region   37192. .37239
                     /rpt_family="AT_rich"
     repeat_region   complement(37278. .37715)
                     /rpt_family="L1PA12"
     repeat_region   38354. .38654
                     /rpt_family="AluSp"
     repeat_region   38678. .38893
                     /rpt_family="L1PA13"
     repeat_region   39541. .39697
                     /rpt_family="MIR"
```



```
     repeat_region   41041. .41076
                     /rpt_family="(TA)n"
     repeat_region   41163. .41480
                     /rpt_family="MER2"
     repeat_region   complement(41481. .41521)
                     /rpt_family="(CA)n"
     repeat_region   42047. .42086
                     /rpt_family="(TA)n"
     repeat_region   complement(42146. .42295)
...
Note: remainder of annotations omitted.

  Query Match           6.8%;  Score 20;  DB 26;  Length 202233;
  Best Local Similarity 100.0%;  Pred. No. 1.23e+00;
  Matches   20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  60769 TATTTAAATAAAAATATTTT 60788
          ||||||||||||||||||||
Cp    124 TATTTAAATAAAAATATTTT 105


RESULT   10
LOCUS       G22210         273 bp    DNA             STS       31-MAY-1996
DEFINITION  human STS WI-16602.
ACCESSION   G22210
NID         g1342536
KEYWORDS    STS sequence; primer; sequence tagged site.
SOURCE      human STSs derived from sequences in dbEST and the Unigene
            collection.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 273)
  AUTHORS   Hudson,T.
  TITLE     Whitehead Institute/MIT Center for Genome Research; Physically
            Mapped STSs
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Thomas Hudson
            Whitehead Institute/MIT Center for Genome Research
            Whitehead Institute for Biomedical Research
            9 Cambridge Center, Cambridge MA 02142 USA
            Tel: 617 252 1900
            Fax: 617 252 1902
            Email: thudson@genome.wi.mit.edu

            Primer A: TTTCAGACTTGGTCTGTGAAGG
            Primer B: ACCGAGCCATCTGTAATCTTATC
            STS size: 139
            PCR Profile:
                 Presoak:
```

```
                Denaturation:
                Annealing: 56 degrees C
                Polymerization:
                PCR Cycles: 35
                Thermal Cycler:
            Protocol:
                Template: 10 ng
                Primer: each 5 pM
                dNTPs: each 4 nM
                Taq Polymerase: 0.025 units/ul
                Total Vol: 20 ul

            Buffer:
                MgCl2: 1.5 mM
                KCl: 50 mM
                Tris-HCL: 10 mM
                pH: 9.3

            Derived from dbEST (genbank accession H67871).
FEATURES             Location/Qualifiers
     source          1. .273
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="573.7 cR from top of Chr8 linkage group"
     STS             6. .144
     primer_bind     6. .27
     primer_bind     complement(122. .144)
BASE COUNT       60 a     42 c     74 g     96 t      1 others
ORIGIN

  Query Match           6.4%;  Score 19;  DB 30;  Length 273;
  Best Local Similarity 100.0%;  Pred. No. 5.52e+00;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      62 TTATCTGTCTCTTGTAGTA 80
           ||||||||||||||||||
Qy     229 TTATCTGTCTCTTGTAGTA 247


RESULT   11
LOCUS       PARMTDI1B     582 bp     DNA             INV       30-SEP-1988
DEFINITION  paramecium species 1,168 mt dna dimer: replication init. region.
ACCESSION   K00915
NID         g342952
KEYWORDS    origin of replication.
SOURCE      paramecium species 1,168 mitochondrial dna.
  ORGANISM  Mitochondrion Paramecium sp.
            Eukaryotae; mitochondrial eukaryotes; Alveolata; Ciliophora;
            Nassophorea; Peniculida; Paramecium.
REFERENCE   1  (bases 1 to 582)
  AUTHORS   Pritchard,A.E., Laping,J.L., Seilhamer,J.J. and Cummings,D.J.
```



```
  TITLE     inter-species sequence diversity in the replication initiation
            region of paramecium mitochondrial dna
  JOURNAL   J. Mol. Biol. 164, 1-15 (1983)
  MEDLINE   83189096
REFERENCE   2  (bases 1 to 582)
  AUTHORS   Pritchard,A.E.
  JOURNAL   Unpublished (1984)
COMMENT     [2]  corrected sequence for [1].
            [2] provided revised sequences for [1] because there were several
            errors in the published sequences. the sequences provided by [2]
            are used in this entry. the revisions have not been noted.
FEATURES             Location/Qualifiers
     source          1. .582
                     /organism="Paramecium sp."
                     /mitochondrion
                     /db_xref="taxon:5887"
BASE COUNT      218 a     53 c     50 g    261 t
ORIGIN

  Query Match           6.4%;  Score 19;  DB 18;  Length 582;
  Best Local Similarity 100.0%;  Pred. No. 5.52e+00;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      89 ATTAAAAATATTTTTATTT 107
           ||||||||||||||||||||
Qy     101 ATTAAAAATATTTTTATTT 119


RESULT   12
LOCUS       MMPSCO       2313 bp     RNA             ROD       03-JUL-1996
DEFINITION  M.musculus mRNA for prostaglandin synthase cyclooxygenase, 3'UTR.
ACCESSION   X98792
NID         g1419017
KEYWORDS    3' UTR; prostaglandin synthase cyclooxygenase.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;
            Murinae; Mus.
REFERENCE   1  (bases 1 to 2312)
  AUTHORS   Hofbauer,R. and Denhardt,D.T.
  TITLE     Cell cycle-regulated and proliferation stimulus-responsive genes
  JOURNAL   Crit. Rev. Eukaryot. Gene Expr. 1 (4), 247-300 (1991)
  MEDLINE   92199443
REFERENCE   2  (bases 1 to 2313)
  AUTHORS   Hofbauer,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUN-1996) R. Hofbauer, Institute of Molecular
            Biology, Vienna Biocenter, University of Vienna, Dr. Bohrg. 9, A-
            1030 Vienna, AUSTRIA
COMMENT     Related sequences: M64291 & M94967.
```



```
FEATURES             Location/Qualifiers
     source          1. .2313
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /cell_type="fibroblast"
                     /clone_lib="lambda ZAP"
                     /clone="2-15b"
                     /chromosome="1"
     mRNA            <1. .>2313
                     /product="prostaglandin synthase cyclooxygenase"
     3'UTR           1. .2313
     polyA_signal    2294
BASE COUNT      779 a    347 c    414 g    773 t
ORIGIN

  Query Match           6.4%;  Score 19;  DB 28;  Length 2313;
  Best Local Similarity 100.0%;  Pred. No. 5.52e+00;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    1269 TTAAAAATATTTTTATTTA 1287
           ||||||||||||||||||||
Qy     102 TTAAAAATATTTTTATTTA 120


RESULT   13
LOCUS       DRO37DC97Z   3225 bp     DNA             INV       02-JUL-1996
DEFINITION  Drosophila melanogaster (subclone D37 5_a4 from P1 DS01068) DNA
            sequence.
ACCESSION   L79874
NID         g1395105
KEYWORDS    Genome Survey Sequences; alcohol dehydrogenase region.
SOURCE      Drosophila melanogaster (tissue library: Subclones in pOT2 from P1
            clone DS01068 (D37)) DNA.
  ORGANISM  Drosophila melanogaster
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Arthropoda;
            Tracheata; Insecta; Pterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 3225)
  AUTHORS   Arcaina,T.T., Bondoc,M.M., Chiang,A., Critz,P.A., Davis,C.A.,
            Doyle,C.M., Ericsson,C.L., Gunning,K.M., Houston,K.A.,
            Jaklevic,M.A., Kadner,K.E., Kim,S.F., Li,M., Lindquist,K.J.,
            Machrus,M.U., Martin,C.H., Mayeda,C.A., Pacleb,J.M.,
            Palazzolo,M.J., Patel,S.G., Santos,R.F., Subramanian,S.,
            Tordilla,M.E., Truong,V., Verhofstadt-Gann,K., Wan,K.H.,
            Whitelaw,K.R., Yee,A., Yeh,R.T. and Yu,C.
  TITLE     Sequencing of the alcohol dehydrogenase region of drosophila
            melanogaster
  JOURNAL   Unpublished (1996)
COMMENT     This subclone overlaps D37 7_c8 and D37 4_e2.
            The P1, from which this subclone is derived, is adjacent to P1
            DS03792 (D31) and DS06238 (D26).
```



```
            Sequence submitted by:
            Human Genome Center and
            Drosophila Genome Center
            Lawrence Berkeley National Laboratory
            Berkeley, CA 94720
            e-mail: seq@genome.lbl.gov.
FEATURES             Location/Qualifiers
     source          1. .3225
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /tissue_lib="Subclones in pOT2 from P1 clone DS01068
                     (D37)"
BASE COUNT      722 a    838 c    809 g    856 t
ORIGIN

  Query Match           6.4%;  Score 19;  DB 18;  Length 3225;
  Best Local Similarity 100.0%;  Pred. No. 5.52e+00;
  Matches   19;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    1745 TATTTTTATTTAAATAACT 1763
           ||||||||||||||||||||
Qy     109 TATTTTTATTTAAATAACT 127


RESULT   14
LOCUS       AF028834     3566 bp     DNA             INV       25-JUN-1998
DEFINITION  Drosophila ananassae vermilion (v) gene, complete cds.
ACCESSION   AF028834
NID         g2612995
KEYWORDS    .
SOURCE      Drosophila ananassae.
  ORGANISM  Drosophila ananassae
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea;
            Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 3566)
  AUTHORS   Stephan,W., Xing,L., Kirby,D.A. and Braverman,J.M.
  TITLE     A test of the background selection hypothesis based on nucleotide
            data from Drosophila ananassae
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 95 (10), 5649-5654 (1998)
  MEDLINE   98245135
REFERENCE   2  (bases 1 to 3566)
  AUTHORS   Stephan,W., Xing,L., Kirby,D.A. and Braverman,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-OCT-1997) Zoology, U. of Maryland, College Park, MD
            20742, USA
FEATURES             Location/Qualifiers
     source          1. .3566
                     /organism="Drosophila ananassae"
                     /db_xref="taxon:7217"
     mRNA            join(<1856. .1881,1971. .2131,2187. .2320,2377. .2986,
```